

©2011

Zhen Hou

ALL RIGHTS RESERVED

SOFTWARE TOOLS FOR MOLECULE-BASED KINETIC MODELING OF
COMPLEX SYSTEMS

by

ZHEN HOU

A Dissertation submitted to the
Graduate School-New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in

Chemical & Biochemical Engineering

written under the direction of

Professor Michael T. Klein

and approved by

New Brunswick, New Jersey

[MAY, 2011]

ABSTRACT OF THE DISSERTATION

Software Tools for Molecule-Based Kinetic Modeling of Complex Systems

By Zhen Hou

Dissertation Director:
Professor Michael T. Klein

Modeling complex process chemistries with complex feedstocks involves several aspects: composition modeling of the complex feedstock, reaction modeling of the complex chemistry, and structure property correlations to provide feed and product property estimation. This thesis was developed to automate these modeling techniques and to provide an integrated approach for combining these modeling aspects into a continuous package.

The first contribution of this thesis is the development of an automated composition modeling tool called the composition model editor (CME). CME uses a statistical hybrid approach to describe a complex feedstock in terms of a set of structural attributes expressed by probability density functions (PDF). Through optimizing a limited set of attribute PDF parameters, CME can obtain a molecular composition array (MCA) for a feedstock based on limited analytical information.

The second contribution of this thesis is the development of a series of automated techniques that are useful for reaction modeling. Firstly, an attribute reaction modeling (ARM) approach is developed for complex process chemistries. ARM can condense a kinetic model by allowing the number of ODEs to be far less than the number of species for a complex system, while maintaining the full molecular detail of the model. Secondly, reaction family and LFER concepts are used to control the number of kinetic

parameters for a complex model. Thirdly, the ability to impose LHHW rate law allows for heterogeneous systems involving with catalysts. Last but not least, various process configurations are addressed to satisfy the kinetic modeling for complex process chemistries.

The third and final contribution of this thesis is the creation of a structure correlation module used to provide data support for kinetic modeling as well as composition modeling. Group contribution methods and the quantum chemistry package are applied to estimate thermodynamic properties. A supplemental database is developed to manage property data in a high efficient way.

The above contributions were then successfully applied to the development of detailed kinetic and feedstock models for complex process chemistries, including complex feedstock characterizations, lignin pyrolysis, and resid pyrolysis.

Acknowledgement

My five-year graduate life at Rutgers was truly enjoyable mainly due to the interactions with many great people. I am grateful for all of their contributions to this work.

At first, my deepest appreciation goes to Mike Klein, an exceptional advisor and mentor, for his continuous guidance, support, encouragement, and trust. What I have learned from Dr Klein is invaluable, both scientifically and professionally. Dr Klein would be the best example for me to learn how to achieve a fantastic career success in my life.

I would also like to thank my advisory committee, Dr. Ben Glasser, Alex Neimark, and Roland Saeger for their contributions to help and insightful comments.

I would like to thank the partners with the Klein research group. They gave me great benefits for the interactions between theoretical methodology and practical applications. Special thanks are due to Mr. Pyl, Steven from Ghent University, Dr. Tanaka, Ryuzo from Idemitsu, and Dr Cheol-Joong Kim from S.K.Energy.

I would also like thank for all members in the Klein research group, past and present, for creating a wonderful work environment during my Ph.D life. I particularly appreciated Dr. Craig Bennett for the help to reaction modeling and the revision on my thesis.

I would like to extend my appreciation to my parents, for their continuous support and encouragement through the years.

Finally, I would show my deepest gratitude to my wife, Lihong, for her continuing love, understanding, and respect. This work would not have been possible without her.

Table of Contents

ABSTRACT OF THE DISSERTATION.....	ii
Acknowledgement.....	iv
Table of Contents.....	v
Lists of tables.....	viii
List of illustrations.....	ix
Chapter 1 Introduction.....	1
1.1 Motivation.....	1
1.1.1 Role of modeling in energy issues.....	1
1.1.2 Limitation of lumped modeling.....	1
1.1.3 Key steps of molecular modeling.....	2
1.1.4 Challenge for complex mixtures and chemistries.....	4
1.2 Limitations of previous work.....	5
1.2.1 Limitations of approaches of Froment, EXXONMOBIL, IFP and UMIST.....	5
1.2.2 Limitations of previous KMT.....	6
1.3 Thesis contributions.....	8
1.4 Thesis Overview.....	9
Chapter 2 Automating the Modeling of Composition.....	12
2.1 Introduction.....	12
2.2 Analytical information of complex feedstocks.....	13
2.3 A statistical view of complex feedstocks.....	14
2.4 Determining the identity of the elements of the MCA.....	16
2.4.1 Bond-Electron matrix representation of molecules.....	21
2.4.2 Core series' structure generator.....	22
2.5 Quantitative determination of molecular compositions.....	26
2.5.1 Probability density functions.....	27
2.5.2 Probability density functions to describe complex mixtures.....	30
2.5.3 Determining appropriate functional forms.....	31
2.5.4 Discretization of the probability density functions.....	34
2.5.5 Renormalization and truncation.....	35
2.5.6 Conditional probability.....	35
2.5.7 Attribute PDF selection.....	36
2.5.8 MCA0 generation.....	37
2.6 Automation of composition modeling-CME.....	39
2.7 Integration between CME and INGen.....	50
2.8 Summary and discussions.....	51
Chapter 3 Automation of reaction equation generation and calculation.....	55
3.1 Introduction.....	55
3.2 Attribute-based reaction model (ARM).....	57
3.2.1 ARM example: lignin.....	58
3.2.2 General illustration of ARM.....	59
3.2.3 Automation of ARM in KME.....	60
3.3 Reaction family and LFER.....	62
3.3.1 LFERs.....	62
3.3.2 Representative results of LFER.....	64
3.3.3 LFERs and reaction families in KME.....	70
3.3.4 Integration between KME and INGen.....	74
3.4 LHHW.....	75
3.5 Deactivation calibration.....	79
3.6 Multiple reactors.....	82

3.7	Basic product separation	83
3.8	Model analysis.....	84
3.8.3	Statistical analysis	85
3.8.4	Sensitivity analysis and model reduction	87
3.9	Summary and discussion	90
Chapter 4	Structure property correlation module.....	94
4.1	Introduction	94
4.2	Pure component property estimation.....	95
4.2.1	Structural properties' estimation	96
4.2.2	Thermodynamic property estimation by group contribution.....	96
4.2.2.1	Selected properties calculated by Joback method	97
4.2.2.2	Selected properties calculated by Rafiqul Gani method.....	100
4.2.2.3	Density calculation for a pure component.....	107
4.2.2.4	Summary of thermodynamic property calculation	108
4.2.3	Selected properties calculated by quantum chemistry.....	109
4.2.3.1	The molecule's conversion from 2D BE matrix to 3D structure.....	109
4.2.3.2	Selected properties calculated by MOPAC	112
4.2.4	Miscellaneous properties:image and chemical name	113
4.2.5	Supplemental property database.....	114
4.2.6	Automation of SPCM for pure components	115
4.3	Bulk property estimation	117
4.4	The reaction's thermodynamic properties of reaction.....	119
4.5	Summary and discussion	120
Chapter 5	Composition modeling of selected complex feedstocks.....	123
5.1	Introduction	123
5.2	Petroleum middle distillates	123
5.2.1	Analytical characterization.....	123
5.2.2	Homologous series determination	126
5.2.3	Structural attribute sampling protocol and PDF form	126
5.2.4	Conditional probability.....	127
5.2.5	Optimization of the PDF parameters	128
5.2.6	Validation results.....	128
5.2.6.1	Case 1 validation results.....	128
5.2.6.2	Case 2 validation results.....	138
5.2.6.3	Conclusion.....	145
5.3	Natural gas condensates	146
5.3.1	Analytical characterization.....	146
5.3.2	Homologous series determination	147
5.3.3	Structural attribute sampling protocol and PDF form	148
5.3.4	Conditional Probability	149
5.3.5	Optimization of the PDF parameters	150
5.3.6	Validation results.....	150
5.3.6.1	Case 1 validation results.....	150
5.3.6.2	Case 2 validation results.....	153
5.3.6.3	Conclusion.....	155
5.4	Petroleum heavy gas oil	155
5.4.1	Analytical characterization.....	156
5.4.2	Homologous series determination	157
5.4.3	Structural attribute sampling protocol and PDF form	157
5.4.4	Conditional probability.....	158
5.4.5	Optimization of the PDF parameters	158
5.4.6	Validation results.....	159
5.4.6.1	Case 2 validation results.....	159

	5.4.6.2	Conclusion.....	166
5.5		Vacuum gas oil.....	166
	5.5.1	Analytical characterization.....	167
	5.5.2	Homologous series determination	167
	5.5.3	Structural attribute sampling protocol and PDF form	168
	5.5.4	Conditional Probability	168
	5.5.5	Optimization of the PDF Parameters.....	169
	5.5.6	Optimization results	170
5.6		Petroleum resids	171
	5.6.1	Analytical characterization.....	171
	5.6.2	Homologous series determination	172
	5.6.3	Structural attribute sampling protocol and PDF form	173
	5.6.4	Conditional probability.....	174
	5.6.5	Optimization of the PDF parameters	174
	5.6.6	Optimization results	175
5.7		Summary and discussions	176
Chapter 6		Modeling lignin pyrolysis with ARM	179
	6.1	Introduction	179
	6.2	The significance of lignin and lignin pyrolysis	180
	6.3	Modeling approach.....	180
	6.4	Modeling Lignin structure and composition	182
	6.5	Lignin reaction pathways and kinetics	187
	6.6	Attribute-based reaction modeling for lignin pyrolysis.....	190
	6.7	Results: application to lignin pyrolysis	195
	6.8	Summary and discussions	198
Chapter 7		Automated modeling of resid pyrolysis with ARM	200
	7.1	Introduction	200
	7.2	Modeling approach.....	200
	7.3	Model development.....	201
	7.3.1	Resid composition modeling	201
	7.3.2	Reactive attributes determination	201
	7.3.3	Resid pyrolysis reaction network	206
	7.3.4	Equations and rate constants of resid pyrolysis.....	209
	7.3.5	Regeneration of the attribute PDF's	212
	7.3.6	Product resampling	213
	7.3.7	Model predictions.....	213
	7.4	Summary and discussion	217
Chapter 8		Summary and discussions	220
	8.1	Summary	220
	8.1.1	Automating the modeling of momposition (CME)	220
	8.1.2	Automation of reaction equation generation and calculation	222
	8.1.3	Structure property correlation module.....	224
	8.1.4	Integration of KMT components	225
	8.1.5	Composition modeling of selected complex feedstocks.....	226
	8.1.6	Modeling lignin pyrolysis with ARM	228
	8.1.7	Automated modeling of resid pyrolysis with ARM	228
	8.2	Suggestions for future	228
	8.2.1	Automating the modeling of composition (CME).....	228
	8.2.2	Automation of reaction equation generation and calculation	230
	8.2.3	Structure property correlation module.....	230

Lists of tables

Table 1.1 Thesis contributions V.S. previous KMT work	9
Table 2.1 Quantitative Attribute PDFs Selection of Hydrocarbon Components in Petroleum	16
Table 2.2 Probability density functions used to model the structural attributes of a complex feedstock.	28
Table 3.1 LHHW formalism (Yang and Hougen, 1950).....	76
Table 3.2 Multiple reactors in proposed KME.....	83
Table 4.1 Pure component structural properties caculated in SPCM.....	96
Table 4.2 Thermodynamics properties in Joback method.....	97
Table 4.3 The structural groups in Jack method.....	98
Table 4.4 Thermodynamics properties in Rafiqul Gani method	100
Table 4.5 The first order groups and their contributions in Gani method	101
Table 4.6 The second order group and their contributions in Gani method	104
Table 4.7 The third order group and their contributions in Gani method.....	106
Table 4.8 The additional parameters in E.q.4.2.....	107
Table 4.9 Internal coordinate representation of ethane	110
Table 4.10 Typical atom data as internal coordinate.....	111
Table 5.1 Bulk properties for petroleum middle distillates	124
Table 5.2 Model predictions for the analytical properties of two middle distillates with case 1.	129
Table 5.3 Typical weight distribution of branched isomers of C ₈ ³	130
Table 5.4 the overall statistical analysis for middle distillate Sample1 and Sample2 with case 1	130
Table 5.5 Model predictions for the analytical properties of two middle distillates with case 2.	139
Table 5.6 the overall statistical analysis for middle distillate Sample1 and Sample2 with case 2.	140
Table 5.7 Bulk properties for natural gas condensates	147
Table 5.8 Model predictions for the analytical properties of two natural gas condensates with case 1	151
Table 5.9 Model predictions for the analytical properties of two natural gas condensates with case 2.....	153
Table 5.10 Bulk properties for the HGO sample.....	156
Table 5.11 Model predictions for the analytical properties of the HGO sample with case 2.	160
Table 5.12 Bulk properties for VGO.....	167
Table 5.13 Model optimization results for the VGO sample	171
Table 5.14 Bulk properties for resids	172
Table 5.15 Model optimization results for the resids' samples	176
Table 6.1 Model Compounds for Lignin Pyrolysis ²	188
Table 7.1 Reactive attributes for various chemistries ¹	202
Table 7.2 The carbon number ranges of the reactive attributes in a resid pyrolysis model.....	204
Table 7.3 The statistics of the reaction network for resid pyrolysis with ARM.....	209
Table 7.4 Product properties of DaQing resid model.....	215
Table 7.5 Product properties of ShengLi resid model.....	216

List of illustrations

Figure 1.1 Molecular-level Modeling steps.....	2
Figure 1.2 KMT automated modeling toolbox.....	6
Figure 2.1 MCA/MCA0 example of reforming feedstock	13
Figure 2.2 Analytical in order Underlying MCA Representation	14
Figure 2.3 Generically Determine Identities in Hydrocarbon	18
Figure 2.4 Determine Aromatic/Resin Cores (Series).....	19
Figure 2.5 Determine Identities for light Oil to Gasoil (Expert determination).....	20
Figure 2.6 Bond-Electron Matrix for pentane	22
Figure 2.7 Series Core Generator	23
Figure 2.8 Multiple Attribute PDFs Sampling Defines Molecules	27
Figure 2.9 Examples of probability density functions.....	29
Figure 2.10 Relative boiling point intensity for kerosene and vacuum resid petroleum fractions	32
Figure 2.11 Examples of exponential, gamma and chi-square distributions	34
Figure 2.12 Overall logic of CME for composition modeling	40
Figure 2.13 automation of Core generator in CME.....	41
Figure 2.14 Identity Setup in CME	42
Figure 2.15 Identity and Sampling Setup in CME	43
Figure 2.16 an example representation of a feedstock's footprint	44
Figure 2.17 Selection cases for MCA0 generation in CME	45
Figure 2.18 Attribute PDF Type Selection.....	46
Figure 2.19 Grouping SC PDFs	47
Figure 2.20 Property Selection in CME	48
Figure 2.21 KME ready results view.....	49
Figure 2.22 PONA Matrix results view.....	50
Figure 3.1 the structure of lignin.....	58
Figure 3.2 Attribute-Based Reaction Modeling	60
Figure 3.3 Conceptual flow of KME with ARM.....	61
Figure 3.4 LFER correlations of experimental hydrogenation rate constants and heat of reaction for aromatic compounds. (Korre ¹⁰ , 1995).....	64
Figure 3.5 LFER correlation of the experimental isomerization and ring opening rate constants with the stability of the carbenium ion intermediates. (Korre ¹⁰ , 1995).....	65
Figure 3.6 LFER correlation of experimental dealkylation rate constant with the heat of formation of the intermediate carbenium alkyl ion. (Data from Mochida and Yoneda ⁴ , 1967)	66
Figure 3.7 LFER correlation of proton affinity (PA) to estimate the adsorption constants K. (Neurock ⁹ , 1992)	67
Figure 3.8 LFER for adsorption constants by Klein group.	68
Figure 3.9 LFER for metal center catalysis by Klein group.....	69
Figure 3.10 LFER for acid center catalysis by Klein group.....	70
Figure 3.11 LFER & RXN Family in KME	73
Figure 3.12 Rate Law by Reaction Family & LFER.....	73
Figure 3.13 Importing an InGen File to KME.....	74
Figure 3.14 Automation of LHHW in KME	79
Figure 3.15 the conceptual flow of deactivation in oncethrough model	81
Figure 3.16 the conceptual flow of deactivation in goal seeking model	82
Figure 3.17 Principles of basic separation model.....	84
Figure 4.1 The role of structure property correlation in KMT	94
Figure 4.2 Structure-Property Correlations for pure Component.....	116
Figure 4.3 Synchronization of the Property Database via a user friendly interface	117
Figure 5.1 Comprehensive 2D GC ²	125
Figure 5.2 Selected homologous series for middle distillates	126
Figure 5.3 sampling protocol for middle distillates.....	127

Figure 5.4 Predicted and experimental carbon number distribution for the paraffin fraction of middle distillate Sample1 with case 1.	131
Figure 5.5 Predicted and experimental carbon number distribution for the isoparaffin fraction of middle distillate Sample1 with case 1.	132
Figure 5.6 Predicted and experimental carbon number distribution for the naphthenic fraction of middle distillate Sample1 with case 1.	133
Figure 5.7 Predicted and experimentally determined carbon number distributions of the aromatic fraction of middle distillate Sample1 with case 1.	134
Figure 5.8 Predicted and experimental carbon number distribution for the paraffin fraction of middle distillate Sample2 with case 1.	135
Figure 5.9 Predicted and experimental carbon number distribution for the isoparaffin fraction of middle distillate Sample2 with case 1.	136
Figure 5.10 Predicted and experimental carbon number distribution for the naphthenic fraction of middle distillate Sample2 with case 1.	137
Figure 5.11 Predicted and experimentally determined carbon number distributions of the aromatic fraction of middle distillate Sample2 with case 1.	138
Figure 5.12 Predicted and experimental carbon number distribution for the paraffin fraction of middle distillate Sample1 with case 2.	140
Figure 5.13 Predicted and experimental carbon number distribution for the isoparaffin fraction of middle distillate Sample1 with case 2.	141
Figure 5.14 Predicted and experimental carbon number distribution for the naphthenic fraction of middle distillate Sample1 with case 2.	141
Figure 5.15 Predicted and experimentally determined carbon number distributions of the aromatic fraction of middle distillate Sample1 with case 2.	142
Figure 5.16 Predicted and experimental carbon number distribution for the paraffin fraction of middle distillate Sample2 with case 2.	143
Figure 5.17 Predicted and experimental carbon number distribution for the isoparaffin fraction of middle distillate Sample2 with case 2.	143
Figure 5.18 Predicted and experimental carbon number distribution for the naphthenic fraction of middle distillate Sample2 with case 2.	144
Figure 5.19 Predicted and experimentally determined carbon number distributions of the aromatic fraction of middle distillate Sample1 with case 2.	145
Figure 5.20 Selected homologous series for natural gas condensates	148
Figure 5.21 sampling protocol for natural gas condensates	149
Figure 5.22 Calculated and experimental group composition data for GC 663 for case 1.....	152
Figure 5.23 Calculated and experimental group composition data for GC659 with case 1	152
Figure 5.24 Calculated and experimental group GC-MS data for GC 663 with case 2	154
Figure 5.25 Calculated and experimental group GC-MS data for GC659 with case 2	155
Figure 5.26 Selected homologous series for HGO	157
Figure 5.27 Predicted and experimental carbon number distribution for the paraffin fraction of HGO sample with case 2.	161
Figure 5.28 Predicted and experimental carbon number distribution for the isoparaffin fraction of HGO sample with case 2.	162
Figure 5.29 Predicted and experimental carbon number distribution for the naphthenic fraction of HGO sample with case 2.	163
Figure 5.30 Predicted and experimentally determined carbon number distributions of the mono,di-aromatic fraction of the HGO sample with case 2.	164
Figure 5.31 Predicted and experimentally determined carbon number distributions of the tri+-aromatic fraction of the HGO sample with case 2.	164
Figure 5.32 Predicted and experimentally determined carbon number distributions of the benzothiophene fraction of the HGO sample with case 2.	165
Figure 5.33 Predicted and experimentally determined carbon number distributions of the di--benzothiophene fraction of the HGO sample with case 2.....	166

Figure 5.34 Selected homologous series for VGO	168
Figure 5.35 Selected homologous series for resids	173
Figure 5.36 sampling protocol for resids.....	174
Figure 6.1 Molecular-level Modeling steps.....	181
Figure 6.2 Lignin monomer alcohols	182
Figure 6.3 the Freudenberg model of lignin structure	183
Figure 6.4 Initial Lignin (Freudenberg) Structure in Terms of MP and PC Attributes	184
Figure 6.5 Attributes Sampling in CME	185
Figure 6.6 Flow sheet of CME strategy for modeling lignin structure.....	187
Figure 6.7 VGE Pyrolysis Pathways	189
Figure 6.8 PDF sampling for molecular compositions representation	191
Figure 6.9 the conceptual flowsheet of KME with ARM for lignin pyrolysis	193
Figure 6.10 KME options.....	194
Figure 6.11 Two ARM Reaction Sets in KME	194
Figure 6.12 Pyrolysis products yield profile along PFR	196
Figure 6.13 Asymptotic yields of key products parametric in the coniferyl/sinapyl alcohol ratio in starting lignin	197
Figure 7.1 Free radical mechanism of alkyl aromatics pyrolysis	206
Figure 7.2 An alkyl-aromatics pyrolysis pathway model (PentaDecyl Benzene-PDB)	208
Figure 7.3 PDF sampling for molecular compositions representation	210
Figure 7.4 The product distribution of DaQing resid model.	214
Figure 7.5 The product distribution of ShengLi resid model.	216
Figure 8.1 Elements of KMT	226

Chapter 1 Introduction

1.1 Motivation

1.1.1 Role of modeling in energy issues

Due to the depletion of traditional oil resources, the supply of energy has been regarded as a paramount global issue. This impending scarcity has led to increased interest in the search for alternate feedstocks as well as the deep utilization of heavy oil. The use of such feeds is complicated by not only the economics of upgrading and conversion, but also the environmental footprint, including the impact on CO₂ issues. These alternate heavy feeds are commonly complex mixtures of complex molecules often including covalent or physically aggregated macromolecules. A US DOE Office of Basic Energy Sciences report on basic research needs for catalysis for energy applications¹ states that it is critical to identify the structures and reaction pathways of these heavy feeds and to develop robust computational tools to model their reaction trajectories. Work towards this goal would contribute not only to the utilization of these feeds but also to the design of catalysts able to produce efficient and environmentally optimal outcomes. Particularly, the role of modeling in the production of these energy and fuels can undertake two aspects. One is to achieve general engineering goals such as: the prediction of the product species, yields and properties; the help for industrial design; linear programming; advanced process control; RT-OPT and so on. The second aspect is that the quantitative understanding of complex chemistries can provide help for the development of new catalysts, new solvents and new processing strategies.

1.1.2 Limitation of lumped modeling

Traditionally, most complex processes were implemented by lumped model schemes, where the molecules are grouped by global properties, such as boiling point or solubility. The limitations of these

global lumped models prevented the fundamental understanding of the complex process chemistry. A lumped model lacks the chemical structure of its species; molecular level information is obscured because of the multi-component nature of each lump. Consequently, this approach can not provide properties beyond the definition of lump. Thus, un-lumped or molecular modeling solutions will need to be employed find this level of detail.

1.1.3 Key steps of molecular modeling

As a significant improvement on lumped models, molecular level models will be an optimal starting point to understand both process and chemistry research and development. Because molecules are the basis for feedstock composition, property calculation, process chemistry, reaction kinetics and thermodynamics, molecule-based models can incorporate information from surface and quantum chemical calculations all the way to processing issues. Molecular level modeling is composed of the four main components listed in Figure 1.1.

1. Structural and Composition Models

- Measurements (GS-MS, NMR, etc.)
- Measurements to Molecules

2. Reaction Modeling

- Computer generated models in seconds

3. Reactivity Correlations

- Fundamentals based Chem Eng LFER kinetics correlations
- Order 10 [O(10)] LFER's for every process chemistry

4. Property Estimation

- Provides commercially relevant product yields
- Molecules to commercial lumps
- End-use vs. internal-use properties

Figure 1.1 Molecular-level Modeling steps

The first component is the structure-composition model for the feed. This information provides the initial conditions for the equations that represent the kinetics. For simple feeds, the molecular details

usually obtained by a single direct measurement. For complex feeds, this step often involves the transformation of a set of measurements into structural representations and quantitative mol fractions. For heavy feedstocks, the measurements are often global in nature, such as NMR, elemental and SIMDIS analyses. These global measurements are often supported by more detailed analyses, such as those from various types of mass spectrometry analyses on separated fractions.

Once the feed stock has been modeled, the second necessary component calls for the creation of a reaction model wherein reactants are transformed into products using experimentally discerned reaction pathways. After the reaction lists are obtained, each reaction must be converted into a mathematic equation which can be compiled into computer code and handled by a numerical solver. Although most complex feeds contain thousands of species and reactions, generally there are only of order 10 different kinds of reactions or reaction families. These reaction families can be used to organize both the reaction pathways and, in favorable circumstances, the reaction kinetics.

Simply applying each of the reaction families to the reactive components of the feed and products can create a network of thousands of reactants and rate laws. In these instances, it is often helpful to reduce the number of model parameters by constraining the kinetic, adsorption and occasionally thermodynamic constants in the rate laws for each reaction family to follow a correlation based on either the Linear Free Energy or van't Hoff relationships. These reactivity correlations cover the third component of molecular modeling. For very complex systems, these correlations can reduce the parameter burden from thousands to of order 10-30. For simpler systems, such reduction may not be necessary and independent individual parameters will be sought.

The final component in building a molecular model is the property estimation function, which has two key roles. First, it provides the connection between the molecular composition and the set of end-use product properties that motivated the development of the model. Secondly, it also serves to provide

intermediate properties that are used in the construction and solution of the model, such as the enthalpy of formation for key intermediate species.

1.1.4 Challenge for complex mixtures and chemistries

Although the potential advantages of molecule-based modeling are clear, the development and operation of molecular models come with large model construction and solution time burdens respectively. In addition, a greater understanding of reactivity information is necessary. The essential challenge of building detailed kinetic models for heavy feeds arises from the staggering complexity of not only the reaction mixtures but the complexity of each molecule within the mixture. Heavy complex feeds often have low volatility and complete sampling from the current analytical chemistry techniques is not possible. Therefore, it is hard to obtain a full set of molecular representations by direct measurements for such complex mixtures. In addition, there will often be thousands of “multi-functional” component species. The sheer size of the thus-implied modeling problem engenders a conflict between the need for molecular detail and the formulation and solution of the model. Generally, such a heavy complex mixture like petroleum resid will contain 50,000 or more molecular species. Traditional deterministic reaction models will consist of one differential equation for each species, and the numerical burden of solving 50,000 simultaneous ODE's is beyond the upper limit of what is considered practical. For some light-middle end of complex feeds containing $O(10^3-10^4)$ species, such as VGO oil, the computational burden can be handled by the current computer technique, but the solution time is still greater than some practical industrial applications would allow: advanced control, RT-OPT, etc. Also, the increasing number of parameters corresponding to the large number of equations is also an intractable issue.

Besides the challenge of the complexity of the heavy feed, the variety of data from complex process configurations in practical process chemistries is a difficult problem. Unlike the experimental data generated by an academic laboratory, the necessary data can come from different catalyst deactivation profiles, multiple reactor types, complex process flows, etc.

Consequently, the issue of complex feed and process chemistries brings out practical questions for molecular level modeling. Can a molecular model be created such that complex process chemistries of complex mixtures overcome the computational burden while maintaining the full molecular information of the model and still fit the real process specifications of the data source? This thesis is motivated to address the above dilemma and provide an integrated approach to fulfill the automation of molecular-based complex modeling.

1.2 Limitations of previous work

1.2.1 Limitations of approaches of Froment, EXXONMOBIL, IFP and UMIST

In the area of automated detailed kinetic modeling for complex process chemistries, several research groups have addressed various detailed modeling concerns with their comprehensive and elegant works such as those of Froment's group, EXXONMOBIL, IFP and UMIST. Froment² and co-workers developed a single-event approach via graph theory that can build fundamental kinetic models at the mechanistic level. However, the model building is computationally expensive, and the memory demand requires that it be limited to a low carbon number range. Quann and Jaffe^{3,4} of EXXONMOBIL developed the Structure Oriented Lumping (SOL) approach that uses vectors for structural groups. SOL does not explicitly represent a molecule's atoms or the atom's bond connectivity; it only characterizes the hydrocarbon conversion. The SOL approach is thus limited to pathways level modeling and can not build mechanistic models. The researchers in UMIST developed a MTHS (molecular type homologous series) matrix to represent the species and built kinetics model based on this MTHS representation. The UMIST (University of Manchester Institute of Science and Technology)⁵ approach omits the isomeric detail and limits application to only predefined molecular types, which is hard to extend to heavy feeds. The researchers at IFP started with Froment's approach and developed a stochastic kinetic modeling based on it. The limitations of the IFP approach are similar to these of Froment's approach.

1.2.2 Limitations of previous KMT

Since 1994, the Klein group has been developing the Kinetic Modeler's Toolbox (KMT) software for the automated building of molecule based models for complex systems. Figure 1.2 shows the three main functional components.

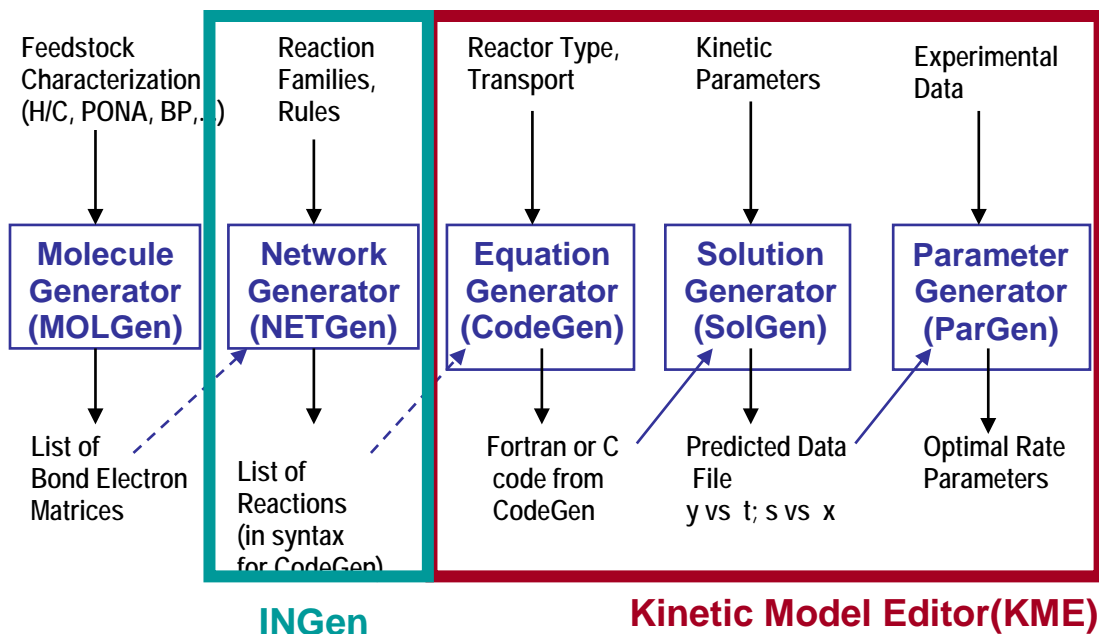


Figure 1.2 KMT automated modeling toolbox

KMT began with MolGen^{6,7} for feedstock characterization. MolGen used a set of probability density functions (PDF) for the elemental structural attributes to represent the feedstock. MolGen stochastically sampled those attributes' values via Monte Carlo simulation and thus provided a qualitative and quantitative representation for complex feedstocks. However, MolGen is entirely dependent on Monte Carlo simulation. Monte Carlo simulation is extremely time consuming for complex feedstocks and lacks any sort of user interaction or direction. Due to the diversity of feedstocks, each complex feedstock would need to be tediously sampled. Since MolGen was not designed as a universal algorithm

capable of modeling many types of feedstocks, different feedstock dependent versions of MolGen exist. In addition, MolGen is an isolated, hard-coded program and thus difficult for users to change to fit their desired feed types. Therefore, it is impractical to use for a large variety of complex feedstocks. As a purely stochastic sampling program, MolGen can only use a quadrature approximation to provide a set of molecules for kinetic modeling. This tremendously heavy task involves a good deal of further programming; and it may lose the necessary molecular information. As a result, MolGen could not be automated to work with the other two components of KMT: INGen and KME. Particularly, it can not provide the initial conditions for subsequent model calibration. The above obstacles block the practical use of MolGen in automating the modeling of complex feedstocks.

NetGen is used to obtain the reaction networks. The reactants are represented by Bond-Electron matrices. Chemical transformations, namely the bond-making and bond-breaking processes, are implemented as reaction matrices. A reactant to product relationship can be derived by performing a matrix addition of the reaction matrix to the reactant matrix in order to obtain a product species matrix. NetGen has the ability to automatically generate a reaction network for a specific chemistry. Based on the work of Broadbelt⁸, Joshi⁹ and Hou¹⁰'s work, Bennett¹¹ developed INGen, an improved version of NetGen, as an interactive, universal reaction network generator for hydrocarbons. INGen's user-friendly automation of reaction network generating makes it possible to integrate into a suite of automated tools for complex feedstocks and reaction modeling.

The final component of KMT is the Kinetics Model Editor (KME)¹² which is most frequently used for model development. KME parses reaction lists from INGen and user entry into compilable C-code and it provides a friendly automated environment for model tuning, simulating and goalseeking. However, KME could not handle the complex process chemistries of complex mixtures that are so often found outside on an academic laboratory. Generally, complex chemistries will require of order 100 or more equations and hundreds of kinetics parameters. Since KME only supported the tuning of individual reactions, calibration proved to be too difficult for complex chemistries. In addition, KME could only

support a single stoichiometric rate law, thereby invalidating its use for most practical process chemistries that involved a heterogeneous catalyst system. Finally, the data from real processes often included an inherent catalyst deactivation profile, and could possibly be obtained by various types of reactors, neither of which could be handled by KME.

Overall, applying these existing tools to heavy feedstocks would bring about tremendous computational and temporal burdens when developing a traditional deterministic model that uses one equation per species. Modeling a complex process with thousands of species requires not only the implementation of reaction equations but the auxiliary acquisition of the necessary data. Creating a simple automated system that overcomes these existing shortcomings would prove to be indispensable.

1.3 Thesis contributions

Although the previous KMT elements provided an excellent foundation for detailed kinetics modeling, they did not satisfy the issues listed above for the complex process chemistries. This thesis approaches each of these issues in the context of building an integrated system designed to help automate the molecular modeling of complex feedstocks and chemistries. The contributions of this thesis and the limitation of the previous KMT are listed in Table 1.1 in terms of the three aspects of molecular modeling: automation of composition modeling, automation of reaction equation generation and calculation, and structure property correlation modeling.

Table 1.1 Thesis contributions V.S. previous KMT work

	Limitation	Contribution
Automation of composition modeling	<p>Molgen:</p> <ul style="list-style-type: none"> Diversity of identities from various feedstocks hard coded, isolated program for different feedstocks Long time consuming of Monte Carlo simulation Pure computer simulation, lack user interaction Weakly connected with INGen and KME ,unable to get KME initial data 	<p>CME:</p> <ul style="list-style-type: none"> Generic represent hydrocarbon mixtures United bundle tool with excel user friendly interface Efficient Combination of fixed identity determination and Monte Carlo simulation computer aid and expert interaction Well communicated with KME and INGen, provide KME initial data Enable BE Matrix representation of molecules
Automation of reaction equation generation and calculation	<ul style="list-style-type: none"> Difficulty for complex system <ul style="list-style-type: none"> $O(10^5)$ equations by deterministic model Difficulty derive from a large number of parameters to tune via individual reactions Only support micro kinetics ratelaw Difficulty derive from process configurations <ul style="list-style-type: none"> Data came from deactivation process Data came from various reactors (PFR, CSTR, Batch etc.) 	<ul style="list-style-type: none"> ARM allows to handle $N_{\text{equations}} < N_{\text{molecules}}$ Uses reaction family ($O(10)$) parameters for every process chemistry Uses LHHW rate law for heterogeneous system Provides Deactivation process Provides multiple reactors Provides model analysis
Structure property correlation model	<ul style="list-style-type: none"> Lacking feed representation properties Lacking end-use properties Lacking reactivity data (dH, dS) 	<ul style="list-style-type: none"> Structure property correlation <ul style="list-style-type: none"> Based on molecule structure Provide end-use properties Provide reactivity data(dH,dS etc.)

1.4 Thesis Overview

The thesis is divided into two parts. The first part illustrates the conceptual methodology and tool development used for this integrated approach. Specifically, Chapter 2 addresses the user friendly automation of composition modeling to obtain both the qualitative and quantitative molecular information for a complex feedstock in the presence of limited analytical data. Chapter 3 chronicles the

addition of several modeling aspects to the KME tool. Chapter 4 develops the structure property correlations used to provide end-use properties and reactivity data by fragmental group contribution and quantum chemistry method. In addition, a supplemental property database is addressed.

The second part of this thesis presents various example applications of this modeling approach, including: composition modeling of selected complex feedstocks (Chapter 5), lignin pyrolysis with ARM (Chapter 6), and resid pyrolysis (Chapter 7). Finally, in Chapter 8, the key contributions of this work are summarized and the suggestions for future work in the area of automated detailed kinetic modeling of complex processes are discussed.

Reference

-
- ¹ Bell, A. T.; Gates, B. C.; Ray, D., Co-Chairs, Basic Research Needs: Catalysis for Energy, Report from the U. S. Department of Energy, Basic Energy Sciences Workshop, August 6-8, 2007.
- ² Froment, G.F., “Fundamental Kinetic Modeling of Complex Refinery Process on Acid Catalysts”, The Kurt Wohl Memorial Lecture, University of Delaware (1999)
- ³ Quann, R. J., and Jaffe, S. B., "Structure Oriented Lumping. Describing the chemistry of complex hydrocarbon mixtures", Ind. Eng. Chem. Res., 31(11), pp. 2483-2497 (1992)
- ⁴ Quann, R. J., and Jaffe, S. B., “Building Useful Models of Complex Reaction Systems in Petroleum Refining”, Chem. Eng. Sci., 51(10), pp. 1615 (1996)
- ⁵ Peng, B.; Towler, G. (Supervisor)., Molecular Modelling of Petroleum Processes., Ph.D. Dissertation, University of Manchester Institute of Science and Technology, Manchester, U.K., 1999; pp22-41.
- ⁶ Campbell, D.M., Stochastic Modeling of Structure and Reaction in Hydrocarbon Conversion, Doctoral Dissertation, University of Delaware (1998)
- ⁷ Trauth, D.M., Structure of Complex Mixtures through Characterization, Reaction, and Modeling, Ph. D. Dissertation, University of Delaware, 1993.
- ⁸ Broadbelt, L. J., Stark, S. M., and Klein, M. T., Computers Chem. Eng., 20, 2, pp. 113-129 (1996)
- ⁹ Joshi, P.V, “Molecular and Mechanistic Modeling of Complex Process Chemistries”, Ph.D. Dissertation, University of Delaware (1998)
- ¹⁰ Gang, Hou, Integrated chemical engineering tools for the building, solution, and delivery of detailed kinetic models and their industrial applications, Doctoral Dissertation, University of Delaware (2001)
- ¹¹ Craig, Bennett, “User-controlled kinetic network generation with INGen”, Doctoral Dissertation, Rutgers University(2009)
- ¹² Wei, W “The Interface of Chemical Engineering and IT in Kinetics Models”, Doctoral Dissertation, Rutgers University(2004)

Chapter 2 Automating the Modeling of Composition

2.1 Introduction

As mentioned in Chapter 1, the first conceptual step of modeling complex process system with complex mixtures is to figure out the representation of composition details that will provide initial data for the latter kinetic modeling. This chapter mainly addresses the first contribution of this thesis, which will use a generic representation called the molecular composition array (MCA) to characterize the detailed molecule-based compositions in a complex process system. The MCA describes the structure of molecules as well as the mole fraction in the hydrocarbon mixtures. Particularly, it uses MCA0 as the sub-items of MCA to describe the qualitative and quantitative representation of reactants in the feedstock, which will serve as the initial conditions to kinetic model bundle such as KME. The MCA0 later can expand to full MCA details in the post-reaction products as the intermediate elements that exchange the full detailed species of complex mixtures in this thesis. An example of MCA0/MCA for a reforming naphtha sample is listed in Figure 2.1. The final goal of this chapter is to illustrate the determination of MCA0 from data provided by analytical chemistry.

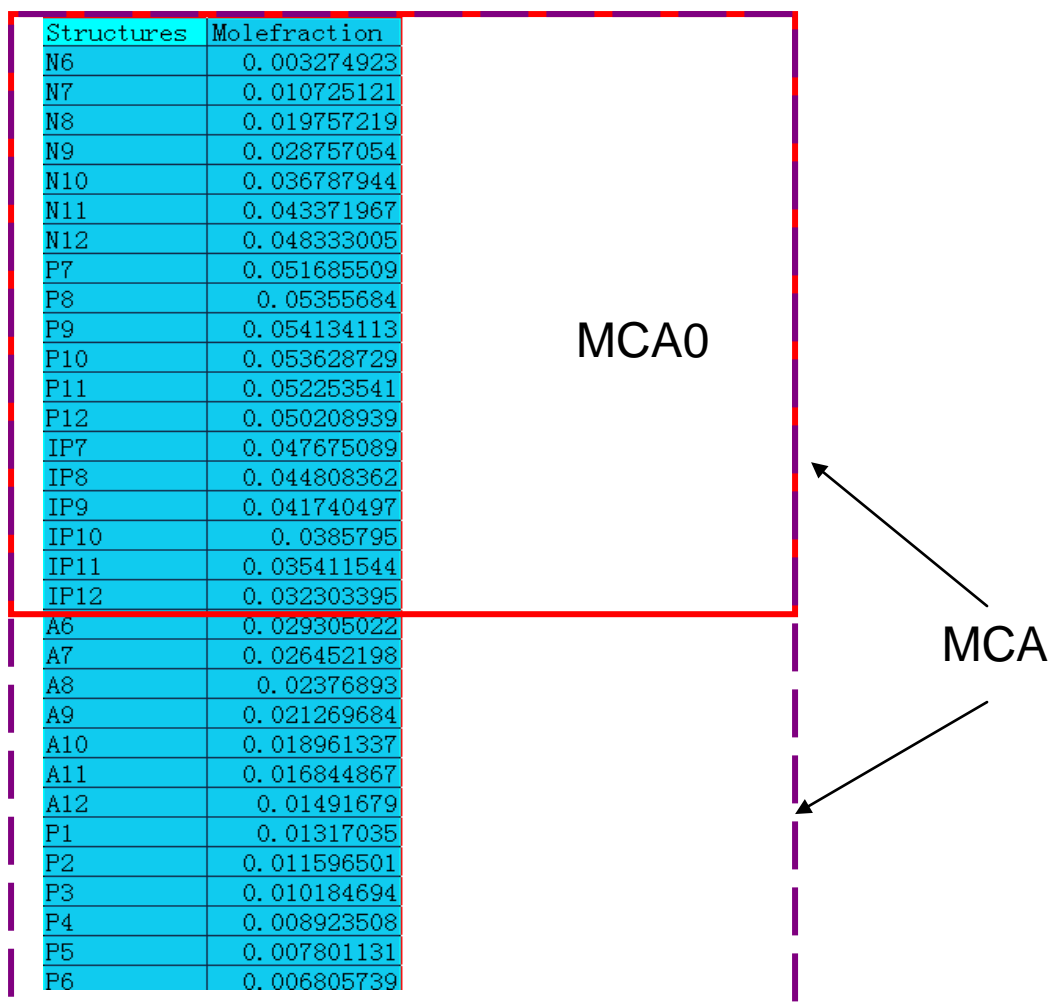


Figure 2.1 MCA/MCA0 example of reforming feedstock

2.2 Analytical information of complex feedstocks

Due to the diversity of complex feedstocks, the available measurements of analytical chemistry are feedstock specific. Typical analytical data and techniques for hydrocarbon from light oil to resid are shown in Figure 2.2. These measurements can be used to obtain some molecularly critical information that can be used to determine the identities of MCA0 of a specific feedstock. Some measurements will be directly applicable on a molecule basis (e.g., elemental analysis). Other measurements will be applicable by an indirect molecular simulation or correlation (e.g., simulated distillation). In addition, some measurements only capture the bulk properties of the feedstock. For these cases, the individual

molecules' properties can be calculated and combined in to bulk properties via mixing rules. A much more complete overview of this method can be found in Campbell's thesis (1998).

● Light Oil	● Gas Oil	● Resid
— MW, Elemental	— MW, Elemental	— MW, Elemental
— PONA fractions and MWs	— PONA fractions and MWs	— SARA fractions and MWs
— Simulated Distillation	— Simulated Distillation	— Simulated Distillation
— GC-MS	— NMR	— Proton NMR
		— ^{13}C NMR
		— Analytical

Figure 2.2 Analytical in order Underlying MCA Representation

Ideally, the preferable way of attaining the identities of MCA0 is directly from the analytical measurements. However, this method only works for light petroleum fractions, such as naphtha with advanced analytical techniques (e.g., Detailed Hydrocarbon Analysis-DHA). Beyond C10, the number of possible isomers alone prevents direct identification. Thus, the most direct measurements of MCA0 for complex feedstocks with a high range of carbon number will be extreme costly if not impossible. Therefore, methods to convert indirect analytical information to MCA0 are the essential issue for the composition characterization of complex process chemistry. To address this methodology, the following section will illustrate how to obtain the qualitative information of MCA first.

2.3 A statistical view of complex feedstocks

The first step in determining the qualitative nature of MCA requires a statistical description of the feedstock. A useful statistical view of the compositions of a complex feedstock is regarded as a collection of elemental structural attributes that indicate irreducible significant chemical criteria. Traditional petroleum hydrocarbons can be described as these structural attributes: number of aromatic

rings, number of naphthenic rings, numbers of alkyl side chains, length of side chains, etc. For example, the alkyl benzene series can be described by number of aromatic rings, numbers of alkyl side chains and length of side chains. Detailed information of structural attributes in traditional hydrocarbons is listed in Table 2.1. Lignin, a component of biomass, can be described by two attributes: methoxy phenols (MP) and propanoid side chains (PC). Any molecule in a feedstock can be viewed as a juxtaposition of those structural attributes. From a statistical view, each of these attributes would be represented by a probability density function (PDF) that can be either a continuous distribution function or discrete values. The PDF provides the probability of finding the value of a given attribute. By sampling those attributes with juxtaposition shown in later section, the values of the structural attributes can be determined for an individual molecule which in turn specifies the identity of the molecule. Based on this statistical view of complex feedstocks, a hybrid approach combining the fixed identity determination and stochastic sampling can be employed to determine the identities of MCA0.

Table 2.1 Quantitative Attribute PDFs Selection of Hydrocarbon Components in Petroleum

Light oil	Middle-heavy oil	Heavy-extreme heavy oil
<p>A set of discrete molecules:</p> <ul style="list-style-type: none"> • Normal paraffin • Iso paraffin with detailed isomers • Naphthenic ring with sidechain and substituent • Aromatic ring with sidechain and substituent 	<p>A set of discrete values for the global fractions of homologous series:</p> <ul style="list-style-type: none"> • Normal paraffin • Iso paraffin with various branches • Naphthenic rings with five member or six member • Aromatic rings • Thiophenic rings <p>A set of continuous attributes' PDFs:</p> <ul style="list-style-type: none"> • Continuous Attribute PDF for different carbon number components for each series (sidechains) • Continuous Attribute PDF for substituents 	<p>Represent molecules by structural attributes's PDFs</p> <p>Nine Resid Attributes:</p> <ul style="list-style-type: none"> • Paraffin Length • Number of Naphthenic Rings • Number of Sidechains • Length of Sidechains • Number of Aromatic Rings • Number of Thiophenic Rings • Number of Naphthenic Rings on an Aromatic Core • Number of Sidechain Sulfurs • Asphaltene Degree of Polymerization

2.4 Determining the identity of the elements of the MCA

With the development of analytical chemistry for heavy feeds, such as HPLC, and the increasing accumulated experience found in literatures, an increasing number of structures have been fully recognized in the complex feedstock. However, for middle to heavy end feedstocks, ring structures and aggregated macromolecule with multiple rings and heteroatoms (S, N, and O) are still difficult to

identify. Unfortunately, those structures significantly affect on the reactivity of the complex process chemistry, while non ring compositions such as paraffin/iso paraffin, alkyl sidechain are relatively easy to handle.

Consequently, unlike the pure stochastic approach of MolGen, a hybrid statistical approach is applied in this thesis. This approach will combine a “fixed” identity determination with stochastic sampling to acquire the qualitative information of feedstocks. This thesis creates a “lumped” structural attribute concept called “core series”. A core series is comprised of the elemental structural attributes that describe ring structures and aggregated molecules in the complex feedstock. The core series is a set of the unique group structures that determines the reaction path and reactivity of the complex system. For example, the core series of alkyl-benzene is one single aromatic ring. A complex feedstock can be described as a set of homologous core series with different sidechains and substituents. The non-ring components can also be considered as a “special” core series in the system. Those core series can be acquired from chemists, literatures and predefined reaction networks (e.g. information from INGen).

For extremely complex feeds (resids, coal etc.), the identity of core series can not be determined explicitly. However, a sampling rule can be established in order to construct the homologous series for the feedstock via the knowledge of chemistry, previous model experience, contemporary analytical techniques, the literature and so on. In general, such a sampling rule can also be applied to the case that the identities of the core series could have been determined explicitly.

A sampling rule for traditional petroleum hydrocarbon mixtures based on elemental structural attributes is listed in Figure 2.3. The first step determines the molecule’s type. Subsequent steps identify the core ring structures through several elemental ring structural attributes based on the specific molecule’ type. Finally, the side chain length/carbon number and the number of side chains are determined.

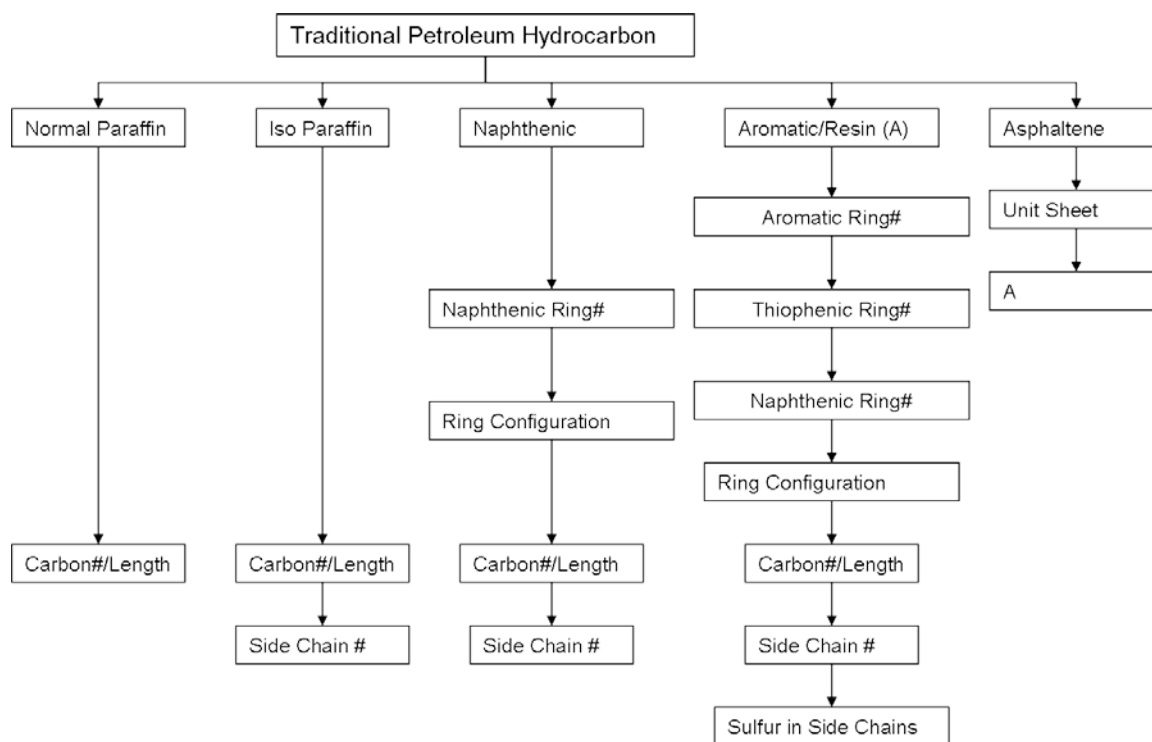


Figure 2.3 Generically Determine Identities in Hydrocarbon

As an example, an aromatic core will be generated using this structural attribute logic. First, the core will be given the number of pure aromatic rings using an attribute PDF. Then the number of heteroatom rings (e.g., thiophenic ring) attached to it will be determined also via a PDF. After that, another PDF will determine whether any naphthenic rings will be added to the core.

After such a set of various ring attributes are determined, several distinct structures can share the same attributes (e.g. naphthalene and biphenyl benzene). Therefore, a further attribute for ring configuration is imposed to further identify those structures. An example of the determination of an aromatic core with four aromatic rings, one thiophenic ring and one naphthenic ring is showed in Figure 2.4. When the core structure is obtained, the full molecule's identity can be determined by those ring structural attributes, the attributes of side chain length/carbon number, and the number of side chains. Non-ring structures are treated as a “special case”, wherein only the carbon number and the number of side chains are considered. When the full identities of the molecules in a feedstock is determined, each molecule in this

feedstock can be built a unique quantitative relationship between the mole fraction and the corresponding attributes' values via juxtaposition as discussed in the later section.

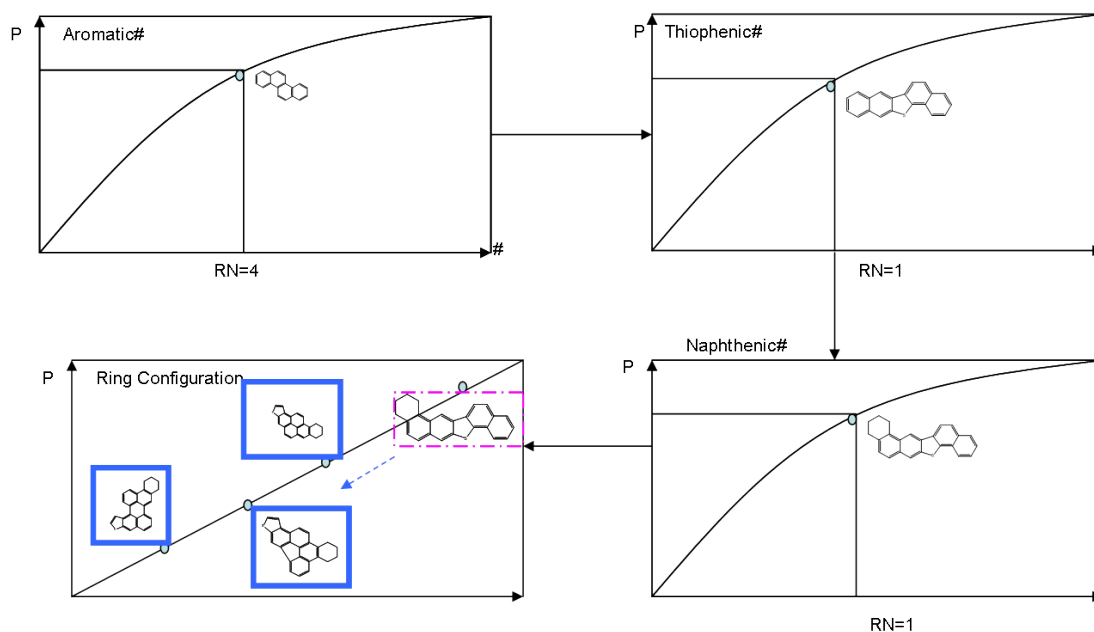


Figure 2.4 Determine Aromatic/Resin Cores (Series)

The sampling rule in Figure 2.3 is a generic protocol for petroleum hydrocarbon mixture. It can be considered as some special cases such as: naphtha oil, gas oil and resids.

Particularly, if the ring structural attributes in Figure 2.3 can be determined explicitly, this sampling logic for that petroleum hydrocarbon mixture can be turned into a set of homologous series with different side chain length and the number of side chain addressed before as shown in Figure 2.5.

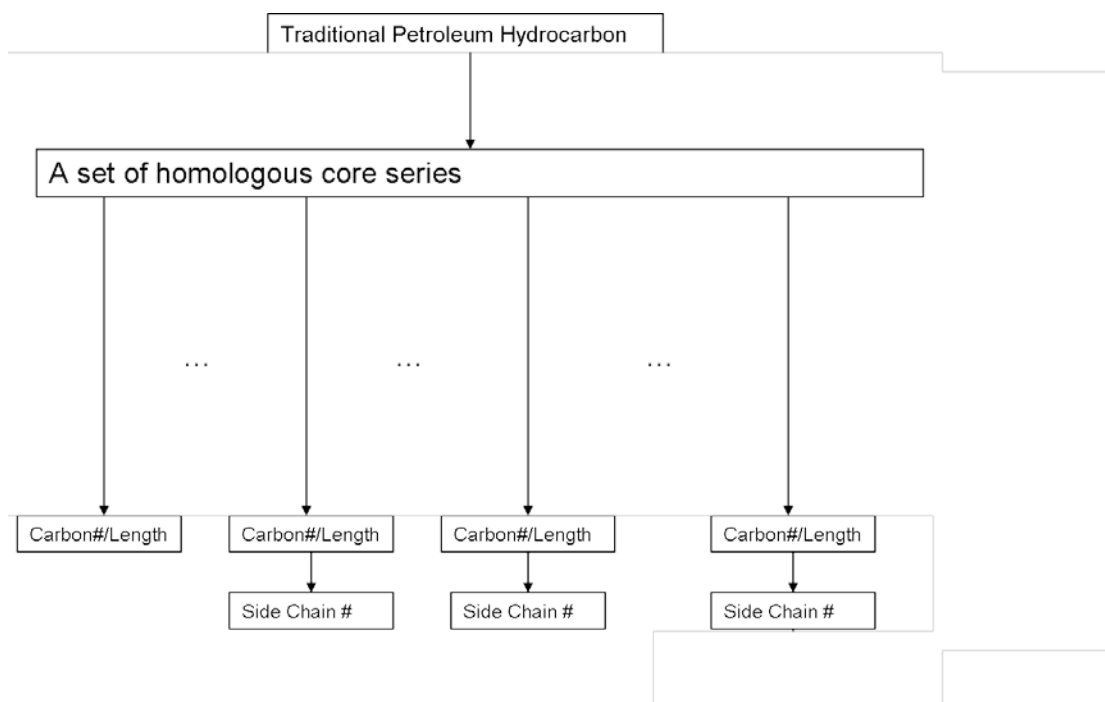


Figure 2.5 Determine Identities for light Oil to Gasoil (Expert determination)

A special case for this fixed identity determination is that a predefined set of molecules is given by an existing reaction network such as that generated by INGen. In this case, not only the homologous series but also the sampling size of the molecules is predetermined. Therefore, an automated identity recognition algorithm can be applied.

Although, the procedures outlined above can handle the creation of fixed identities for feedstocks ranging from light to extremely complex, there exists the possibility that the identities or sample rules simply can not be determined. Therefore, as a supplemental support, a method of stochastic sampling via Monte Carlo simulation can be applied. After having done such, the results of the Monte Carlo be added to modelers' experience and therefore employed as a fixed structure sets in future.

Although, the examples and discussions in above sections focus on petroleum feedstocks, the same methodology (with different attributes) can be applied to new feedstocks such as biomass.

Overall, this approach can determine the identity of most feedstocks in a straightforward manner. The computation involved in feedstock modeling will mainly focus on the optimization of quantitative information, as will be discussed later. Although this approach uses a two step method to determine the qualitative and quantitative information of a complex feedstock, it avoids the necessity for a computationally intense algorithm.

Although the methodology for the determination the identity of MCA has been demonstrated, it is necessary to show how a computer system can handle and identify those structures. Each molecule and core series group that can be regarded as a specific molecule in MCA will be represented digitally as Bond-Electron (BE) matrix. The BE matrix is not only the basis for the communication among the three aspects of this thesis as mentioned in Chapter 1, but also provides the key to determining the identity such as isomers and ring configurations.

2.4.1 Bond-Electron matrix representation of molecules

A molecule can be represented by a graph, the atoms being the nodes and the bonds being the edges. A more mathematically tractable implementation for a chemical species is through its bond-electron matrix, where the ij entries denote the bond order between connected atoms i and j . The BE matrix formation of every molecule/ is stored as an adjacency list in its own file. Each atom in the molecule is expressed in one row in the adjacency list. The adjacency list indicates the sequence numbers and provides the element name for each atom. In addition, it describes the bond connectivity by pairs of data in the parenthesis. For atom i in the i th row, each parenthesis represented a bond to atom i : the first value is the identification number of the bonded atom, the second value is the bond order of that bond. An example of pentane is shown in Figure 2.6¹.

The BE matrix and adjacency list represent the two dimensional (2D) structure of any given molecule in MCA. In addition, they are used to interact with INGen and structure property correlations as will be

mentioned later. Because a large identities for core series will need to be determined, it is necessary to provide an efficient way to construct the BE Matrix and the adjacent list for those core series.

Connectivity Matrix Adjacency List

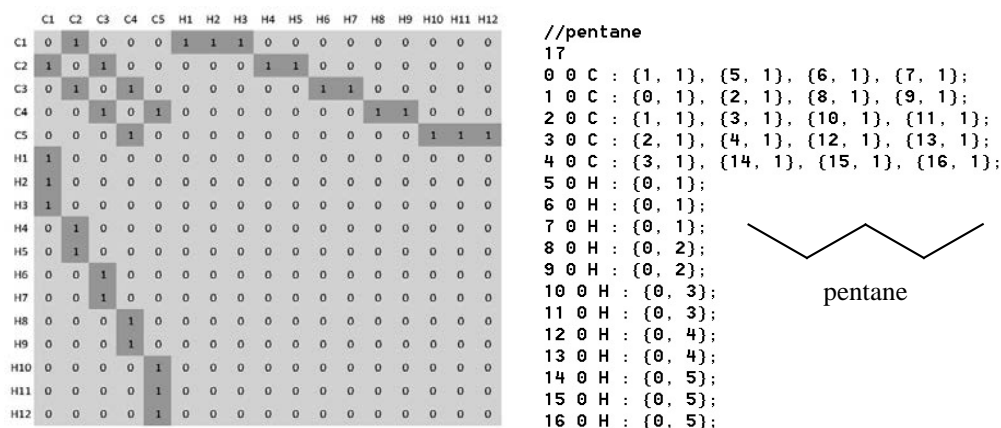


Figure 2.6 Bond-Electron Matrix for pentane¹

2.4.2 Core series' structure generator

To construct the BE Matrix and the adjacent list of core series, a core structure generator has been developed for traditional petroleum hydrocarbon mixtures. It is tremendously difficult to specify a BE Matrix for a complex core structure directly from a user inference. For example, the pentane shown in Figure 2.6 contained 17 atoms and thus a 17*17 matrix has to be specified for this simple structure. A complex molecule can contain hundreds of atoms and be much more complex than the case of pentane, so it is impractical to specify the elements of the BE matrix for the complex molecule. A simplified approach must be applied to make it practical.

Based on the accumulated experience of the field over the past few decades^{2,3}, a set of sub BE matrices that can be defined as “lumped” incremental vectors in order to describe the typical key structures occurring in hydrocarbon mixtures. The full BE matrix is then generated based on the combination of incremental vectors selected by the user. ExxonMobil's SOL^{2,3} provides the leading example for the use

of structural incremental vectors in defining a molecule's structure. A similar set of structural increments has been employed in the core structure generator described in this thesis. As shown in Figure 2.7, there are 15 structural increments designed for the description of ring structures and four descriptors for non-ring components for the case of petroleum hydrocarbon mixtures.

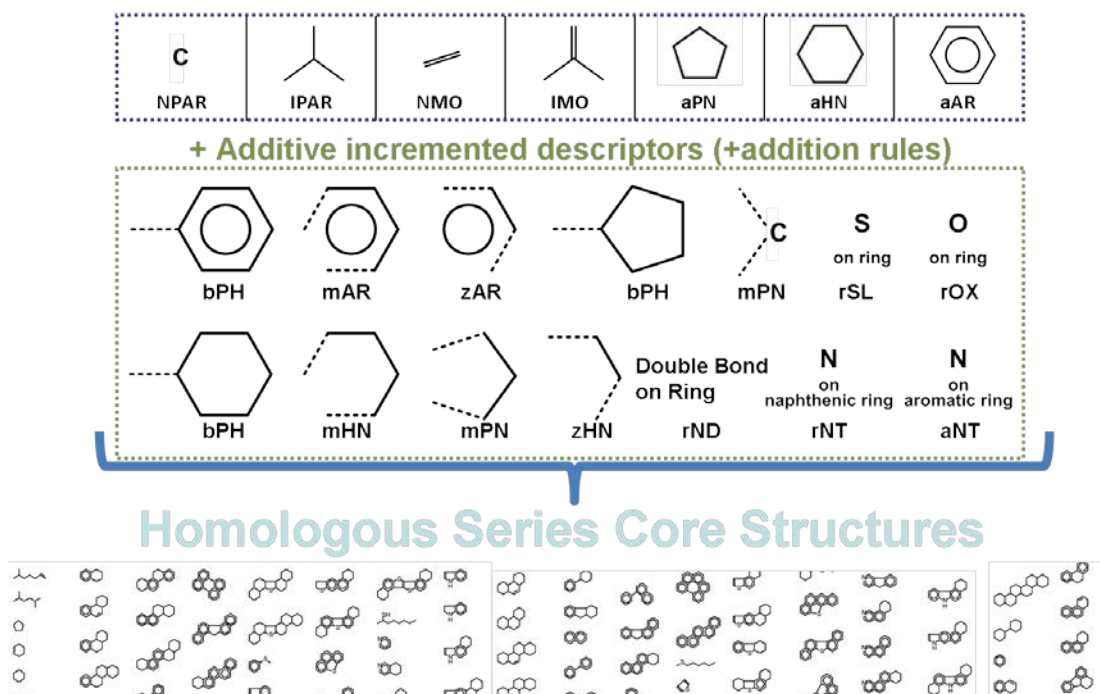


Figure 2.7 Series Core Generator

The meaning of each incremental descriptor is given as follows:

- NPAR and NMO are normal paraffin and normal mono-olefin increments used for non-ring structures.
- IPAR and IMO are iso paraffin and iso mono-olefin increments used for non-ring structures.

The other 15 descriptors are used describe ring structure:

- aAR is a single six carbon aromatic ring. The molecule defined by an aAR group that exists by itself is benzene. All multiple ring aromatic compounds must contain at least one aAR group; the other rings may be described by other incremental ring structures as given below.
- mAR is an incremental structure that must be attached to either an aAR or another mAR and cannot exist by itself.
- zAR is a two carbon aromatic increment that results from a pen-condensed multiple ring structure

such as pyrene. This increment can occur when another aromatic ring forms in the “bay region” of an angular multiple ring aromatic molecule such as phenanthrene.

- aHN and aPN are six and five carbon naphthenic ring increments. The aHN and aPN rings are also, like aAR, structural increments that can exist independently in this case as cyclohexane and cyclopentane respectively. All multiple ring naphthenic molecules contain at least one aHN or aPN, where further ring structure is defined by other incremental naphthenic ring structures.
- mHN, mPN, zHN, and zPN are the additional four, three, two, and one carbon naphthenic ring increments that like mAR and zAR above describe attached ring structures.
- bPH is a biphenyl bridge between any two nonincremental rings (aAR, aHN, or aPN).
- rSL, rNT, and rOX are respectively sulfur, nitrogen, and oxygen heteroatoms that exist as substitutes in a naphthenic ring.
- aNT is a nitrogen group substitution in an aromatic ring (as in pyridine or quinoline).
- rND specifies the number of double bonds within a naphthenic ring (as in cycloene or cyclodiene).

To simplify the structures of ring isomers, the following set of conventions is used to determine typical ring structures:

- All benzenoid multiple ring aromatics or their partially hydrogenated and naphthenic analogues assume a catacondensed arrangement of rings (e.g., phenanthrene, not anthracene, is represented by three condensed aromatic ring cores).
- The external rings of partially saturated aromatic compounds are saturated first.
- An mPN ring is always an external ring.
- Heteroatoms are placed in the smallest naphthenic ring.
- Double bonds are first associated with the heteroatom-bearing ring.
- In multiple biphenyl ring sets, condensed rings only appear on the end. Only single ring structure such as (aAR, aHN or aPN) can be attached in the middle. Aromatic rings have a higher priority than naphthenic rings for inclusion in the middle.

Through the configuration of these incremental descriptors via a user interface, the structure of any core and its BE matrix can be efficiently specified for hydrocarbon mixtures. Thus a set of homologous series' core structures can be conveniently determined as shown the latter half of Figure 2.7. The identities for the entire set of molecules defined by a set of homologous series can be determined by adding side chain information based on carbon and branch counts, thereby allowing the BE matrix of any molecule to be generated.

The side chain information for a specific homologous series assumes that there is only one side chain with carbon length larger than one. The remaining (if any) side chains each only contain a methyl group. Consequently, the identities for a set of representative molecules of a complex feedstock and their digital format-BE Matrix can finally be obtained finally the approach of fixed identity determination addressed in the previous section. The collection of all those identities is called the footprint of the feedstock.

The idea behind this core structure generation technique not only was to provide a quick and easy way to determine the footprint, but also to provide an integral path forward for the computer based sampling rules, for the Monte Carlo simulation of those implicit identities.

The core structure generator works well for general petroleum hydrocarbon mixtures, which have been the main source of energy and fuel for the last hundred years. However, its current incremental descriptors and rules may not cover new feedstocks (e.g., biomass, coal, shale). One solution would be to add new incremental descriptors as they are discovered via accumulated future research. Another flexible solution would be to draw out the core structures with third-party chemical software (e.g., CambridgeSoft, ChemDraw) that supports the Chemical Markup Language (CML). The integration of CML to BE matrix parsing into a toolset exists as a supplemental approach in this thesis.

Therefore, the qualitative determination of the footprint for traditional and new feedstocks has been successfully outlined. The next step is to determine the quantitative information for those complex feedstocks.

2.5 Quantitative determination of molecular compositions

The quantitative representation of MCA0 will be a combination of multiple attributes' values as juxtaposition shown in Figure 2.8.

This example molecule contains five attributes (the number of aromatic rings, the number of thiophenic rings attached on aromatic rings, the number of naphthenic rings attached on aromatic rings, the number of side chains and the carbon number/length of side chain) and neglects ring configuration. Each attribute is described by a continuous PDF function. The number on the X axis shown in parenthesis indicates the specific values of those attributes (e.g., $X_{51(1)}$ indicates a single aromatic ring). The value of the Y axis is the corresponding probability (P) of a specific attribute PDF of a given X. Through a juxtaposition of those attributes, the identity of a specific molecule is determined by various X of attributes and simultaneously a unique quantitative information relationship between those attributes' values and the mole fraction of the molecule by the joint probability of the probabilities corresponding to those given X values of the attributes.

Since the mole fraction is came from the combination of attributes' PDFs values, it is important to understand these attribute PDF functions.

Juxtaposition sampling

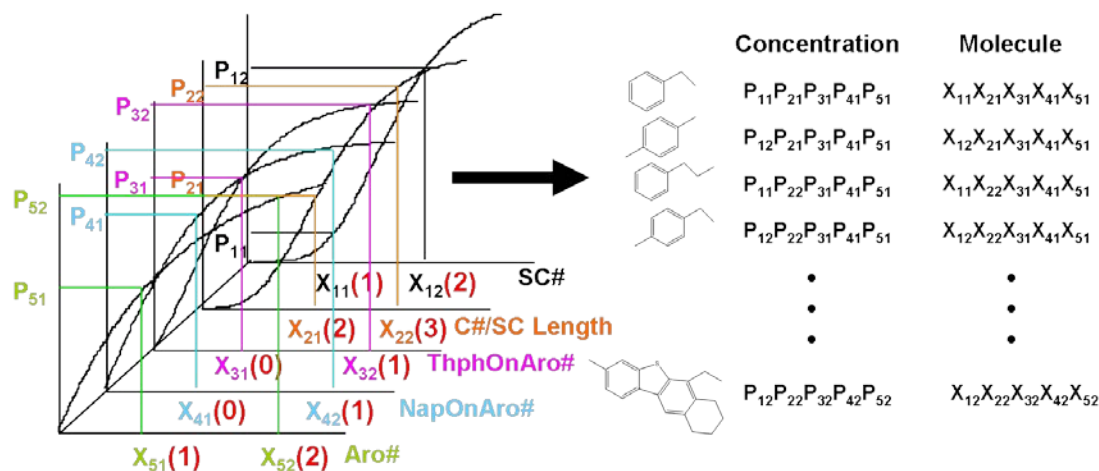


Figure 2.8 Multiple Attribute PDFs Sampling Defines Molecules

2.5.1 Probability density functions

In order to understand how to represent a feedstock statistically, it is necessary to first understand what a probability density function is. A PDF may either be discrete (integer values of x only), or continuous (any real value of x). A discrete PDF is defined by the following equations:

$$0 \leq f(x) \leq 1 \quad (2.1)$$

$$\sum f(x) = 1 \quad (2.2)$$

A is defined similarly, except that Eq. 2.3 is applicable instead of equation

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad (2.3)$$

The common functional forms of continuous PDFs are the normal distribution, the gamma distribution, and the exponential distribution as shown in Table 2.2. In addition, a PDF does not need to be defined with a statistical continuous distribution, but also can be discrete forms. The common discrete PDFs are discrete uniform distribution, the binomial distribution, and the Poisson distribution. For example, a

distribution of test scores is computed simply by counting the number of each score and dividing by the population of the class. Figure 2.9 shows the graphs of some PDFs mentioned above.

Table 2.2 Probability density functions used to model the structural attributes of a complex feedstock¹⁴.

	$p_i = f(x_i, \alpha, \beta, \gamma)$ $x_i = \text{attribute}$ $\alpha, \beta, \gamma = \text{pdf parameters}$	
Exponential:	$p_i = \frac{e^{\left(-\frac{x_i - \gamma}{\Theta}\right)}}{\Theta}$	2 parameters (γ, Θ) $\gamma < x_i < \square$ $\Theta = \mu - \gamma$ $\sigma = \mu - \gamma$ $\gamma = \text{minimum}$
Gamma:	$p_i = \frac{\left((x_i - \gamma)^{(\alpha-1)} * e^{\left(-\frac{(x_i - \gamma)}{\Theta}\right)} \right)}{\left(\Gamma(\alpha) * \Theta^\alpha \right)};$	3 parameters (γ, α, Θ) $\gamma < x_i < \square$ $\Theta = \sigma^2 / (\mu - \gamma)$ $\alpha = (\mu - \gamma)^2 / \sigma^2$ $\gamma = \text{minimum}$
Chi - Square:	$p_i = \frac{\left((x_i - \gamma)^{\left(\frac{r}{2} - 1\right)} * e^{\left(-\frac{(x_i - \gamma)}{2}\right)} \right)}{\left(\Gamma\left(\frac{r}{2}\right) * 2^{\frac{r}{2}} \right)};$	2 parameters (γ, r) $\gamma < x_i < \square$ $r = \mu - \gamma$ $\sigma = (2(\mu - \gamma))^{0.5}$ $\gamma = \text{minimum}$

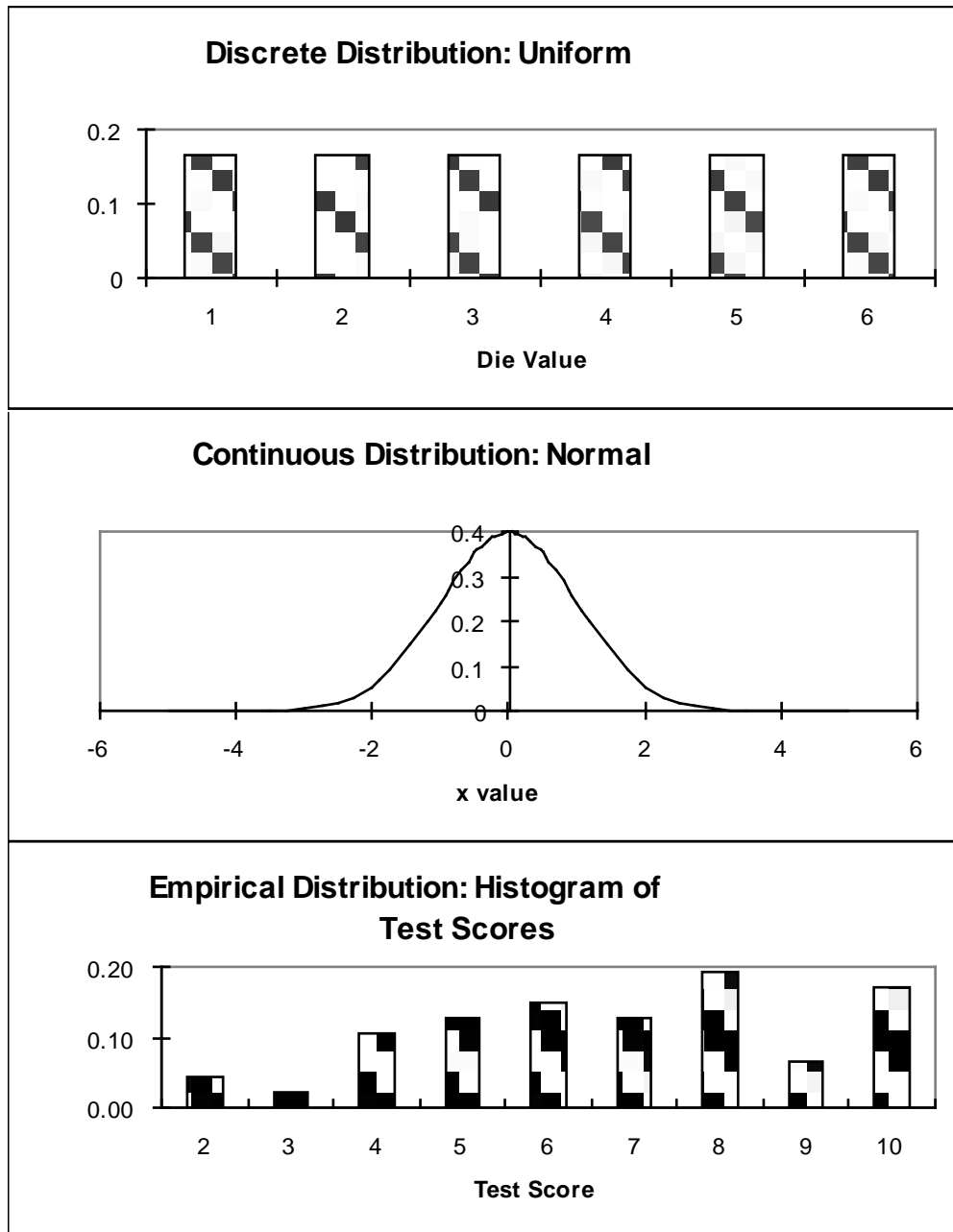


Figure 2.9 Examples of probability density functions¹⁴

Besides the understanding of the rigorous definition of a PDF, there are several other important practical aspects need to be considered when using them to define a complex feedstock. The scientific relevant is very important as the historical proofs to use these attribute PDFs for defining a complex feedstock by the statistical approach mentioned above. How to select an appropriate functional form of those PDFs

for modeling the feedstock is quite important too. Finally, some technical issues such as PDF discretion, truncation, and conditional probability must be considered for the accurate representation of a complex feedstock.

2.5.2 Probability density functions to describe complex mixtures

It is a quite long scientific history to describe complex mixtures by using PDFs. It is started from the research on polymer chemistry. The products of the formation and degradation of polymers have both been modeled with statistical distributions. Flory used a modified gamma distribution to fit the molecular size distribution of condensation polymers in 1936⁴. Similarly, Libanati studied the degradation of an infinite polymer⁵ by using a log normal distribution to describe the molecular weights.

Similarly, this logic was extended to the petroleum field later, since petroleum crude can be viewed as a breakdown of the infinite polymer kerogen. Consequently, petroleum reservoirs were studied by this idea. Whitson used a gamma distribution to fit the molar and weight distribution of the C7+ fraction of crude oil⁶. His results showed the shape of a gamma distribution is quite similar to the shape of a log normal distribution, which gave a support to represent crude components with PDFs. In addition, Whitson had extended the distribution beyond experimental measurements that demonstrated higher accuracy predictions in the equation of state via fitting the molar distribution to a PDF. Similarly, Shibata et al. used mixed distributions to enhance phase equilibrium calculations for a petroleum reservoir⁷.

In addition to above modeling works, a direct experimental confirmation showed that those statistical PDFs can be used to model petroleum. Peterson et al. used high-temperature gas chromatography to measure the weight percent distribution of carbon number up to C80+ for 17 North Sea oils⁸. An exponential distribution fit up to C20 showed that it could be used to predict the amount of heavier

components with high accuracy. In this model, Peterson used normal paraffin standards for calibration. Boduszynski⁹ demonstrated that boiling points with increasing carbon number have the wide divergence for different compound classes. From Boduszynski's work, it showed that the boiling point distribution may be followed by an exponential distribution after C20.

Trauth extended the use of PDFs to model not only the molecular weights or boiling points, but to the structural attributes described in the previous section^{10,11}. A statistical modeling project about the depolymerization of coal¹² which dealt with the thermal pyrolysis gave the experimental proof for the validation of Trauth's statement. This result showed that a gamma distribution can accurately fit the molecular weight distribution of the products as with the degradation of an infinite polymer. Because products are formed primarily by bond fission reactions during pyrolysis process, this result indicated that the individual structural attributes also would be well represented by gamma distributions. In addition, Trauth also demonstrated that using a gamma distribution for each of the structural attributes of a petroleum resid yielded a molecular weight distribution that was also gamma¹³. Campbell¹⁴ and Hou¹⁵ extended a gamma distribution to whole petroleum fractions.

2.5.3 Determining appropriate functional forms

As discussed in the previous section about the introduction of PDFs, there are many different functional forms for PDFs. It is very important for an optimal representation of a feedstock to select the appropriate forms of these PDFs. The most important factor to be considered is to capture the shape of the object to be modeled by the PDF qualitatively. In addition, it is often desirable to select the flexible distributions that there are only slight deviations to model a particular functional form accurately. The number of parameters for optimization is another considerable factor, which has a great effect on the computational burden especially for modeling a complex feedstock with many PDFs.

The molecular weight or boiling point distributions generally follow a smooth curve for petroleum fractions. Figure 2.10 shows the relative boiling point distributions of a petroleum kerosene and a petroleum resid. Lighter fractions such as the kerosene are constrained by both a minimum and a maximum boiling point and generally show a normal or skewed normal distribution of boiling point. However, petroleum resid can only be constricted by a minimum boiling point. As shown in Figure 2.10, a boiling point distribution of petroleum resid generally shows a rapid rise followed by a slow decrease that mathematically can be expressed by a gamma type distribution. Gamma distributions or exponential distributions can accurately model such both polymers and heavy components of fossil fuels, which is investigated from past research works.

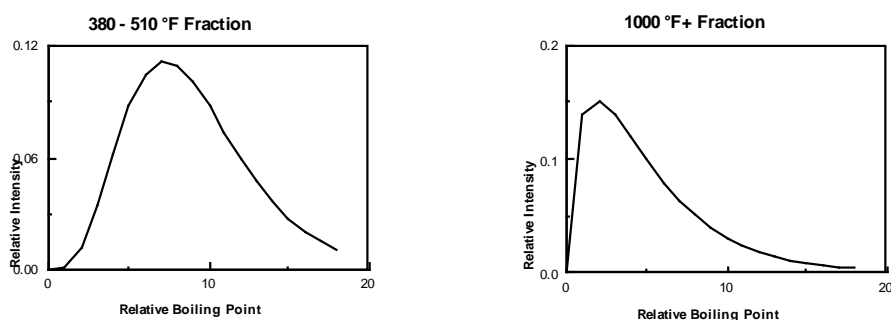


Figure 2.10 Relative boiling point intensity for kerosene and vacuum resid petroleum fractions¹⁴

The boiling point distribution is tightly linked to the structure of the molecules in a feedstock. Generally, boiling point distributions are very closely correlated to the molecular weight or carbon number of the molecules in a feedstock. Moreover, as shown in the previous section, the molecular attributes are implicitly related to the carbon number. Therefore the attribute PDFs can use the same type of functional forms that are well fitted with boiling point.

The above semi-theoretical arguments have been supported by empirical experience. Trauth⁸ showed good results of modeling a petroleum resid with a set of experimentally determined analytical properties by using the structural attributes with gamma and gamma-like distributions. What's more, this flexibility of the gamma distribution made it can express from an exponential distribution to a delta

function, and can also approximate a normal distribution, which also allows for the modeling of lighter feedstocks, even if the boiling point distributions for such feedstocks may not be considered to be typically gamma.

A final factor to be considered when selecting a functional form is the number of parameters for this form. It is true that the gamma distribution shows the advantage in flexibility, but it requires three parameters. Besides of gamma distribution, other functional forms as shown in Table 2.2 with fewer parameters are also tested to model a series of petroleum resids. The comparison results of gamma, chi-square and exponential forms are listed in Figure 2.11. The chi-square distribution is a special case of the gamma distribution where the standard deviation equals one half of the mean. As mentioned previously, the gamma distribution can match the exponential distribution for certain values of parameters. The priority of a chi-square or exponential distribution is that they have one fewer parameter than gamma to be optimized. This can be quite important because it can reduce computational burden of later optimization. The choice of the PDF form needs to balance the flexibility (gamma) and the simplification of calculation (chi-square/exponential) for a real system. This choice will be determined during exploring real applications.

Besides the above continuous functional forms, there are two discrete PDFs that can be chosen for special attributes. One is the uniform, or equal-molar, distribution shown in Figure 2.9. This simple form is used for the attributes that can not be determined from analytical measurements (e.g., isomers of a paraffin with the same carbon number and the same number of side chains but different branch positions). The other discrete form is that of a histogram distribution, or discrete values, shown in Figure 2.9. This form can be used for an attribute with a limited number of values or one that is incompatible with a continuous function (e.g., global fractions of PONA/SARA for overall molecule types).

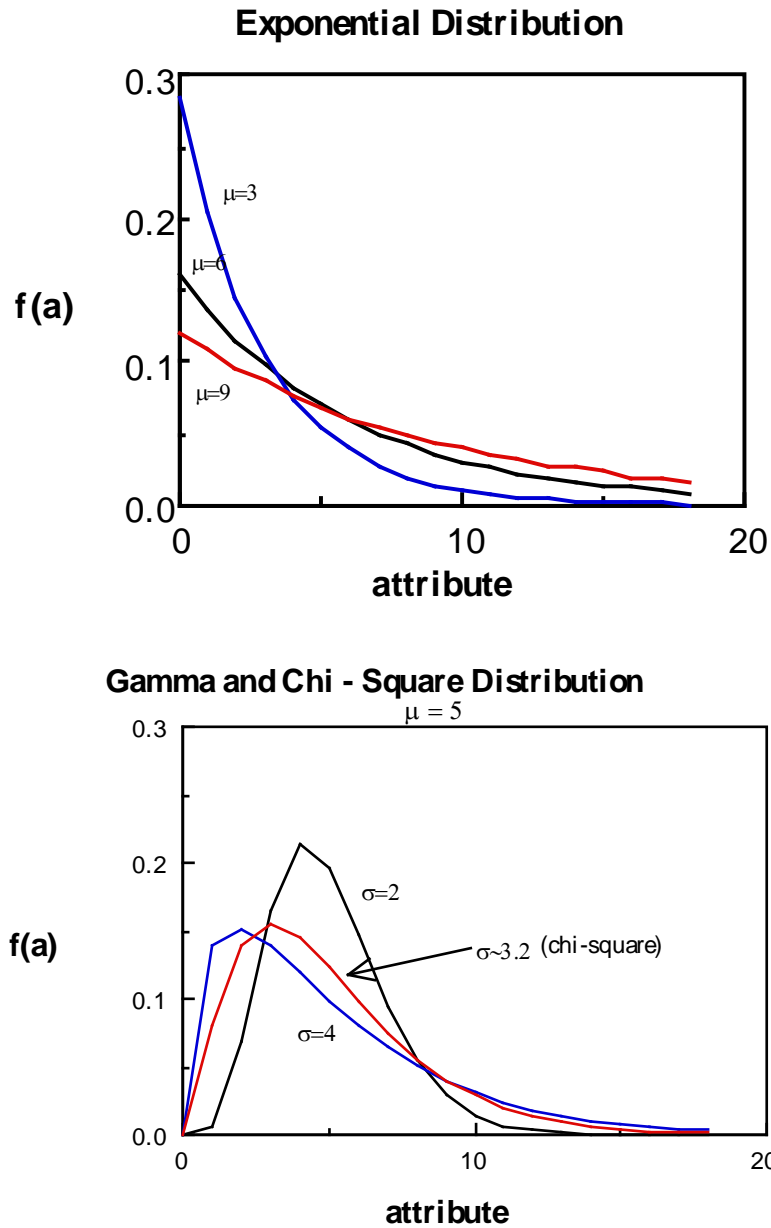


Figure 2.11 Examples of exponential, gamma and chi-square distributions¹⁴

2.5.4 Discretization of the probability density functions

Although PDFs such as the gamma distribution and the exponential distribution can be used to model complex feedstocks accurately, both of these distributions are continuous. However, real feedstocks are composed of attributes with discrete integer values. Therefore, it is necessary to transform these continuous distributions into discrete distributions.

In order to transform a continuous distribution into a discrete distribution, it is necessary to divide the distribution into intervals and evaluate a representative value for each interval. If the length of the interval is sufficiently small, then the midpoint represents the average value of the interval. If not, then the transformation results in a poor approximation of the continuous distribution. Trauth showed that the shape of a continuous gamma curve was well maintained using such a discretization¹⁰.

2.5.5 Renormalization and truncation

Just as the attributes of the molecules in a complex feedstock are all discrete integer values, they are also finite. Therefore, it is necessary to truncate the distributions at some physically reasonable value. Once a distribution has been truncated, it is also necessary to renormalize the distribution so that the probabilities add up to one.

A truncation criterion may be set by specifying the minimum contribution each new interval must make to the cumulative distribution expressed on a fractional basis. Trauth found that a value of 0.01 worked well¹⁰. Thus, if the value of the current interval divided by the sum of the values of all previous intervals is less than 0.01, the distribution is truncated and renormalized. Eq. 2.4 mathematically shows the truncation criterion:

$$\frac{P_i}{P_i + \sum_{j=1}^{i-1} P_j} < 0.01 \quad (2.4)$$

2.5.6 Conditional probability

Conditional probability is the last important consideration on constraining structural attribute PDFs to describe the MCA of a complex feedstock. Conditional probability is the constraints from physical properties of a feedstock (e.g., boiling point cut, carbon number or molecular weight). Through these constraints, the value of one attribute might follow the modification of another attribute PDF. In other

words, once an attribute value has been determined, the probability of another attribute having a particular value may be altered.

For example, all molecules in a petroleum resid that is physical defined only by a minimum boiling point must have the boiling point above that minimum boiling point, which leads some conditional probability. The three aromatic series requires the longer length of side chain than the four aromatic series to satisfy a given minimum boiling point cut such as 1000°F. Therefore, the PDF for the length of sidechains must be different corresponding with the two different attribute values of number of aromatic rings.

The implementation of conditional probability with physical constraints is very important in obtaining a good molecular representation. In addition to initial boiling point criteria, it may be necessary to apply conditional probability to limit the size of a molecule or to limit the number of heteroatoms in a molecule.

2.5.7 Attribute PDF selection

The selection of attributes can be classified by the complexity of the feedstocks. For simpler feeds such as naphtha oil, the attributes can be described as a set of discrete molecules that can be directly attained from analytical measurements. For the mid-heavy end of a complex feedstock such as gasoil, the attributes can be comprised of a set of homologous core series with different sidechains and substituents. The global probability of these series can be represented as a set of discrete values. The side chains and substituents of each series will be represented as a set of continuous attribute PDFs. For the heavy end of a complex feedstock such as resid, discrete attributes would contain a large number of core series. Such a selection would carry a huge computational burden. Instead, the core series can be decomposed into a combination of continuous PDFs for the elemental structural attributes by using the general sampling rules discussed in the previous section. The side chains and substituents of each series would then also be represented as a set of continuous attribute PDFs. This scheme of cascading PDFs

for heavy-end oil can also be extended for the whole spectrum of petroleum hydrocarbons from light to heavy.

For the identities as sampled stochastically by Monte Carlo simulation, the quantitative information will be byproduct of the sampling calculation. Examples of attribute PDFs selections for petroleum hydrocarbon conversion are listed in

Table 2.1 .

Once the PDFs have been selected, the next step is to finally obtain the values of MCA0.

2.5.8 MCA0 generation

If MCA0 is quantitatively predetermined, it can be used directly for either deterministic modeling or parsed into the structural attributes used for Attribute Reaction Modeling (ARM) as described in the next section. This attribute parsing follows by a straightforward mathematic method as shown in Eq. 2.5:

$$P_{i,j}(x_j) = C_{i,j}(x_j) / \sum C_{i,j} \quad (2.5)$$

Where $P_{i,j}(x_j)$ is the probability of attribute value x_j for attribute type i and $C_{i,j}(x_j)$ is the concentration (mol/L) of the molecule containing attribute i with value j (e.g., length of side chain is the attribute type i , and j is the exact value of certain molecule's length of side chain).

However, the full quantitative molecular compositions of most complex mixtures can not be acquired by analytical chemistry directly. Therefore, an optimization loop is employed to find an optimal MCA0. As discussed previously, the quantitative molecular information can be found by a combination of discrete values and continuous attribute PDFs. Below a chi-square objective function encompassing various analytical terms (Eq. 2.6) can be minimized in order to determine the optimal discrete values

and attribute PDF parameters. Thus, the optimal attribute values are determined, and the optimal MCA0 is set.

$$\begin{aligned} \chi^2 = & \left(\frac{MW_{\text{exp}} - MW_{\text{pred}}}{W_{MW_{\text{exp}}}} \right)^2 + \left(\frac{HtoC_{\text{exp}} - HtoC_{\text{pred}}}{W_{HtoC_{\text{exp}}}} \right)^2 + \left(\frac{Harom_{\text{exp}} - Harom_{\text{pred}}}{W_{Harom}} \right)^2 + \\ & \left(\frac{Halpha_{\text{exp}} - Halpha_{\text{pred}}}{W_{Halpha}} \right)^2 + \left(\frac{Nitrogen_{\text{exp}} - Nitrogen_{\text{pred}}}{W_{Nitrogen_{\text{exp}}}} \right)^2 + \left(\frac{Sulf_{\text{exp}} - Sulf_{\text{pred}}}{W_{Sulf_{\text{exp}}}} \right)^2 + \quad (2.6) \\ & \frac{1}{\#Frac} \sum_{i=1}^{\#Frac} \left(\frac{SimDis_{\text{exp}} - SimDis_{\text{pred}}}{W_{SimDis_{\text{exp}}}} \right)^2 + \frac{1}{\#Comps} \sum_{i=1}^{\#Comps} \left(\frac{SARA_{\text{exp}} - SARA_{\text{pred}}}{W_{SARA}} \right)^2 \end{aligned}$$

The analytical items in Eq. 2.6 could be different for various feedstocks (e.g., PONA for middle-distillates vs. SARA for resids). Therefore, the optimization parameters and computational burden can be different based on various feedstocks. The following discussion analyzes the previously mentioned two cases of attribute PDF selection: 1. a set of discrete values for the homologous core series and continuous attribute PDFs for the side chain information; 2. a set of continuous elemental structural attributes for both the core series and the side chain information.

For case 1, there are #CoreSeries -1 parameters for the global core series fractions and 2*#CoreSeries for the side chain carbon number PDFs of each series if each PDF is defined as gamma). In addition, there are also 4 parameters for the number of side chains given that there are separate gamma PDFs for ring and non ring attributes. Therefore (#CoreSeries -1 +2* #CoreSeries +4) parameters are necessary in total. Light-middle oil fractions often have O (10~20) series, and therefore O (30) parameters, which is a practical number for optimization. Using discrete values for the homologous series provides more flexibility in the discrimination of isomer details than the constraint of continuous attribute PDF function. In addition, it is possible that direct measurements of light feedstock can provide help in simplifying the optimization. However, this method is not applicable for heavy end oil fractions, as they may contain large number of core series. Therefore a given high end fraction can contain over O (100-200) parameters and thus not of practical computational burden.

For case 2, the core series is derived from several continuous structural attributes (except the fraction of overall PONA/SARA that indicates the molecules' types). If gamma PDFs are selected, the maximum number of parameters for core series is O (10~20) (depending on how many core series there are). The number of parameters for side chain information is the same as described in case 1. However, this method can use a side chain attribute PDF "family" concept that will lump together PDFs for certain core series groups (e.g., all cores with the same aromatic ring number). From a statistical view, this concept is applicable for middle-heavy oil. The number of the parameters for side chain information can be constrained to O (10). As a result, the total parameters for case 2 can be limited to O (30), which is a practical number for optimization. This method provides a feasible solution for complex heavy mixtures in which the effect of isomer details is trivial.

For the middle-heavy fractions, both of these two cases may be applicable. Chapter 5 will give further the discussion as to which is better while providing validation to both approaches.

2.6 Automation of composition modeling-CME

Following the methodology laid out in this chapter, a user-friendly program with an excel interface called the Composition Model Editor (CME) has been developed for qualitatively and quantitatively obtaining MCA0. CME consists of four functional executable modules: CoreGen, FootGen, CompGen and PropGen. CoreGen is the core series structure generator. FootGen generates the entire qualitative footprint of the complex feedstock from one of three separate aspects: a predefined set of molecules from INGen¹, Monte Carlo simulation from MolGen^{13,14}, or a homologous core series defined by CoreGen. CompGen is the optimization package that determines the quantitative information of the feedstock and thus obtains an optimal MCA0. PropGen discussed later in Chapter 4 provides property calculations to CompGen. The overall logic of CME is shown as Figure 2.12.

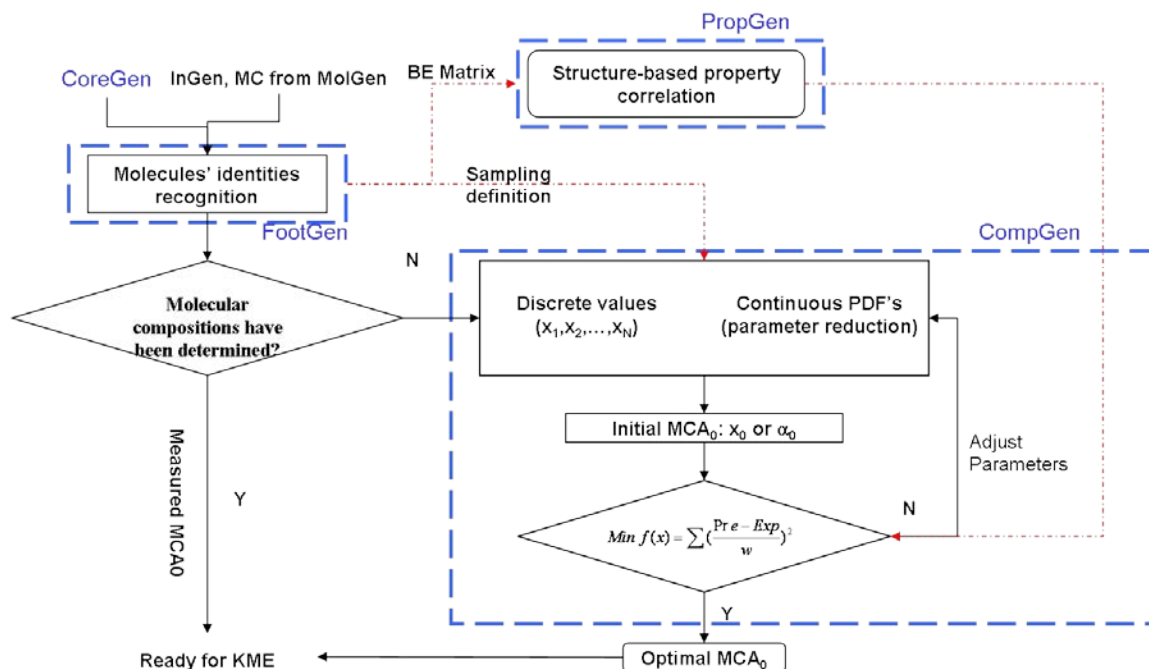


Figure 2.12 Overall logic of CME for composition modeling

Upon launch, CME can be run using any of the three footprint generation modes: a predefined set of molecules from INGen, a stochastic sample from MolGen (only used for cases containing uncertain identities), and generation from CoreGen (provided that a useful set of homologous core series can be inferred). For example, if CorGen made is used to create a VGO model, a set of homologous core series can be entered by the excel interface shown in Figure 2.13. By specifying the core series' structural increments, CME will generate the BE Matrix and adjacency list for each core series structure. In addition, a visualization of the structure is generated in order to provide an intuitive way to view each series.

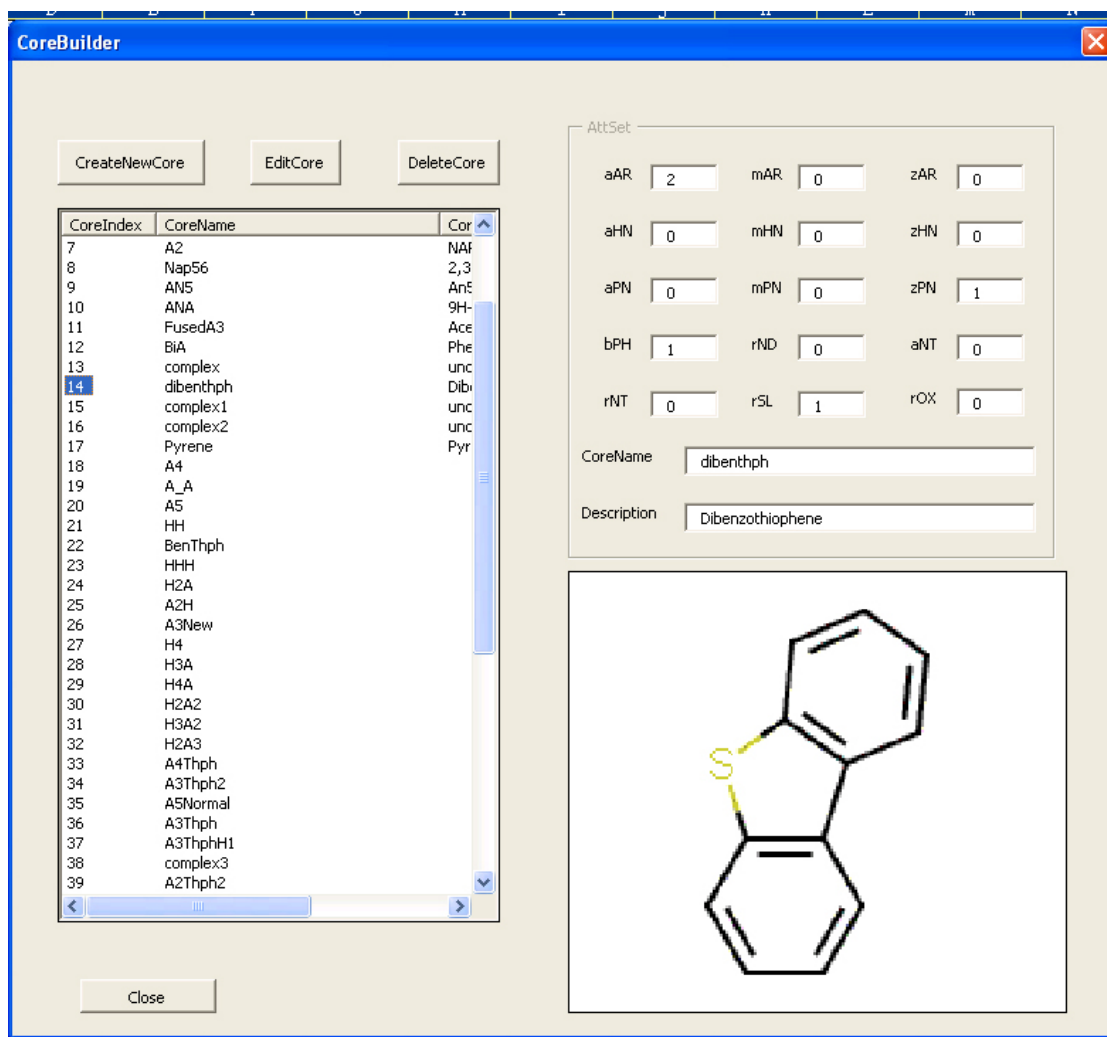


Figure 2.13 automation of Core generator in CME

The next step is the selection of the fixed set of core series for this model (Figure 2.14). In addition, to simple selection, the maximum number of side chains for each core series must be specified.

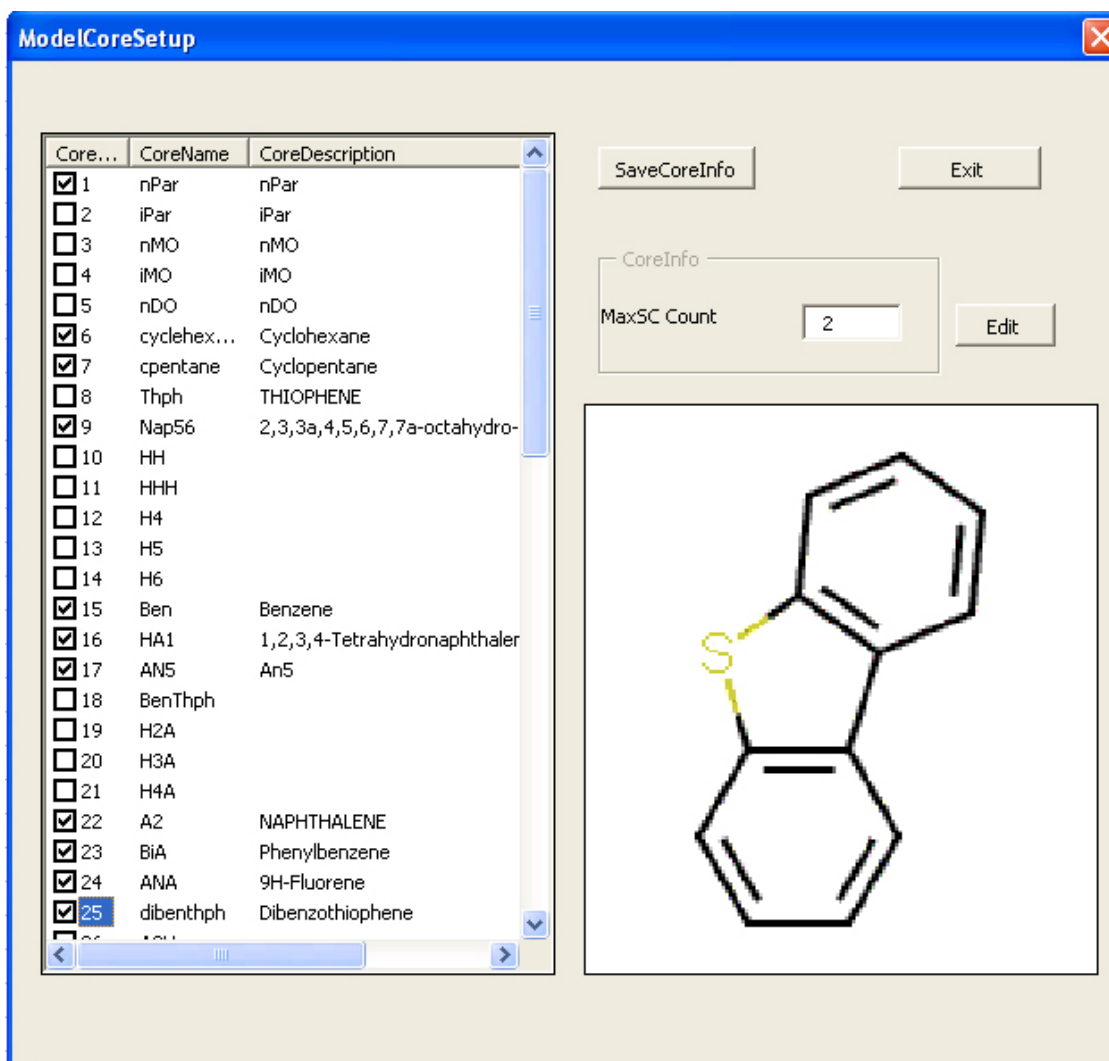


Figure 2.14 Identity Setup in CME

After core series selection, CME will generate the entire representation for this feedstock as shown in Figure 2.15. Conditional probability based on the boiling point range is then specified in order to constrict the side chain carbon number range for each core series.

CME will then launch FootGen to automatically recognize the structures generated above. Thus, a unique quantitative mapping relationship between attributes and mole fractions can be described for the whole footprint and is illustrated in Figure 2.16.

Frame1

GetCoreFromExpert

MaxBPCut(K) 826

MinBPCut(K) 540

GetCorefromInGen

Change

Automated Identity recognition and implemented sampling definition

Conditional constraint by boiling point

CoreIndex	CoreName	CoreDe
0	nPar	nPar
1	aAR0_mAR0_zAR0_aHNO_mHNO_zH...	Cyclohex
2	aAR0_mAR0_zAR0_aHNO_mHNO_zH...	Cyclohex
3	aAR0_mAR0_zAR0_aHNO_mHNO_zH...	2,3,3a
4	aAR1_mAR0_zAR0_aHNO_mHNO_zH...	BENZENE
5	aAR1_mAR0_zAR0_aHNO_mHNO_zH...	1,2,3,4
6	aAR1_mAR0_zAR0_aHNO_mHNO_zH...	2,3-Dih
7	aAR1_mAR1_zAR0_aHNO_mHNO_zH...	NAPHT
8	aAR2_mAR0_zAR0_aHNO_mHNO_zH...	Phenyl
9	aAR2_mAR0_zAR0_aHNO_mHNO_zH...	9H-Flu
10	aAR2_mAR0_zAR0_aHNO_mHNO_zH...	Dibenz
11	aAR1_mAR2_zAR0_aHNO_mHNO_zH...	Phenar
12	aAR1_mAR1_zAR0_aHNO_mHNO_zH...	Acenap
13	aAR1_mAR2_zAR1_aHNO_mHNO_zH...	PYRENE
14	aAR1_mAR3_zAR0_aHNO_mHNO_zH...	Chryse
15	aAR2_mAR2_zAR0_aHNO_mHNO_zH...	Peryler
16	aAR1_mAR3_zAR1_aHNO_mHNO_zH...	Benzo[

Selected series structure in model

OK

CoreName aAR1_mAR0_zAR0_aHNO_mHNO_zHNO
aPNO mPNO zPNO bPHO aNTO rNTO rOXO

CoreDescription BENZENE

Edit




Figure 2.15 Identity and Sampling Setup in CME

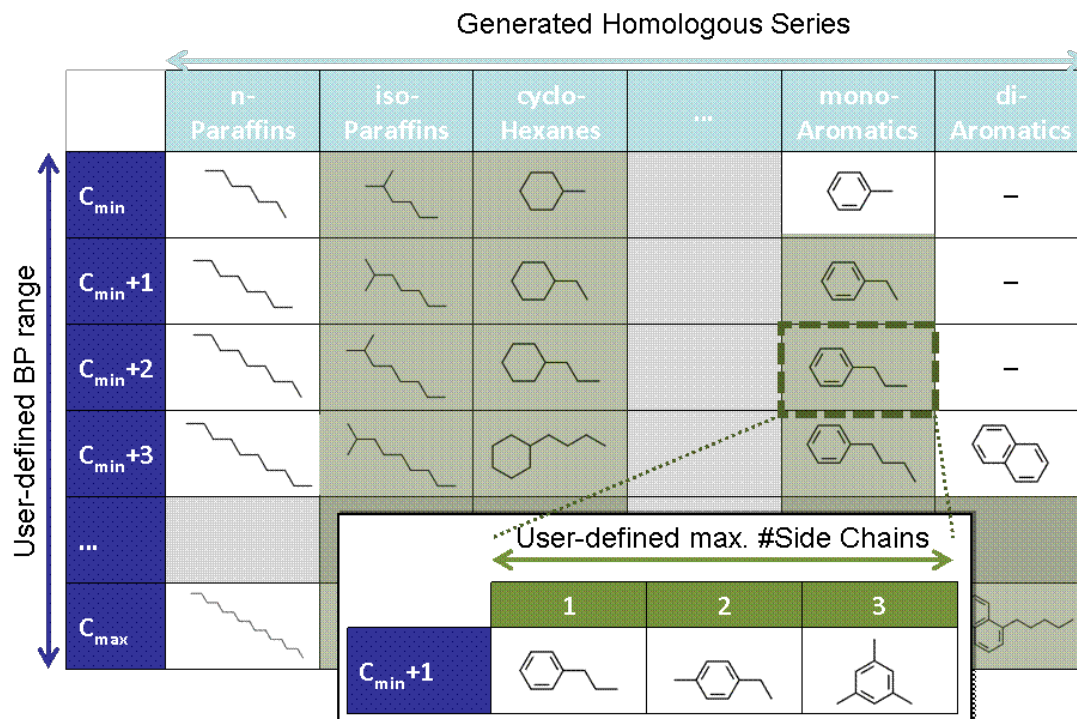


Figure 2.16 an example representation of a feedstock's footprint

After the footprint is set up, CME can run in either case mentioned in the previous section. An example of this choice is shown in Figure 2.17. If case 2 is selected, CME will continue to the configuration screen for the attribute PDF type (Figure 2.18), and the PDF family definition screen shown in Figure 2.19 (if necessary).

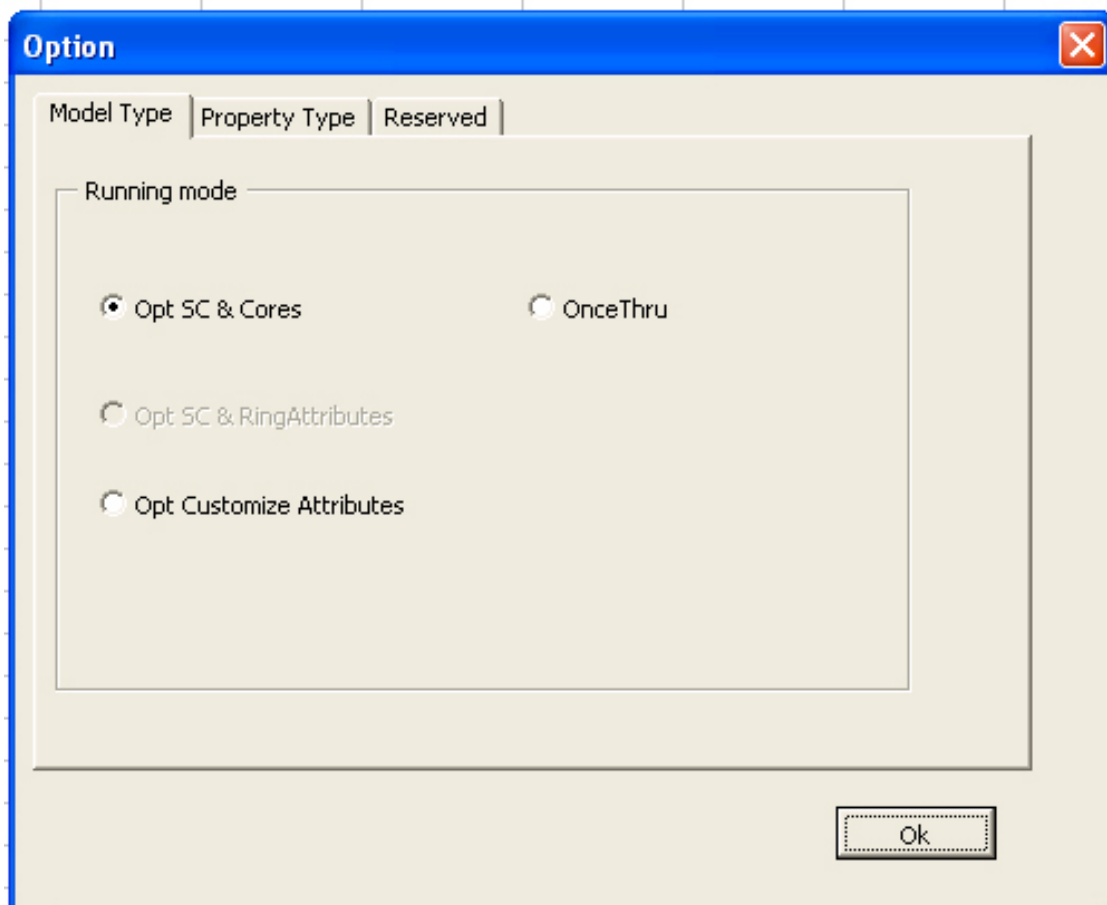


Figure 2.17 Selection cases for MCA0 generation in CME

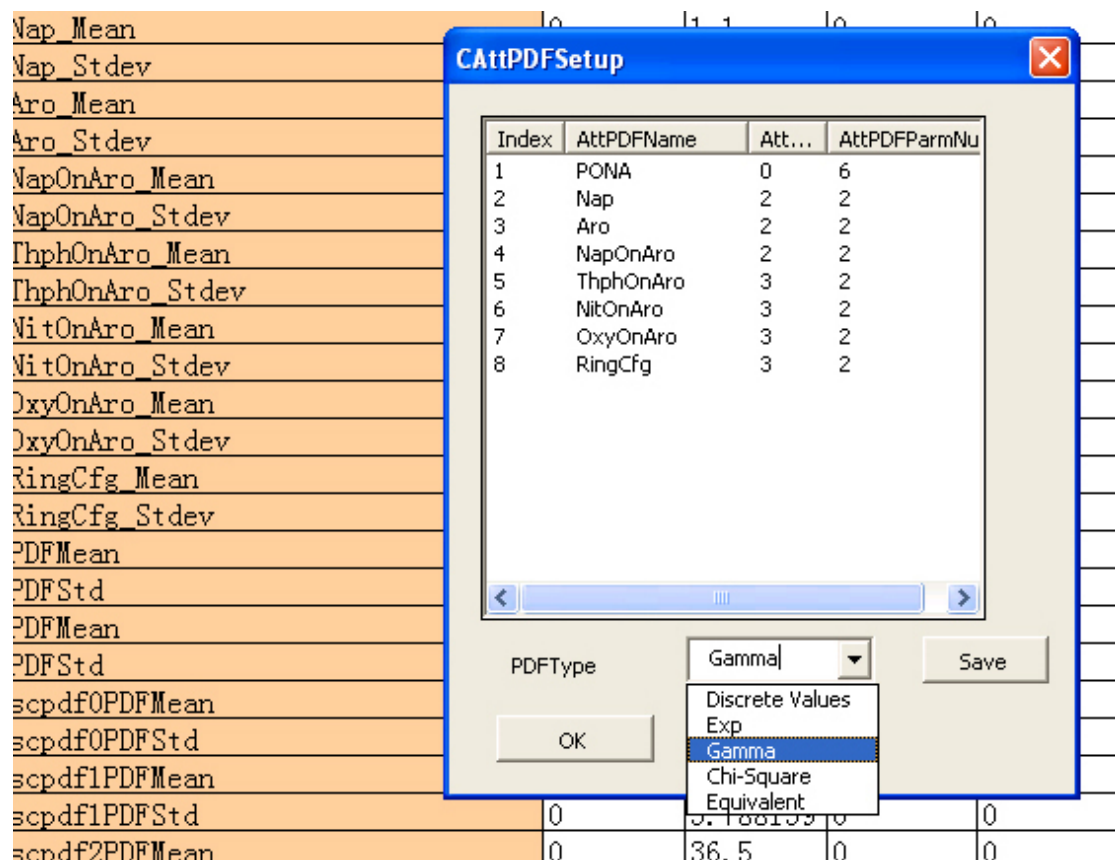


Figure 2.18 Attribute PDF Type Selection

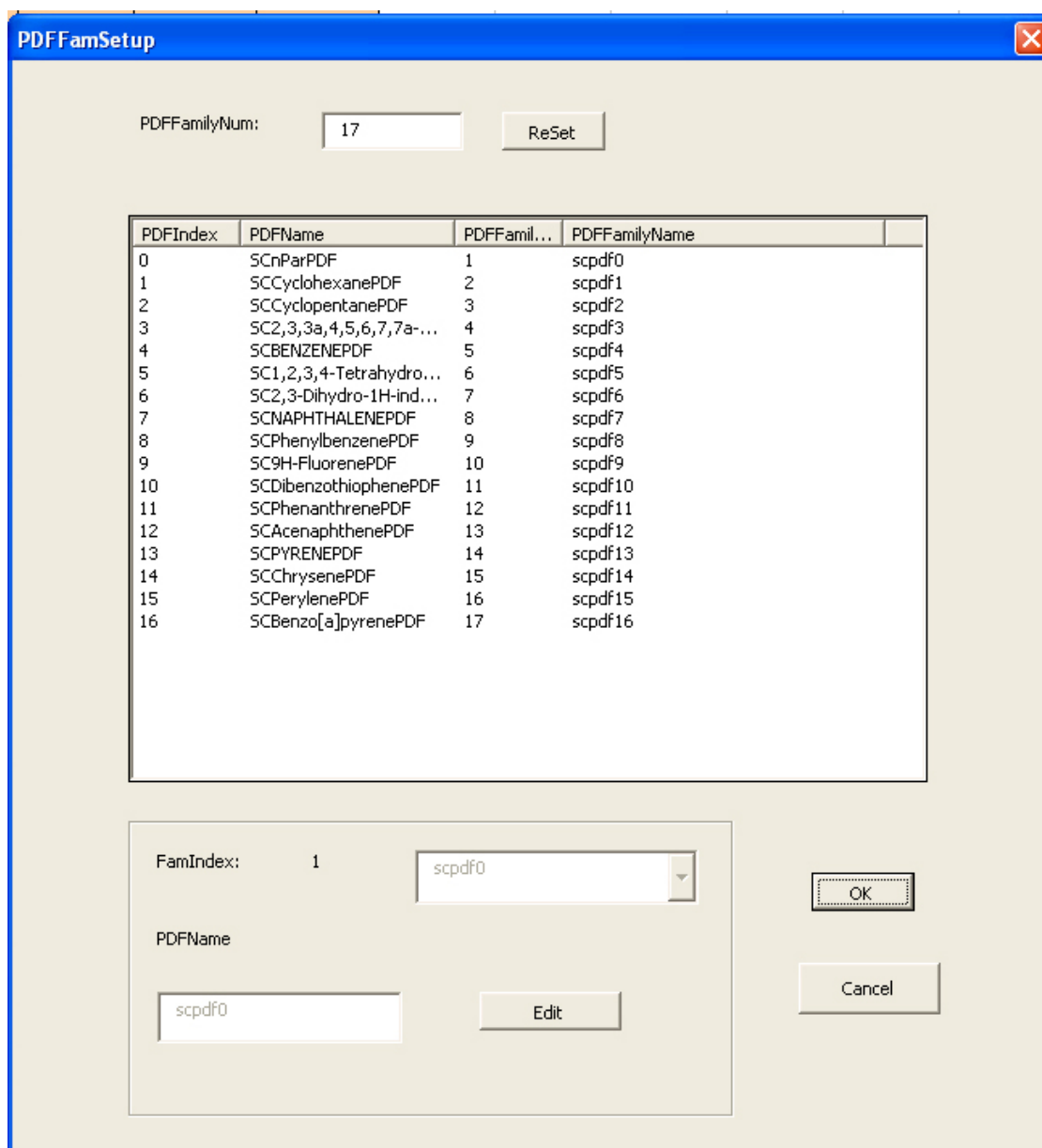


Figure 2.19 Grouping SC PDFs

After the PDFs have been set up, CME will use PropGen to integrate the properties of each molecule into the optimization objective function. Figure 2.20 shows the selection of the objective function, and thus the relevant properties, for either Light to VGO or HGO to Resids.

Finally, CME will launch CompGen to compute the optimization and obtain the optimal MCA0. The MCA0 shown in Figure 2.21 and Figure 2.22 can then be integrated with KME for kinetic modeling (as

discussed in later chapters). Figure 2.21 and Figure 2.22 also demonstrate the visualization and property analysis obtained by the structure-property correlations (as addressed in chapter 4) and therefore provides window into the species nature of MCA0.

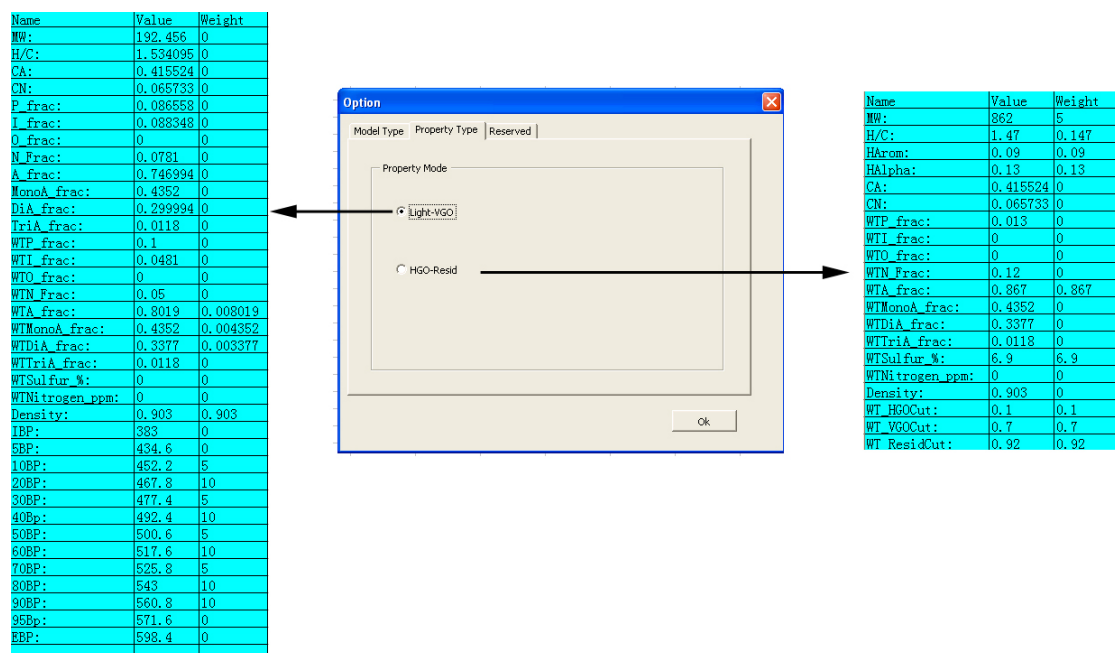


Figure 2.20 Property Selection in CME

	A	B	C	D	E	F	G	H	I	J	K	L	M
115	species113	0											
116	species114	0											
117	species115	0											
118	species116	3.97605E-38											
119	species117	0											
120	species118	3.97605E-38											
121	species119	3.97605E-38											
122	species120	2.31979E-37											
123	species121	3.97605E-38											
124	species122	5.40198E-39											
125	species123	7.65635E-38											
126	species124	2.31979E-37											
127	species125	2.27367E-37											
128	species126	2.31979E-37											
129	species127	7.65635E-38											
130	species128	2.31979E-37											
131	species129	3.89701E-38											
132	species130	5.52332E-37											
133	species131	3.85143E-36											
134	species132	3.85143E-36											
135	species133	5.52332E-37											
136	species134	3.22253E-36											
137	species135	6.53955E-25											
138	species136	3.22253E-36											
139	species137	6.53955E-25											
140	species138	3.85143E-36											
141	species139	6.53955E-25											
142	species140	3.81544E-24											
143	species141	6.53955E-25											
144	species142	8.88482E-26											
145	species143	5.74535E-25											
146	species144	3.81544E-24											
147	species145	3.73959E-24											
148	species146	3.81544E-24											
149	species147	5.74535E-25											
150	species148	3.85143E-36											
151	species149	3.81544E-24											
152	species150	6.40954E-25											
153	species151	4.14472E-24											
154	species152	2.89012E-23											
155	species153	4.22879E-24											
156	species154	2.46725E-23											
157	species155	4.14472E-24											
158	species156	2.89012E-23											

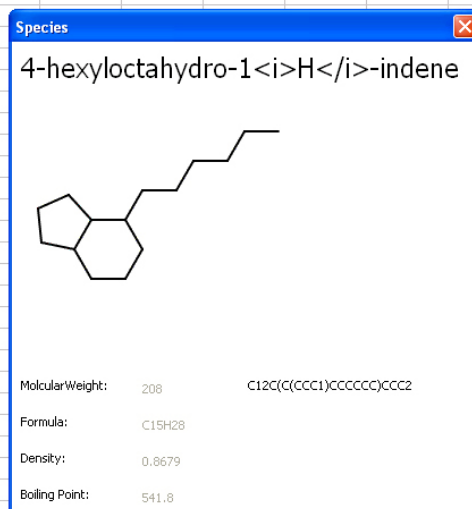


Figure 2.21 KME ready results view

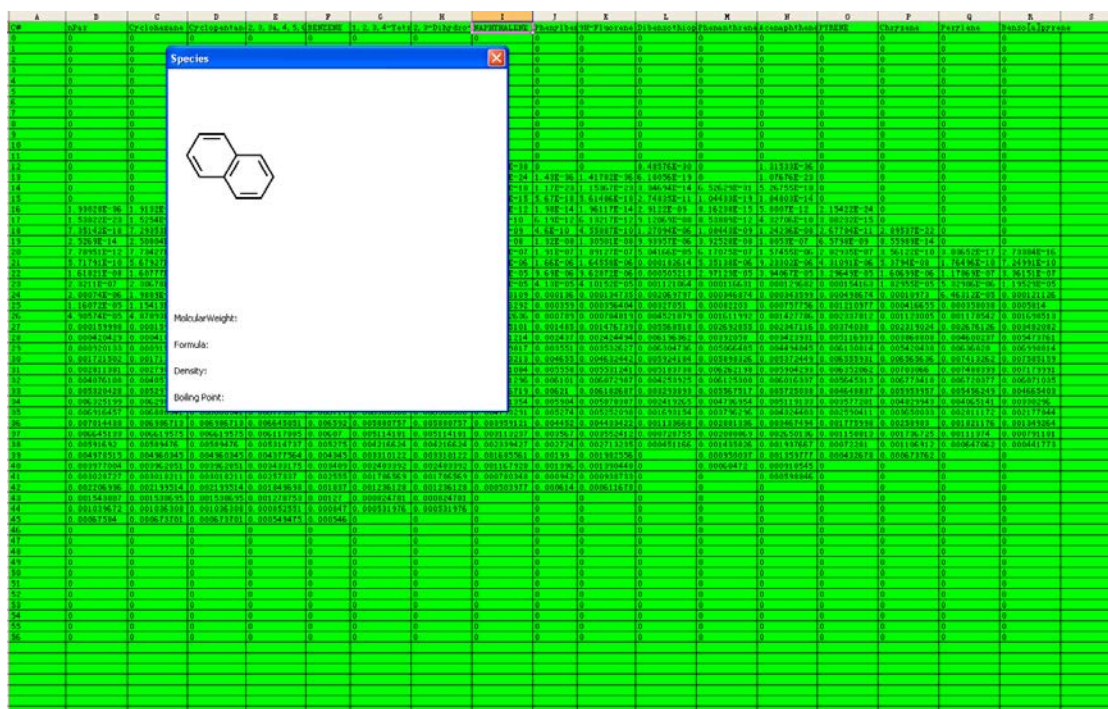


Figure 2.22 PONA Matrix results view

Once the MCA0 of feedstock is modeled, it serves as the initial values necessary for reaction modeling.

The full and final MCA will include the products of those reactions.

2.7 Integration between CME and INGen

The first goal of composition modeling is the determination of the identities of the system's species. The scale of the model is determined by a mixture of necessary detail, data, and the feasibility of resultant computational burden.

As discussed in previous sections, the identities of a feedstock could be either determined by CME or imported from INGen. Whether INGen or CME should provide the qualitative determination of the model species is a sort of the “chicken” or “egg” problem. If the species' identities are to come from CME first, CME will use the analytical information along with conditional probability to constrain the

size of the MCA0. As a result, the MCA0 from CME will provide the seeding information for INGen. INGen will then determine the scale and details of the reaction networks.

However, the determination of the core series in CME needs to consider the reactivity characteristics used by INGen. In addition, the set of molecules in the reaction network generated by INGen can provide more isomeric detail than CME. For instance, CME's CoreGen specifies the identity of the core structure, the sidechain length and the number of sidechains, but not the substituted position. INGen has this ability and can provide detailed isomeric information to CME.

As discussed above, CME provides constraint based on analytical information, and INGen provides greater structural details and reactivity prerequisites. Therefore, it was necessary to find an appropriate way to harmonize them. User knowledge and input through extensive interaction between CME and INGen reconciled this issue but required a seamless integration between the two programs. That integration was realized through the incorporation of a common set of BE Matrices.

2.8 Summary and discussions

It is possible to develop a statistical representation of a complex feedstock by constraining the representation to a set of analytical characterizations. The answer for this representation can be defined as MCA, a matrix containing both qualitative and quantitative information of the molecules.

However, it is important to consider the computational burden a real world (especially for heavy complex feeds) might occur. Although it is clear that a stochastic sampling method can get a statistical representation for a complex feedstock, based on previous researchers' results, it is unnecessary to model every feedstock in this way. Moreover, this stochastic method will cast a tremendous computational burden and thereby bring a formidable challenge in employing on the complex feedstocks found in practical applications. To address this issue, a hybrid statistical approach has been developed in this chapter. In most cases, a set of finite molecular identities called the footprint can be

determined in a straightforward process. The way to specify qualitative information in a complex feedstock can be derived explicitly or implicitly from several aspects, such as: core series determination, literature, automated reaction network generation etc.. Stochastic sampling can still be used as a supplemental computer aid for the few cases which contain uncertain identities. Therefore, in most of cases, non-deterministic methods will only be used for the calculation and optimization of quantitative information and will vary from feed to feed. Thus, a huge computational burden is avoided, making this approach more feasible for practical application.

To obtain the MCA0 (the initial identities and concentrations) of a complex feedstock, a sampling protocol based on structural attributes of molecules has been laid forth in this chapter. In addition, the structural attribute PDFs used for quantitative calculation have been extensively discussed in this chapter. The combination of attribute sampling with imposed attribute PDF constraints can be used to optimize the MCA0 where the objective function is given in terms of available analytical information.

A user-friendly program with an Excel-VBA interface called CME was developed to fulfill this new approach. CME automates the composition modeling for complex feedstocks varying from light to extremely heavy.

Although the examples of this hybrid approach emphasize on complex petroleum feedstocks, the methodology should hold true for others (such as the increasingly important biomass).

A biomass can be modeled in CME via imposing new attribute definition and sampling protocol, altering the PDFs, and creating an appropriate objective function. A specific example of biomass processing (lignin pyrolysis) is discussed in Chapter 6. Unlike petroleum feedstocks, oxygen containing compounds such as alcohols, phenols, etc. are involved in these biofuels. Exploring new oxygen-related chemical attributes would be a key outstanding task. In addition, other complex feedstocks such as coal, shale oil, etc. should be addressed in future.

The final goal of this chapter was to illustrate a method for obtaining the MCA0, a molecular representation of a complex feedstock. MCA0 can then be used to provide the input for a reaction network generator, or to provide the initial concentrations for a molecular level reaction system model. Both aspects will be discussed in detail in the next chapter.

Reference

-
- ¹ Craig, Bennett, "User-controlled kinetic network generation with InGen", Doctoral Dissertation, Rutgers University(2009)
- ² Quann, R. J., and Jaffe, S. B., "Structure Oriented Lumping. Describing the chemistry of complex hydrocarbon mixtures", Ind. Eng. Chem. Res., 31(11), pp. 2483-2497 (1992)
- ³ Quann, R. J., and Jaffe, S. B., "Building Useful Models of Complex Reaction Systems in Petroleum Refining", Chem. Eng. Sci., 51(10), pp. 1615 (1996)
- ⁴ Flory, P.J., Molecular Size Distribution in Linear Condensation Polymers. J. Am. Chem. Soc., 1936. 58: p. 1877.
- ⁵ Libanati, C., Monte Carlo Simulation of Complex Reactive Macromolecular Systems, in Chemical Engineering. 1992, University of Delaware.
- ⁶ Whitson, C.H., Characterizing Hydrocarbon Plus Fractions. Soc. Pet. Eng. J., 1983: p. 683.
- ⁷ Shibata, S.K., S.I. Sandler, and R.A. Behrens, Phase Equilibrium Calculations for Continuous and Semicontinuous Mixtures. Chemical Engineering Science, 1987. 42(8): p. 1977-1988.
- ⁸ Pederson, K.S., A.L. Blilie, and K.K. Meisingset, PVT Calculations of Petroleum Reservoir Fluids Using Measured and Estimated Compositional Data for the Plus Fraction. Ind. Eng. Chem Res., 1992. 31: p. 1378-1384.
- ⁹ Boduszynski, M.M., Composition of Heavy Petroleum. 1. Molecular Weight, Hydrogen Deficiency, and Heteroatom Concentration as a Function of Atmospheric Equivalent Boiling Point up to 1400 °F (760 °C). Energy and Fuels, 1987. 1: p. 2-11.
- ¹⁰ Petti, T.F., et al., *CPU Issues in the Representation of the Molecular Structure of Petroleum Resid Through Characterization, Reaction, and Monte Carlo Modeling*. Energy and Fuels, 1994. 8(3): p. 570-575
- ¹¹ Trauth, D.M., et al., Representation of the Molecular Structure of Petroleum Resid through Characterization and Monte Carlo Modeling. Energy and Fuels, 1994. 8(3): p. 576-580.
- ¹² Darivakis, G.S., W.A. Peters, and J.B. Howard, Rationalization for the Molecular Weight Distribution of Coal Pyrolysis Liquids. AIChE J., 1990. 36(8): p. 1189.
- ¹³ Trauth, D.M., Structure of Complex Mixtures through Characterization, Reaction, and Modeling, Ph. D. Dissertation, University of Delaware, 1993.
- ¹⁴ Campbell, D.M., Stochastic Modeling of Structure and Reaction in Hydrocarbon Conversion, Doctoral Dissertation, University of Delaware (1998)
- ¹⁵ Gang, Hou, Integrated chemical engineering tools for the building, solution, and delivery of detailed kinetic models and their industrial applications, Doctoral Dissertation, University of Delaware (2001)

Chapter 3 Automation of reaction equation generation and calculation

3.1 Introduction

The most important three aspects of a rigorous molecule-based detailed kinetic model for complex process chemistries are its reaction network, the equation generation and solution techniques of its networks, and the kinetic rates. The construction of the reaction network is discussed in Bennett (2009)¹. The two aspects will be addressed in this chapter, and will serve as the second proposed contribution of this thesis. This chapter will emphasize and to address the dilemma of the complexity of the reaction modeling including attribute-based reaction model, reaction family and LFER, LHHW support, deactivation calibration, multi reactor support and model analysis. The automation techniques of these aspects have been employed on the Kinetic Modeler's Editor (KME) software introduced in chapter1.

A traditional deterministic model has $N + 2$ equations, one material balance for each of the N components along with energy and momentum balances. This downside has been addressed through the development of software tools for the construction, solution and editing of large models. At a certain level of complexity, however, the number of components N (e.g. over 50,000) is so large that even automated construction tools will not overcome the hardware and software barriers and execution time demanded. In these instances the essential question is whether the N -component mixture can be modeled with fewer than N material balance equations. An approach called attribute-based reaction modeling (ARM) has been developed toward this end and will be discussed in later section.

Another inherent problem associated with detailed kinetic modeling is that thousands of reactions require thousands of rate constants associated with the reaction network. It is impractical, if not impossible, to experimentally determine the rate constant for each reaction through thousands of model

compound experiments. In addition, the concurrent reconciliation of so many rate constant parameters with experimental data falls outside the scope of any currently available optimization algorithms. It is thus clear that approaches aiming at reducing the complexity and number of model parameters would be of great value in the formulation and use of detailed molecular models of complex mixtures.

There are various levels of approaches to handle this problem. A first approximation is that each reaction family is assumed to have the same kinetic parameters since all the reactions undergo similar intra-molecular rearrangement. However, this is generally insufficient considering the perturbation on the reaction rates of a wide range of molecular structures even for the same reaction family. The other extreme is the direct computation of rate constants from first principles through the implementation of quantum chemical calculations. While the quantum mechanics theory has grown into an unprecedented level, the molecular size it can handle or calculate in a feasible amount of time is thus far still limited. A practical resolution of this conflict emerges from a further scrutiny of the composition and reactions of complex feedstocks. Much of the complexity is statistical. Each of the $O(10^5)$ species in the complex mixture falls into one of a handful of compound classes (e.g., paraffins, olefins, naphthenes, aromatics, alkylaromatics); and these in turn react in a manageable set of reaction families (e.g., hydrogenation, isomerization, dealkylation). Thus, the complexity is really the simultaneous reaction of sets of many similar compounds. Within each set, compounds differ only in substitutes, and differences in reactivity are attributable to these substituents. This suggests the use of linear free energy relationship (LFER) corresponding with reaction families, as an organizational and predictive technique for organizing or assembling kinetic rates for detailed kinetic modeling. An automation technique of this will be discussed in this chapter.

The organization of rate laws for the reaction equations in complex process chemistries is another issue to be addressed in this chapter. Most complex process chemistries are heterogeneous phase and involved with a catalytic process. An efficient way to describe the rate laws of these chemistries is the

Langmuir-Hinshelwood-Hougen-Watson (LHHW) formalism. An automation technique of LHHW in KME will be addressed in later section.

Ultimately, the detailed kinetic model must be able to model the system and process configurations that are found outside of bench scale including: catalyst deactivation, multiple reactors, and separate process. An automation technique for these advanced configurations is also covered in this chapter.

Finally, when modeling a complex system, it is necessary to provide auxiliary analytical tools such as: statistical analysis, sensitivity analysis and model reduction. These auxiliary tools will be discussed in the last section of this chapter.

3.2 Attribute-based reaction model (ARM)

As addressed in Chapter 1, a deterministic model comprising one equation per species will bring about an impractical computational task when dealing with complex process chemistries with complex mixtures of numerous species. To address this dilemma, an Attribute-based Reaction Model (ARM) has been developed for the kinetic modeling of heavy hydrocarbon conversion. This is a hybrid statistical-molecular approach to represent the reactions as well as the molecular structures of heavy hydrocarbon mixtures. Unlike the deterministic model, this approach models the dynamics of the molecular composition with fewer than one equation per molecule. For that purpose, attention is turned to the reactions of attributes instead of species. The molecular compositions of the feedstock parse to a set of structural attributes as illustrated in Chapter 2. The reactant attributes derive from those structural attributes directly and can be modified based on the characteristics of reactivity. The conversion pathways and kinetics are subsequently modeled in terms of the reactions of these attributes. The molecular composition of the converted mixture is assembled by sampling the updated PDF's of the post-reaction attributes. This provides the molecular basis for the estimation of the properties of the upgraded mixture.

3.2.1 ARM example: lignin

For example, lignin pyrolysis is a good case to show how ARM works. The structure of lignin can be reduced to a set of single-ring aromatics each with two attributes as shown in Figure 3.1. The first attribute is the type of propanoid side chain (PC) attached to each aromatic ring and the second attribute is the nature of the phenolic or methoxyphenol (MP) substituent on each aromatic ring.

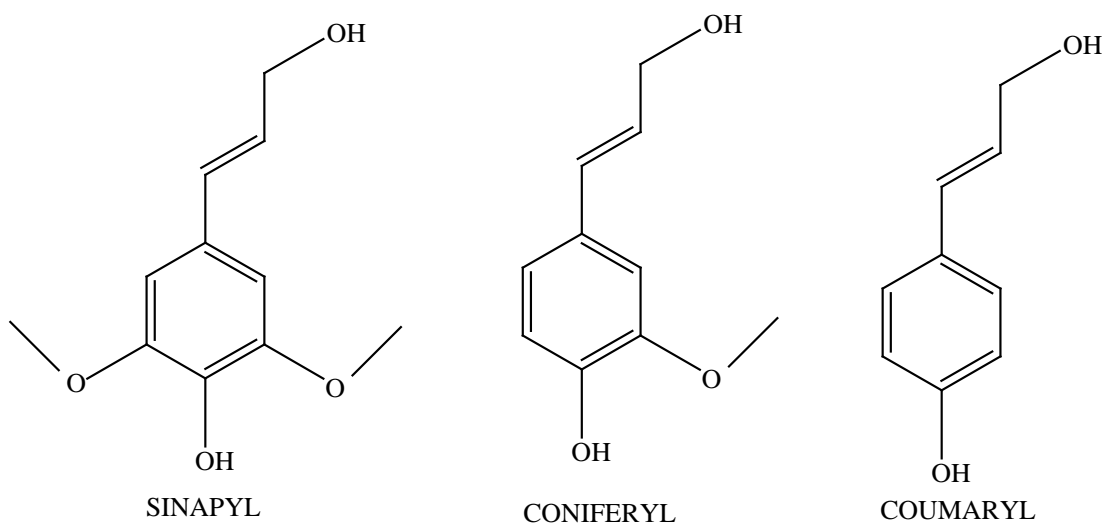


Figure 3.1 the structure of lignin

During lignin pyrolysis, the aromatic rings remain unchanged and the conversion is derived from the independent reactions of the PC and MP attributes. Therefore, it is possible to use PC and MP attributes as the reactants in ARM. The equations of reaction can be written as Eq. 3.1.

Attribute 1 Submodel :

$$\begin{aligned}
 \frac{dMP_1}{dt} &= R_{11} \\
 \frac{dMP_2}{dt} &= R_{12} \\
 &\dots \\
 \frac{dMP_M}{dt} &= R_{1M}
 \end{aligned}
 \tag{3.1}$$

Attribute 2 Submodel :

$$\begin{aligned}
 \frac{dPC_1}{dt} &= R_{21} \\
 \frac{dPC_2}{dt} &= R_{22} \\
 &\dots \\
 \frac{dPC_N}{dt} &= R_{2N}
 \end{aligned}
 \tag{3.2}$$

The classic Freudenberg model² of lignin contained 24 values of MP attributes and 26 values of PC attributes. ARM will use only 24+26=50 equations instead of the 24*26=624 equations required by deterministic modeling, and thus it reduces the computation burden significantly. The detailed illustration of this lignin example will be discussed in Chapter 6.

3.2.2 General illustration of ARM

A general illustration of the ARM approach is shown in Figure 3.2, where a molecular composition with N attribute types with M attribute values will have M*N molecules. The associated deterministic reaction model will thus have M^N balance equations. This number can exceed 50,000 for petroleum resids and similar large feed stocks and is too large for current practical applications. The essential idea behind the ARM approach is to treat the reactions of the attributes independently. The ARM is thus a hybrid model of N submodels, each submodel having the number of equations equal to the number of values the attribute can take. Therefore, the overall problem scales as N×M instead of M^N, which leads to a considerable savings of CPU time. This is the approach followed in the lignin pyrolysis model, where N = 2.

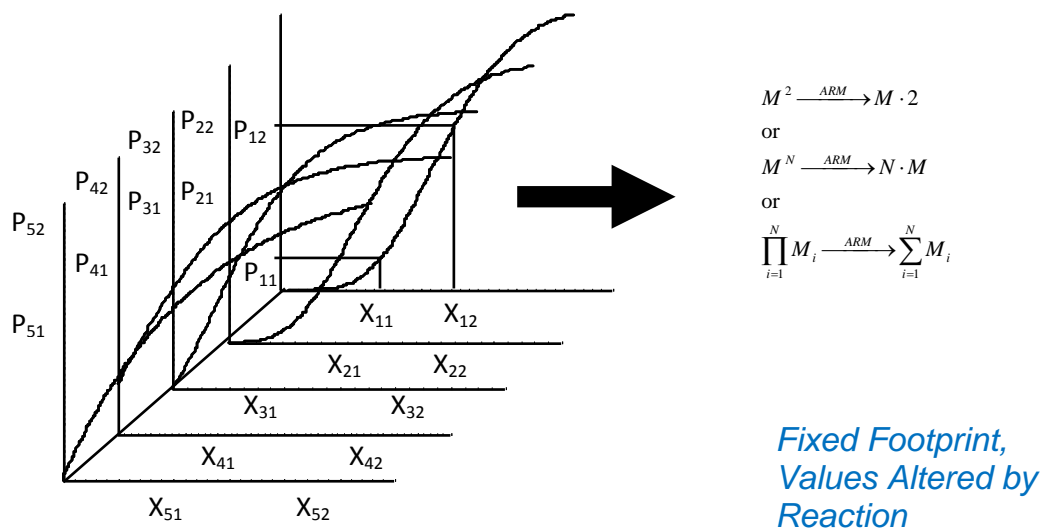


Figure 3.2 Attribute-Based Reaction Modeling

3.2.3 Automation of ARM in KME

This ARM approach was implemented in KME. The previous versions of KME were basic one-dimensional model bundles of tools used only for deterministic modeling. Essentially, extension to include the ARM strategy involves the creation of N independent kinetic models with shared information, such as reaction conditions, inlet flows, and the like, and a final product parser that creates the molecular composition as the juxtaposition of attributes. The computational flow of KME is illustrated in Figure 3.3.

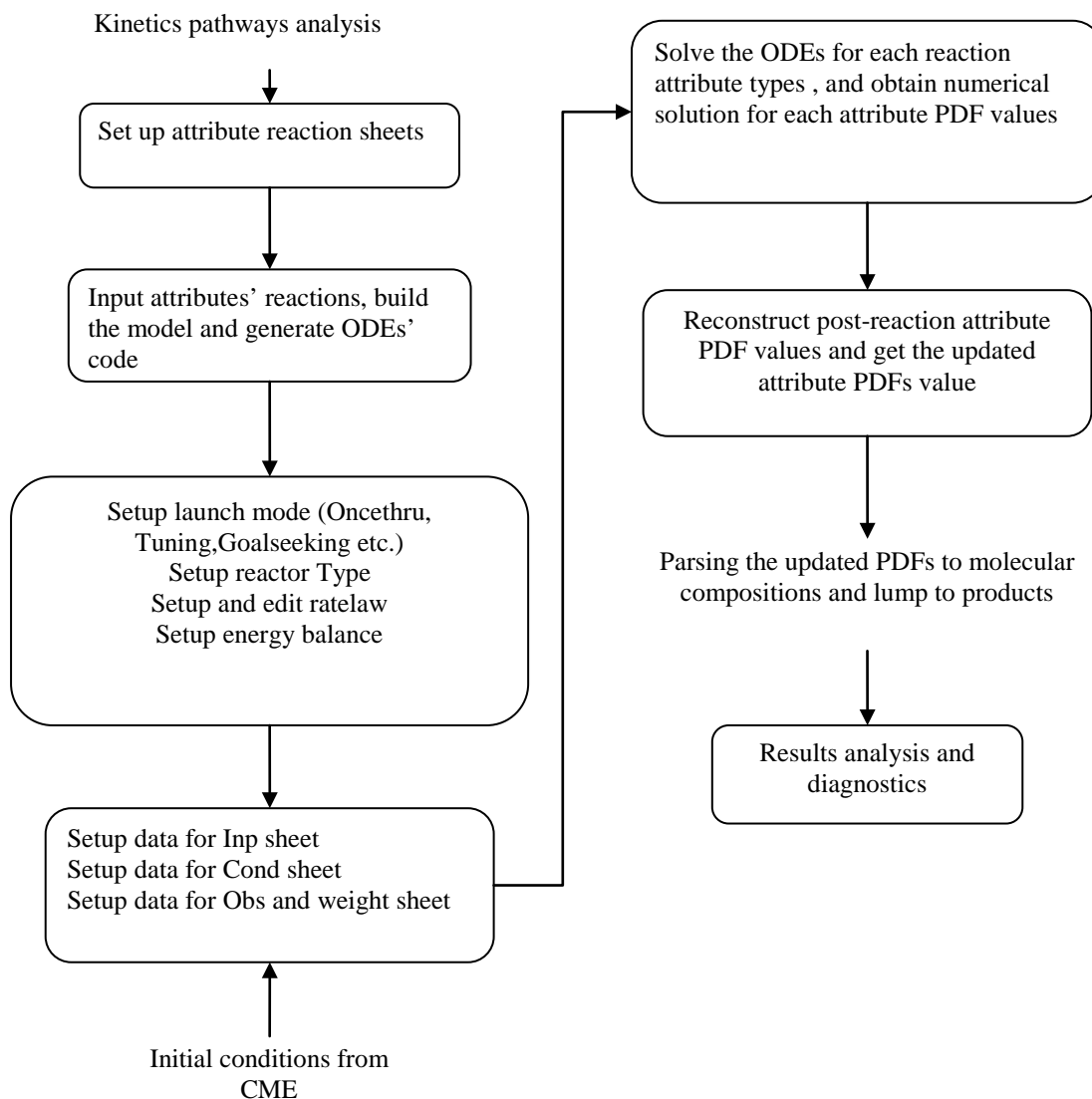


Figure 3.3 Conceptual flow of KME with ARM

The ARM model in KME comprises the following steps. First, the reactions of each attributes are set up independently. It then generates the ODE's for all attributes' submodels. KME obtains the initial attributes' values from CME. The user then enters the input and output data, along with the reaction conditions and rate parameters, which allow the model to be run. The submodels of each attribute types will be run respectively. The reaction-altered attributes are then juxtaposed to create the conversion mixtures and thus attain the properties of the final product.

3.3 Reaction family and LFER

Detailed kinetic modeling for complex process often contains thousands of reactions and therefore thousands of rate constants associated with the reaction network. It is impossible, or merely impractical, to either evaluate the rate constant for each reaction through thousands of model compound experiments or reconcile so many rate constant parameters at the same time with the experimental data using any currently available optimization algorithm. A conceptual approach to this problem is the use of reaction family and LFERs in order to reduce the complexity and number of model parameters. A reaction family is a series of analogous reactions that share the same reaction mechanism.

In general, the concept that there is a relationship between the properties of compounds and their molecular structures is inherent to chemistry. It is the basic tenet of chemistry to attempt to identify these relationships between molecular structure and property or activity and to quantify them. Hammett³ organized reaction families and substituent effects for homogeneous systems. The Evans-Polyani⁴ relationship is another classical example. Mochida and Yoneda^{5,6,7} first demonstrated the use of LFERs for catalytic reactions. Klein and coworkers (Neurock^{8,9,10}; Korre¹¹; Broadbelt¹²) have extensively developed the LFERs to correlate reaction rates of various metal-, acid-catalytic chemistries as well as free-radical chemistry. LFERs have also been developed as a convenient way to organize the equilibrium and adsorption constants (Ho, 1994; Korre¹⁰).

3.3.1 LFERs

A linear free energy relationship (LFER) can be derived from transition state theory, wherein a general kinetics rate expression of reaction *i* in a reaction family can be expressed as Eq. 3.2.

$$\ln \frac{k_i}{k_0} = \frac{\Delta S_i^+ - \Delta S_0^+}{R} - \frac{\Delta H_i^+ - \Delta H_0^+}{RT} \quad (3.2)$$

In Eq. 3.2, subscript (0) refers to an arbitrary reference reaction, while subscript (i) refers to any other reaction in the family, ΔS_i^+ and ΔH_i^+ are the entropy and enthalpy differences between transition state complexes and reactants. Particularly, for the reactions within the same family, as long as the reaction center is spatially separated from the substituents (other groups, heteroatoms, alkyl chains), the sterics of the transition state will not change significantly. Rather, the substituents will affect only the activation energy. Therefore, $(\Delta S_i^+ - \Delta S_0^+)$ can be considered negligible, and Arrhenius factors, A_i , of any reactions in the same family will be equivalent to A_0 . Therefore, an approximation of Eq. 3.2 is given by in Eq. 3.3 which is only a function of activation enthalpy.

$$\begin{aligned}\Delta S_i^+ - \Delta S_0^+ &= \Delta(\Delta S^+) = 0 \\ \frac{A_i}{A_0} &= 1 \\ \ln \frac{k_i}{k_0} &= -\frac{1}{RT}(\Delta H_i^+ - \Delta H_0^+) = -\frac{1}{RT}\Delta(\Delta H^+) \\ \ln \frac{k_i}{k_0} &= -\frac{1}{RT}\Delta(\Delta H^+) = -\frac{(E_i^* + RT - E_0^* - RT)}{RT} = -\frac{1}{RT}\Delta(E_i^*) \\ \text{generally, } \Delta(\Delta H^+) &\approx (RI)_i \\ \ln k_i &= a + b(RI)_i\end{aligned}\tag{3.3}$$

In E.q.3.3, a and b are correlation constant parameters, $\Delta(\Delta H^+)$, $\Delta(\Delta S^+)$ and $\Delta(E_i^*)$ refer to the differences in activation enthalpy, entropy and energy differences between reaction (i) and reaction (0). In Eq. 3.3, the rate constant of any reactions in a specific reaction family, k_i , can be retrieved via a linear relationship with the differences in activation enthalpy. For complex system, $\Delta(\Delta H^+)$ can be estimated by a reaction index (RI) that is a quantitative variable inherent to the properties of the reactants. A set of representative examples of LFER are shown as the next section.

3.3.2 Representative results of LFER

Figure 3.4 (Korre¹⁰, 1995) shows very good correlations between the experimental hydrogenation rate constants with the heat of reaction as the reactivity index for various aromatics, including benzenic, naphthalenic, and phenanthrenic compounds. The highlighted rings denote the site of hydrogenation. Although the hydrogenation of aromatic compounds seems to provide special challenges of accounting for the multiplicity of the reaction sites per molecule, the overall heat of reaction has served as a very good reactivity index to characterize these reactions.

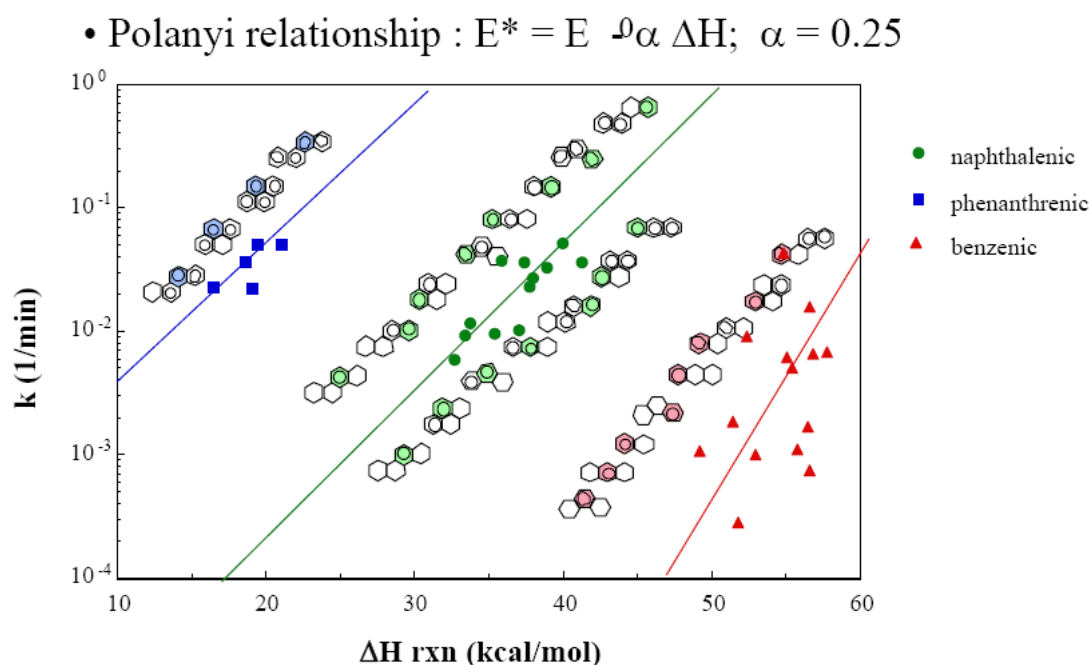


Figure 3.4 LFER correlations of experimental hydrogenation rate constants and heat of reaction for aromatic compounds. (Korre¹⁰, 1995)

Figure 3.5 shows LFER correlations for the kinetics of acid center transformation reaction families (both isomerization and ring opening) with the heat of formation of the carbenium ion intermediate as the reactivity index. Both the isomerization and ring opening reactions share a similar underlying reaction mechanism, and both have the same slope in the LFER diagram. However, it is clear that the isomerization reaction is much slower than the ring opening reaction.

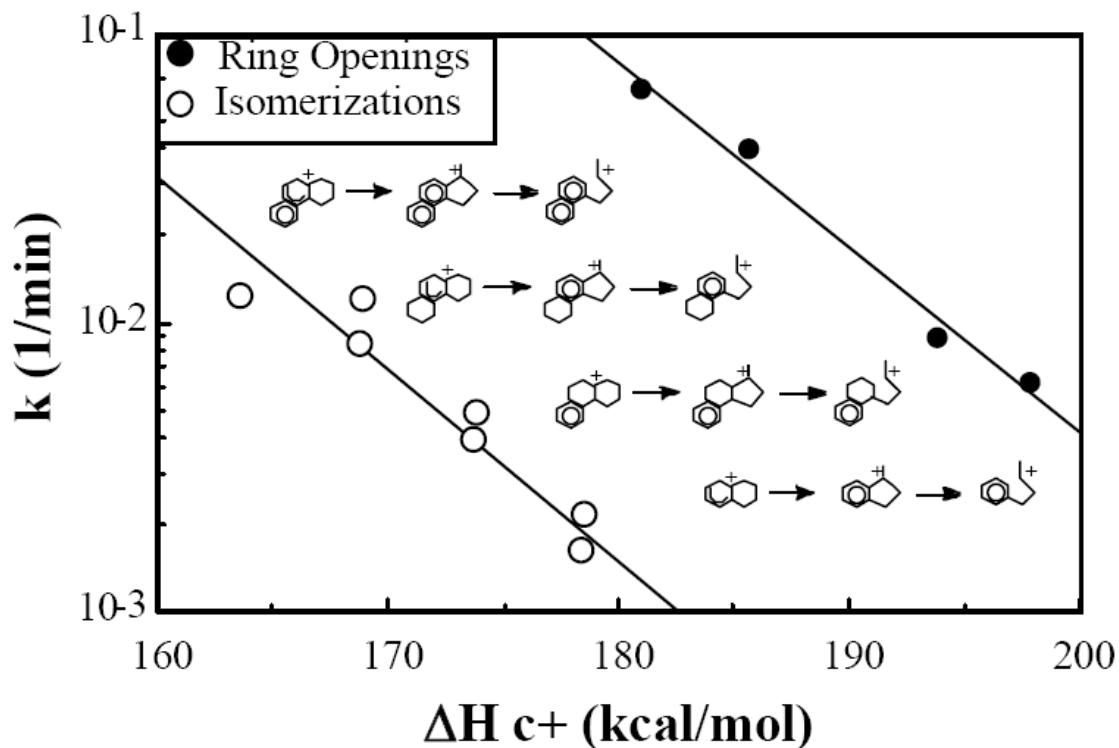


Figure 3.5 LFER correlation of the experimental isomerization and ring opening rate constants with the stability of the carbenium ion intermediates. (Korre¹⁰, 1995)

Figure 3.6 shows the classical example of LFER correlations by Mochida and Yoneda between the dealkylation rate parameters and the stability of the alkyl ion based on a carbenium ion chemistry reaction mechanism. Neurock^{7,8,9} has also used the heat of reaction of the rate-determining step based on the carbenium ion chemistry reaction mechanism as a reactivity index and successfully correlated with the dealkylation kinetics.

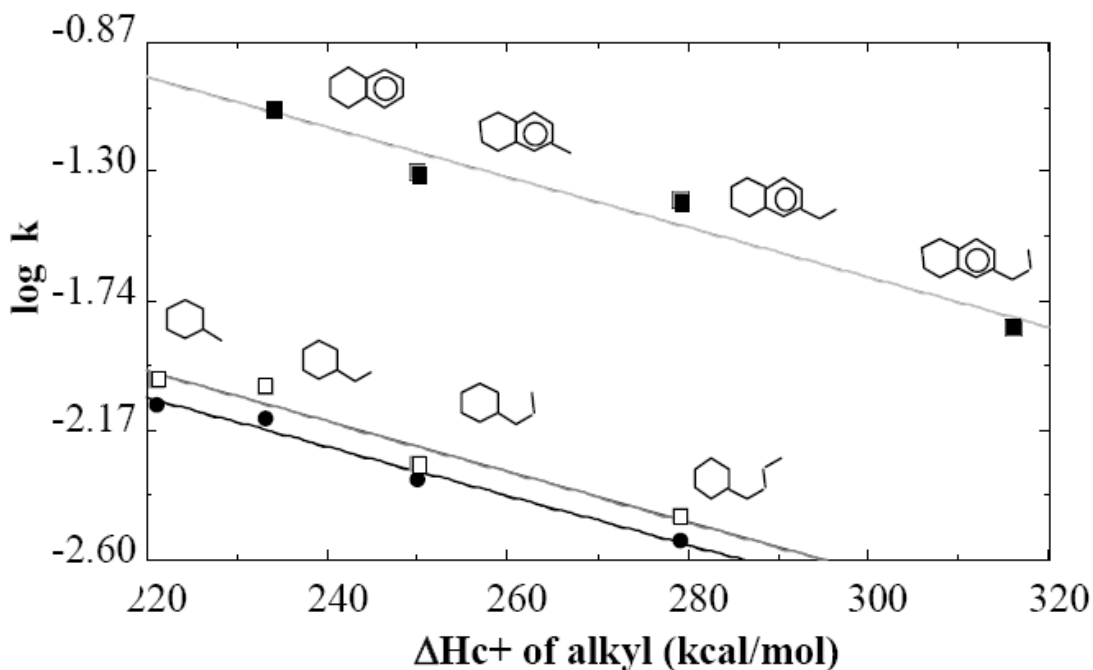


Figure 3.6 LFER correlation of experimental dealkylation rate constant with the heat of formation of the intermediate carbenium alkyl ion. (Data from Mochida and Yoneda⁴, 1967)

Besides using LFER on kinetics rates, the Klein group has also applied this concept to equilibrium constants and adsorption constant.

Figure 3.7 (Neurock⁹, 1992) shows an LFER correlation used to estimate adsorption constants. The experimental adsorption coefficients were determined through experiments where competitive inhibition data were represented by the LHHW adsorption constant K_i (La Vopa and Satterfield, 1988¹³; Neurock and Klein⁹). The proton affinity (PA) was taken as a measure of gas-phase basicity and computed according to the thermochemical cycle: $M + H^+ \rightarrow MH^+$. Thus, the PA is given by E.q. 3.4 and can be computed using MNDO (modified neglect of diatomic overlap) calculations. As we can see from Figure 3.7, the MNDO calculated PAs correlate linearly with the experimental adsorption parameters.

$$PA = -\Delta H = \Delta H_0(M) + \Delta H_0(H^+) - \Delta H_0(MH^+) \quad (3.4)$$

This good correlation shows the organizational capability of LFERs and suggests an efficient methodology for the estimation of kinetic parameters for many similar compounds that have not been studied experimentally. In other words, finding the adsorption constant of a similar molecule would simply require a PA calculation. From the chemical significance perspective, the correlation between K and PA actually points to an acid-base interaction between the catalyst and reactant¹².

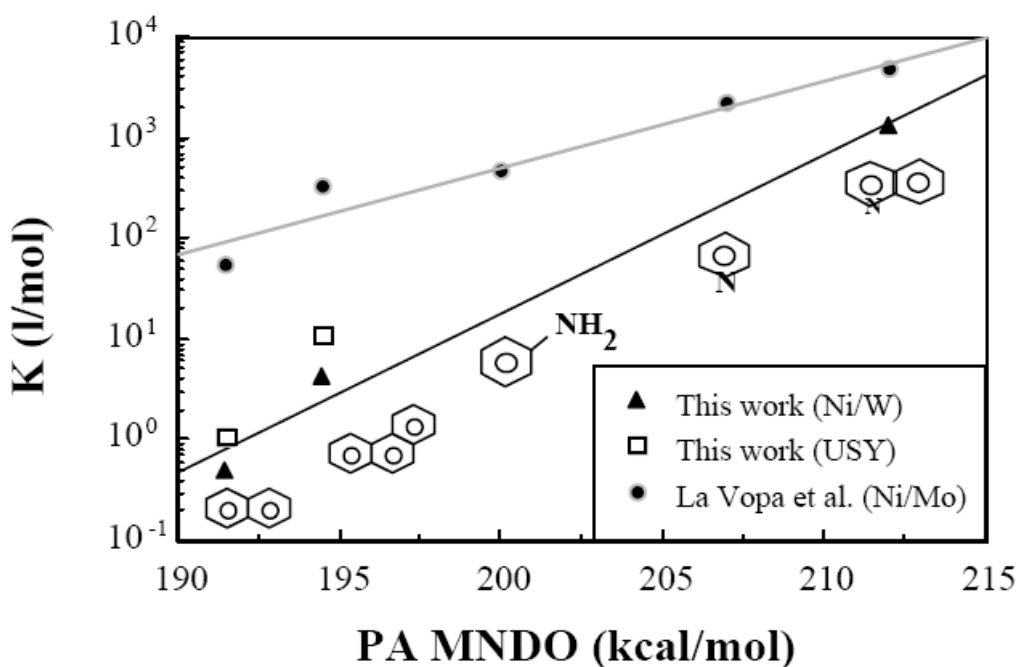


Figure 3.7 LFER correlation of proton affinity (PA) to estimate the adsorption constants K . (Neurock⁹, 1992)

The above examples qualitatively demonstrate the feasibility of using LFER concept as a reaction engineering tool to organize the rate parameters. The following figures (Figure 3.8, Figure 3.9, and Figure 3.10) summarize a comprehensive set of quantitative correlations and LFER parameters developed in the past few years by the Klein group.

Adsorption LFER: Metal Sites

$$\ln K_i (\text{l/mol}) = 2.04 + 0.654 N_{\text{AR}} + 0.0964 N_{\text{SC}}$$

N_{AR} = Number of aromatic rings

- Activity: Catalyst/PNA acid-base interaction

N_{SC} = Number of saturated carbons

- Availability: Size of the molecule

28 adsorption constants ----> 3 QS/RC parameters

Figure 3.8 LFER for adsorption constants by Klein group¹⁴.

LFER's for Metal Center Transformations

HYDROGENATION REACTION FAMILY:

	<u>Regressed</u>	<u>Variables</u>	<u>QS/RC Parameters</u>
Rate:	80	ΔH_{rxn} , n, T	3
Equilibrium:	80	ΔH_{rxn} , n, T	2
Adsorption:	136	NAR, NSC, T	3

$$\ln k = -(6.61 + 5.76n) + \frac{0.228\Delta H_{\text{rxn}}^0}{RT}$$

$$\ln K_{\text{eq}} = -7.02n + (0.547 - \frac{1}{RT})\Delta H_{\text{rxn}}^0$$

$$\ln K_{\text{ads}}^{\text{H}} = 1.32 + \frac{0.887N_{\text{AR}} + 0.123N_{\text{SC}}}{RT}$$

Figure 3.9 LFER for metal center catalysis by Klein group¹⁴

LFER's for Acid Center Transformations

ACID CENTER TRANSFORMATIONS REACTION FAMILY:

	<u>Regressed</u>	<u>Variables</u>	<u>QS/RC Parameters</u>
Rate:	72	ΔH_{c^+} , T	5
Equilibrium:	72	ΔH_{rxn} , T	2
Adsorption	136	NAR, NSC, T	3

$$\ln k = A - B\left(0.585 - \frac{1}{RT}\right) + \left(1 - \frac{1}{0.585RT}\right)\Delta H_{\text{f}}^{\text{c}^+}$$

Isomerizations: A=-2.89, B=-305. Ring Openings: A=lnP_{H2}-4.64, B=-271.

$$\ln K_{\text{eq}} = -1.78 + \left(1.47 - \frac{1}{RT}\right)\Delta H_{\text{rxn}}^0$$

$$\ln K_{\text{ads}}^{\text{A}} = 0.182 + \frac{1.93N_{\text{AR}} + 0.187N_{\text{SC}}}{RT}$$

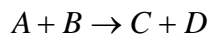
Figure 3.10 LFER for acid center catalysis by Klein group¹⁴

Based on the above summaries, an automated technique for incorporating LFERs and reaction families was developed in KME.

3.3.3 LFERs and reaction families in KME

In KME, the kinetic rate of an irreversible reaction is expressed by a function of Arrhenius factors A_i and active energy E_i^* , as shown in E.q.3.5a. For a reversible reaction, an equilibrium constant K_{eq_i} is applied to the backward reaction in E.q. 3.5b.

For an irreversible reaction i



$$r_i = k_i (\lg A_i, E_i^*) \cdot (y[A] \cdot y[B]) \quad (3.5a)$$

For an reversible reaction i



$$r_i = k_i (\lg A_i, E_i^*) \cdot (y[A] \cdot y[B] - y[C] \cdot y[D] / K_{eq_i}) \quad (3.5b)$$

Here,

$$k_i = 10^{\lg A_i} \cdot \exp\left(-\frac{E_i^*}{RT}\right)$$

In this thesis, the Evans-Polanyi principle will be used as an example of how to apply LFER and reaction family in KME. The Evans-Polanyi principle (E.q.3.6) is a classic methodology for activation energy via LFER.

$$E_i^* = E_0^* + \alpha \cdot \Delta H_{rxn} \quad (3.6)$$

Through experimental observations, it was determined that the activation energies of many reactions in the gas phase linearly correlate with the enthalpy change during the reaction. E_0^* refers to an arbitrary reference activation energy which defines the intercept ; α is a parameter that indicates the slope. As a result, the rate constants in a reaction family can be expressed as a function of the enthalpies of reaction, which are more easily accessible since they are differences between heats of formation of stable compounds.

$$\ln \frac{k_i}{k_0} = -\frac{1}{RT} \Delta(E_i^*) = -\frac{\alpha}{RT} (\Delta H_{rxn}) \quad (3.7)$$

$$\ln k_i = \ln k_0 - \frac{\alpha}{RT} \cdot (\Delta H_{rxn}) = A_0 - \frac{E_0^* + \alpha \cdot \Delta H_{rxn}}{RT}$$

From Eq.3.7, the rate of reactions in the same family is the function of A_0 , E_0 and ΔH_{rxn} .

For a reversible reaction, if k_i is the forward reaction, the rate of backward reaction can be expressed as the combination of k_i and equilibrium constant K_{eq} , given by E.q. 3.5b. K_{eq} can be expressed via Van

Hoff Law as a function of ΔH_{rxn} , ΔS_{rxn} and temperature, or correlated from an LFER as a function of ΔH_{rxn} and temperature as shown in Figure 3.8 to Figure 3.10.

The enthalpy and entropy of a reaction are inherent thermodynamics properties of the reactants and can be measured directly or calculated from the structure-property correlation as illustrated in Chapter 4. These properties are independent of the catalyst, so for an entire reaction family only three parameters will require tuning (A_0 , E_0 and α).

Generally, every process chemistry contains of order 10 reaction families and thus the number of tuning parameters can be of a computationally practical order 10-30 when using the LFER approach. This is a practical computational burden. As a product of this thesis, KME now provides an Excel-VBA interface for automating reaction family and LFER approximation. First, it will set up the reaction family information as shown in Figure 3.11. KME will automatically generate the LFER rate expression for each reaction family (Figure 3.12) by applying the expressions given by E.q. 3.5, E.q. 3.6 and E.q. 3.7. By default, the equilibrium constant is implemented by Van Hoff equation, but KME provides a flexible way to modify it to a LFER expression as a correlation of the species' reaction index (e.g., ΔH_{rxn}).

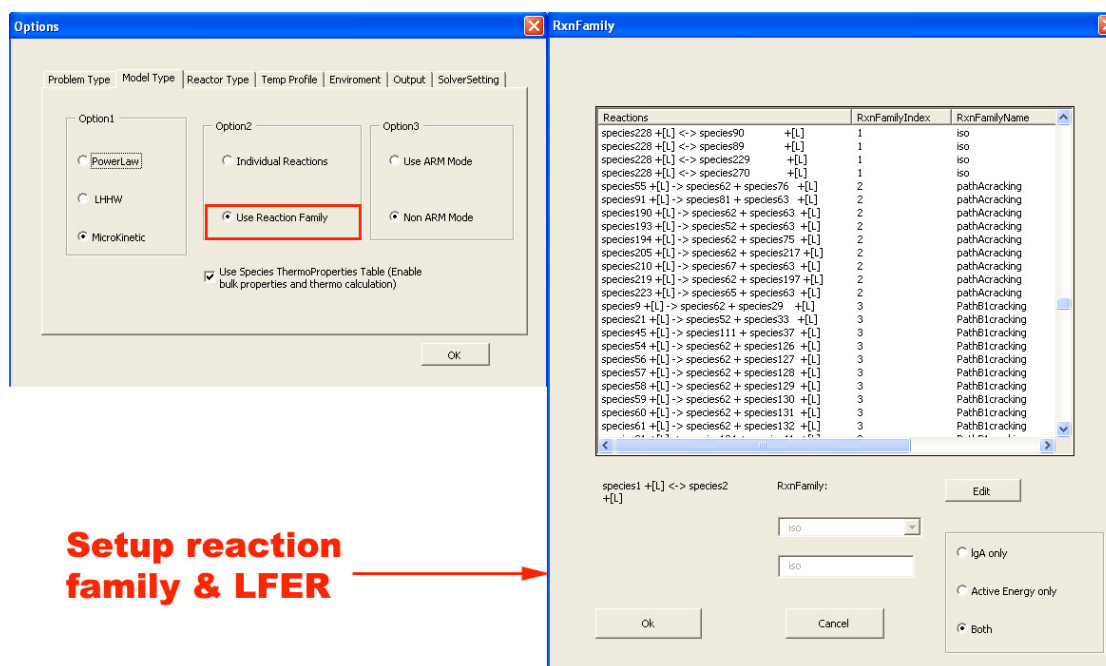


Figure 3.11 LFER & RXN Family in KME

A	B	C	D
10	8	species1 +[L] <-> species4	+ [L] rxn[7] = k(IgA1, GetE(deltaH[7], dS1, E1)) * y(species1) * y(L) - y(species4) * y(L) / Keq[7];
11	9	species1 +[L] <-> species5	+ [L] rxn[8] = k(IgA1, GetE(deltaH[8], dS1, E1)) * y(species1) * y(L) - y(species5) * y(L) / Keq[8];
12	10	species1 +[L] <-> species6	+ [L] rxn[9] = k(IgA1, GetE(deltaH[9], dS1, E1)) * y(species1) * y(L) - y(species6) * y(L) / Keq[9];
13	11	species1 +[L] <-> species7	+ [L] rxn[10] = k(IgA1, GetE(deltaH[10], dS1, E1)) * y(species1) * y(L) - y(species7) * y(L) / Keq[10];
14	12	species2 +[L] <-> species8	+ [L] rxn[11] = k(IgA1, GetE(deltaH[11], dS1, E1)) * y(species2) * y(L) - y(species8) * y(L) / Keq[11];
15	13	species2 +[L] <-> species9	+ [L] rxn[12] = k(IgA1, GetE(deltaH[12], dS1, E1)) * y(species2) * y(L) - y(species9) * y(L) / Keq[12];
16	14	species2 +[L] <-> species10	+ [L] rxn[13] = k(IgA1, GetE(deltaH[13], dS1, E1)) * y(species2) * y(L) - y(species10) * y(L) / Keq[13];
17	15	species2 +[L] <-> species11	+ [L] rxn[14] = k(IgA1, GetE(deltaH[14], dS1, E1)) * y(species2) * y(L) - y(species11) * y(L) / Keq[14];
18	16	species2 +[L] <-> species12	+ [L] rxn[15] = k(IgA1, GetE(deltaH[15], dS1, E1)) * y(species2) * y(L) - y(species12) * y(L) / Keq[15];
19	17	species2 +[L] <-> species13	+ [L] rxn[16] = k(IgA1, GetE(deltaH[16], dS1, E1)) * y(species2) * y(L) - y(species13) * y(L) / Keq[16];
20	18	species2 +[L] <-> species14	+ [L] rxn[17] = k(IgA1, GetE(deltaH[17], dS1, E1)) * y(species2) * y(L) - y(species14) * y(L) / Keq[17];
21	19	species2 +[L] <-> species15	+ [L] rxn[18] = k(IgA1, GetE(deltaH[18], dS1, E1)) * y(species2) * y(L) - y(species15) * y(L) / Keq[18];
22	20	species2 +[L] <-> species16	+ [L] rxn[19] = k(IgA1, GetE(deltaH[19], dS1, E1)) * y(species2) * y(L) - y(species16) * y(L) / Keq[19];
23	21	species2 +[L] <-> species17	+ [L] rxn[20] = k(IgA1, GetE(deltaH[20], dS1, E1)) * y(species2) * y(L) - y(species17) * y(L) / Keq[20];
24	22	species3 +[L] <-> species2	+ [L] rxn[21] = k(IgA1, GetE(deltaH[21], dS1, E1)) * y(species3) * y(L) - y(species2) * y(L) / Keq[21];
25	23	species3 +[L] <-> species9	+ [L] rxn[22] = k(IgA1, GetE(deltaH[22], dS1, E1)) * y(species3) * y(L) - y(species9) * y(L) / Keq[22];
26	24	species3 +[L] <-> species4	+ [L] rxn[23] = k(IgA1, GetE(deltaH[23], dS1, E1)) * y(species3) * y(L) - y(species4) * y(L) / Keq[23];
27	25	species3 +[L] <-> species20	+ [L] rxn[24] = k(IgA1, GetE(deltaH[24], dS1, E1)) * y(species3) * y(L) - y(species20) * y(L) / Keq[24];
28	26	species3 +[L] <-> species21	+ [L] rxn[25] = k(IgA1, GetE(deltaH[25], dS1, E1)) * y(species3) * y(L) - y(species21) * y(L) / Keq[25];
29	27	species3 +[L] <-> species22	+ [L] rxn[26] = k(IgA1, GetE(deltaH[26], dS1, E1)) * y(species3) * y(L) - y(species22) * y(L) / Keq[26];
30	28	species3 +[L] <-> species23	+ [L] rxn[27] = k(IgA1, GetE(deltaH[27], dS1, E1)) * y(species3) * y(L) - y(species23) * y(L) / Keq[27];
31	29	species3 +[L] <-> species24	+ [L] rxn[28] = k(IgA1, GetE(deltaH[28], dS1, E1)) * y(species3) * y(L) - y(species24) * y(L) / Keq[28];
32	30	species3 +[L] <-> species25	+ [L] rxn[29] = k(IgA1, GetE(deltaH[29], dS1, E1)) * y(species3) * y(L) - y(species25) * y(L) / Keq[29];
33	31	species3 +[L] <-> species26	+ [L] rxn[30] = k(IgA1, GetE(deltaH[30], dS1, E1)) * y(species3) * y(L) - y(species26) * y(L) / Keq[30];
34	32	species3 +[L] <-> species27	+ [L] rxn[31] = k(IgA1, GetE(deltaH[31], dS1, E1)) * y(species3) * y(L) - y(species27) * y(L) / Keq[31];
35	33	species3 +[L] <-> species16	+ [L] rxn[32] = k(IgA1, GetE(deltaH[32], dS1, E1)) * y(species3) * y(L) - y(species16) * y(L) / Keq[32];
36	34	species3 +[L] <-> species9	+ [L] rxn[33] = k(IgA1, GetE(deltaH[33], dS1, E1)) * y(species3) * y(L) - y(species9) * y(L) / Keq[33];
37	35	species3 +[L] <-> species4	+ [L] rxn[34] = k(IgA1, GetE(deltaH[34], dS1, E1)) * y(species3) * y(L) - y(species4) * y(L) / Keq[34];
38	36	species8 +[L] <-> species10	+ [L] rxn[35] = k(IgA1, GetE(deltaH[35], dS1, E1)) * y(species8) * y(L) - y(species10) * y(L) / Keq[35];
39	37	species8 +[L] <-> species44	+ [L] rxn[36] = k(IgA1, GetE(deltaH[36], dS1, E1)) * y(species8) * y(L) - y(species44) * y(L) / Keq[36];
40	38	species8 +[L] <-> species45	+ [L] rxn[37] = k(IgA1, GetE(deltaH[37], dS1, E1)) * y(species8) * y(L) - y(species45) * y(L) / Keq[37];
41	39	species8 +[L] <-> species46	+ [L] rxn[38] = k(IgA1, GetE(deltaH[38], dS1, E1)) * y(species8) * y(L) - y(species46) * y(L) / Keq[38];
42	40	species8 +[L] <-> species47	+ [L] rxn[39] = k(IgA1, GetE(deltaH[39], dS1, E1)) * y(species8) * y(L) - y(species47) * y(L) / Keq[39];
43	41	species8 +[L] <-> species48	+ [L] rxn[40] = k(IgA1, GetE(deltaH[40], dS1, E1)) * y(species8) * y(L) - y(species48) * y(L) / Keq[40];
44	42	species8 +[L] <-> species49	+ [L] rxn[41] = k(IgA1, GetE(deltaH[41], dS1, E1)) * y(species8) * y(L) - y(species49) * y(L) / Keq[41];
45	43	species8 +[L] <-> species50	+ [L] rxn[42] = k(IgA1, GetE(deltaH[42], dS1, E1)) * y(species8) * y(L) - y(species50) * y(L) / Keq[42];
46	44	species8 +[L] <-> species51	+ [L] rxn[43] = k(IgA1, GetE(deltaH[43], dS1, E1)) * y(species8) * y(L) - y(species51) * y(L) / Keq[43];
47	45	species9 +[L] <-> species54	+ [L] rxn[44] = k(IgA1, GetE(deltaH[44], dS1, E1)) * y(species9) * y(L) - y(species54) * y(L) / Keq[44];
48	46	species9 +[L] <-> species55	+ [L] rxn[45] = k(IgA1, GetE(deltaH[45], dS1, E1)) * y(species9) * y(L) - y(species55) * y(L) / Keq[45];

Rate law in family & LFER expression

IgA

Green correlation

Figure 3.12 Rate Law by Reaction Family & LFER

3.3.4 Integration between KME and INGen

KME can be run in either deterministic or ARM mode based on the critical number of reaction equations. INGen provides the reaction network along with each reaction's family. A function was developed for KME wherein the reactions and families can be directly imported from INGen as shown in Figure 3.13. For the ARM approach, this communication must be executed for each attribute type.

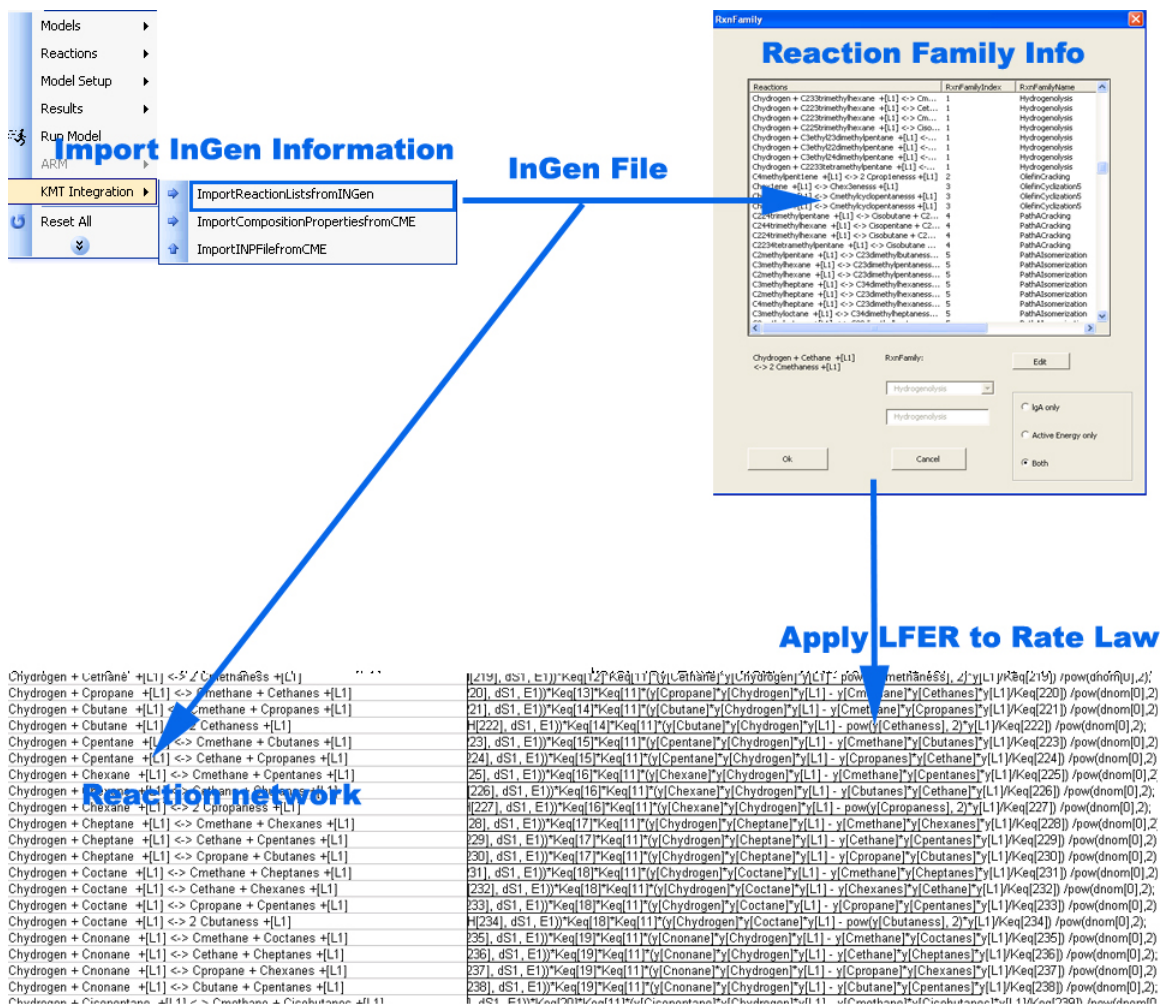


Figure 3.13 Importing an InGen File to KME

After showing how reaction families and LFERs can reduce the number of kinetic rate parameters and thus the complexity of the model, the next section adds to the complexity by accounting for heterogeneous systems by the inclusion of LHHW rate laws.

3.4 LHHW

Most complex process chemistries involve a catalyst and are thus a heterogeneous reaction system. To date, the most popular and successful way to describe the kinetics of heterogeneous catalysis is through the Langmuir-Hinshelwood-Hougen-Watson (LHHW) formalism. The general formula is shown as Eq 3.8.

$$r = \frac{(Kineticgroup)(Drivingforcegroup)}{(Adsorptiongroup)^n} \quad (3.8)$$

For the sake of complete reference of kinetic modeling in this work, we have also quoted the LHHW formalism here in Table 3.1.

Table 3.1 LHHW formalism (Yang and Hougen, 1950)¹⁵

Kinetic Groups				
Adsorption of A controlling	k_A			
Adsorption of B controlling	k_B			
Desorption of R controlling	$k_R K$			
Adsorption of A controlling with dissociation	k_A			
Impact of A controlling	$k_A K_B$			
Homogeneous reaction controlling	k			
	Surface Reaction Controlling			
	$A \leftrightarrow R$	$A \leftrightarrow R + S$	$A + B \leftrightarrow R$	$A + B \leftrightarrow R + S$
Without dissociation	$k_{sr} K_A$	$k_{sr} K_A$	$k_{sr} K_A K_B$	$k_{sr} K_A K_B$
With dissociation of A	$k_{sr} K_A$	$k_{sr} K_A$	$k_{sr} K_A K_B$	$k_{sr} K_A K_B$
B not adsorbed	$k_{sr} K_A$	$k_{sr} K_A$	$k_{sr} K_A$	$k_{sr} K_A$
B not adsorbed, A dissociated	$k_{sr} K_A$	$k_{sr} K_A$	$k_{sr} K_A$	$k_{sr} K_A$

Driving-Force Groups				
Reaction	$A \leftrightarrow R$	$A \leftrightarrow R + S$	$A + B \leftrightarrow R$	$A + B \leftrightarrow R + S$
Adsorption of A controlling	$p_A - \frac{p_R}{K}$	$p_A - \frac{p_R p_S}{K}$	$p_A - \frac{p_R}{K p_S}$	$p_A - \frac{p_R p_S}{K p_S}$
Adsorption of B controlling	0	0	$p_S - \frac{p_R}{K p_A}$	$p_S - \frac{p_R p_S}{K p_A}$
Desorption of R controlling	$p_A - \frac{p_R}{K}$	$\frac{p_A}{p_S} - \frac{p_R}{K}$	$p_A p_S - \frac{p_R}{K}$	$\frac{p_A p_S}{p_S} - \frac{p_R}{K}$
Surface reaction controlling	$p_A - \frac{p_R}{K}$	$p_A - \frac{p_R p_S}{K}$	$p_A p_S - \frac{p_R}{K}$	$p_A p_S - \frac{p_R p_S}{K}$
Impact of A controlling (A not adsorbed)	0	0	$p_A p_S - \frac{p_R}{K}$	$p_A p_S - \frac{p_R p_S}{K}$
Homogeneous reaction controlling	$p_A - \frac{p_R}{K}$	$p_A - \frac{p_R p_S}{K}$	$p_A p_S - \frac{p_R}{K}$	$p_A p_S - \frac{p_R p_S}{K}$

Table 3.1 LHHW formalism (Yang and Hougen, 1950) (Continued)

Replacements in the General Adsorption Groups				
$(1 + K_A p_A + K_B p_B + K_R p_R + K_S p_S + K_I p_I)^n$				
Reaction	$A \leftrightarrow R$	$A \leftrightarrow R + S$	$A + B \leftrightarrow R$	$A + B \leftrightarrow R + S$
Where adsorption of A is rate controlling, replace $K_A p_A$ by	$\frac{K_A p_R}{K}$	$\frac{K_A p_R p_S}{K}$	$\frac{K_A p_R}{K p_B}$	$\frac{K_A p_R p_S}{K p_B}$
Where adsorption of B is rate controlling, replace $K_B p_B$ by	0	0	$\frac{K_B p_R}{K p_A}$	$\frac{K_B p_R p_S}{K p_A}$
Where dsorption of R is rate controlling, replace $K_R p_R$ by	$KK_R p_A$	$KK_R \frac{p_A}{p_S}$	$KK_R p_A p_B$	$KK_R \frac{p_A p_B}{p_S}$
Where adsorption of A is rate controlling with dissiciation of A, replace $K_A p_A$	$\sqrt{\frac{K_A p_R}{K}}$	$\sqrt{\frac{K_A p_R p_S}{K}}$	$\sqrt{\frac{K_A p_R}{K p_B}}$	$\sqrt{\frac{K_A p_R p_S}{K p_B}}$
Where equilibrium adsorption of A takes place with dissociation of A, replace $K_A p_A$ by (and similarly forother components adsorbed with dissociation)	$\sqrt{K_A p_A}$	$\sqrt{K_A p_A}$	$\sqrt{K_A p_A}$	$\sqrt{K_A p_A}$
Where A is not adsorbed, replace $K_A p_A$ by (and similarly for other components that are not adsorbed)	0	0	0	0
Exponents of Adsorption Groups				
Adsorption of A controlling without dissociation	n=1			
Desorption of R controlling	n=1			
Adsorption of A controlling with dissociation	n=2			
Impact of A without dissociation $A+B \leftrightarrow R$	n=1			
Impact of A without dissociation $A+B \leftrightarrow R + S$	n=2			
Homogeneous reaction	n=0			
Surface Reaction Controlling				
	$A \leftrightarrow R$	$A \leftrightarrow R + S$	$A + B \leftrightarrow R$	$A + B \leftrightarrow R + S$
No dissociation of A	1	2	2	2
Dissociation of A	2	2	3	3
Dissociation of A (B not adsorbed)	2	2	2	2
No dissociation of A (B not adsorbed)	1	2	1	2

This thesis developed a flexible approach to apply LHHW in KME. For example, consider the following reaction involved with catalyst site [L]. KME's automation of LHHW generates the following rate equation (E.q. 3.9) for the surface limit case.

$$r = \frac{k_{SR} K_A K_B [L]_O (A \cdot B - C \cdot D / K)}{(1 + K_A A + K_B B + K_C C + K_D D)^2} \quad (3.9)$$

Where, K_i (i refer to A, B, C, and D) is the adsorption equilibrium constant for each reactants on L, K is the overall equilibrium constant of the reaction; and k_{SR} is the rate constant of the surface reaction.

As shown in Figure 3.14, KME allows a user to enter the pathway reactions of a specific chemistry in an excel sheet called "LHHW Reactions". Then, KME will ask the modeler to determine the catalyst site each reaction occurs upon. Finally, KME will automatically generate the LHHW rate expression for each reaction with catalyst site [Li], where [Li] is the indexed catalyst site in the model. KME will use the variable "dnom" to indicate the denominator of the equation and K_{eq} to indicate the adsorption equilibrium constants for each species. The driving force is assumed to be surface controlled automatically, but it is easy to adjust the denominator for other limiting cases by editing the dnom expression. In addition, K_{eq} can be edited for certain correlations analogous to the reaction family and LFER concepts shown in the previous section.

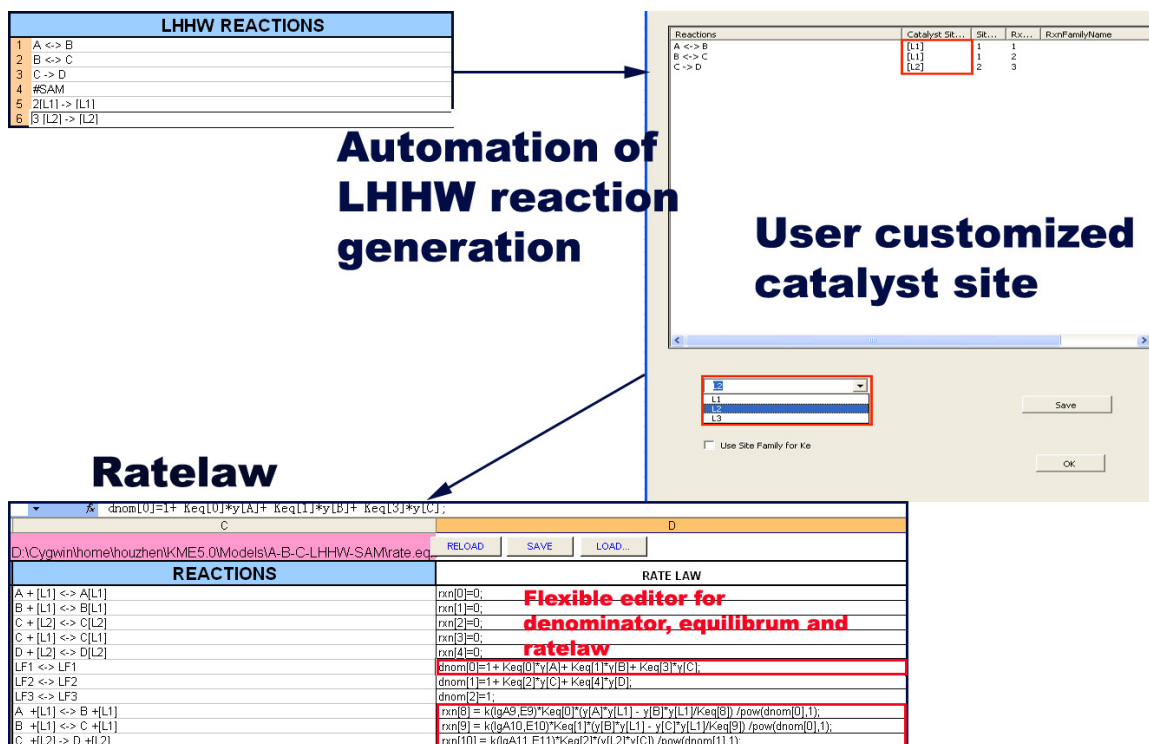


Figure 3.14 Automation of LHHW in KME

The combination of the reaction family and LFER concepts with LHHW can provide a practical methodology for providing the kinetic rate organization and evaluation for complex heterogeneous process chemistries.

3.5 Deactivation calibration

Most complex catalytic process chemistries experience the deactivation of the catalyst. Often the data acquired from those chemistries are not sampled from a steadystate, but from a deactivation period. Thus a detailed molecular kinetic model should consider such situations. A deactivation or catalyst coking reaction can be described by a selectivity activity modifier (SAM) reaction, $r_c = \sum k_i c_i$. Such reactions express how the chemistry of the system can change the catalyst. Coke deposition is a function of both the Time on Stream (TOS) and the local composition of the mixtures in the reactor. Typically, in a complex process, the characteristic reaction time of coke deposition or catalyst deactivation (units in

months or years) is significantly larger than that of the main reaction network (units in seconds). Due to this significant difference in the two time scales, a decoupled model can be used to validly describe the coke formation in order to avoid the numerical issues associated with solving PDE equations. During a catalyst deactivation model calculation, the total time on stream (TOS) is evenly divided into small time zones. During one simulation iteration, at a certain TOS point, the coke deposition rate is considered constant, and the concentration profile along the reactor is evaluated. At the next iteration of TOS, the coke deposition rate is updated by the SAM model. The coke level can be calculated via Eq. 3.10 and the species' concentration within reactor is updated. Thus, the coke levels can be updated sequentially along the TOS. At the end of all the iterations along the time on stream, the model can generate a three dimensional coke profile along both the TOS and the reactor length.

$$Coke(TOS1) = Coke(TOS0) + r_c(TOS0) * (TOS1 - TOS0) \quad (3.10)$$

The observed data of deactivation can be classified into two cases. In the first case, the data are comprised of the specific deactivation profile of species during the TOS. The conceptual flow is shown as Figure 3.15. The observed data and the weight for tuning are set up first. Then an optimization loop is applied. During each iteration, the catalyst site and the observed/weight data are updated from the deactivation measurement profile. KME then retrieves the data from inp, cond, obs and weight sheets to evaluate the model along the reactor length at this TOS point. The coke profile is then updated after the reactor simulation. The optimization process continues as KME enters into the next iteration. By minimizing the objection function for the deactivation profile in this optimization loop, KME will calculate an optimal set of parameters for this deactivation model.

Once Through Modeling Modes

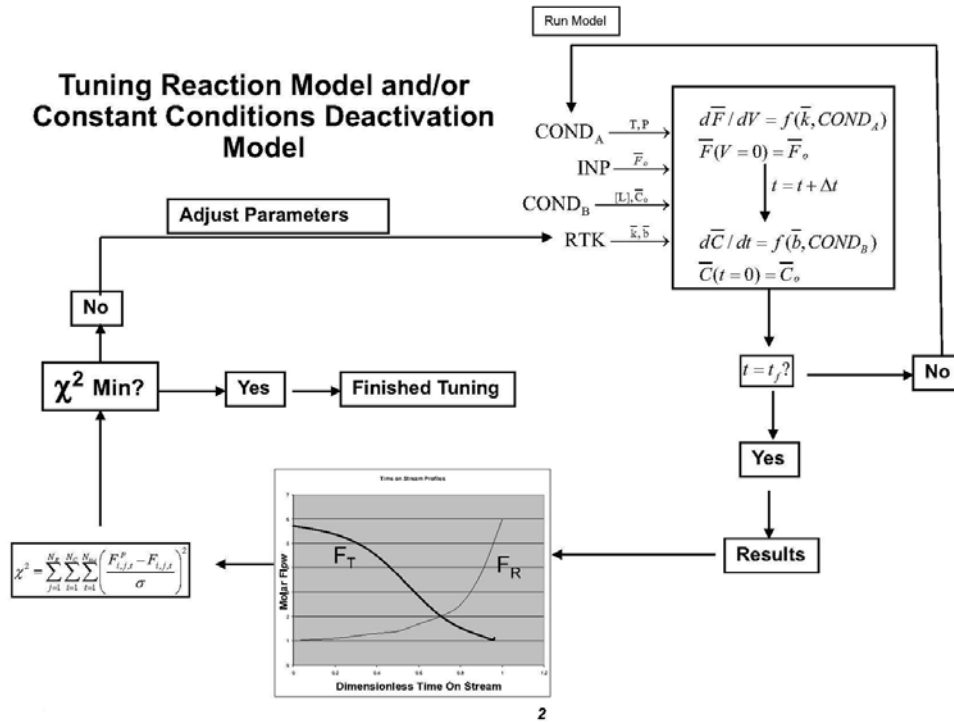


Figure 3.15 the conceptual flow of deactivation in onethrough model

However, industrial processes will generally not allow an unsatisfactory conversion due to catalyst deactivation. In combating such a situation, the operation conditions (such as reaction temperature) will be changed in order to achieve the target conversion. In such a case, the observed deactivation data is often a set of operating temperatures over the TOS, along with an optimal product yield, conversion or some other properties. The conceptual flow of this case is illustrated in Figure 3.16.

After inputting the observed temperature profile of deactivation, KME employs an optimization loop. For each iteration, KME updates the catalyst site and observed goalseeking temperature at this TOS point. KME will receive all necessary data and evaluate a goalseeking optimization to get a targeted conversion at this TOS point. The coke profile is updated after this optimization. Then KME entered to

Table 3.2 Multiple reactors in proposed KME

Reactor Type	Governing equation ($\frac{dF_A}{dV} = -r_A$)
CSTR	$r_A = \frac{C_{A0} - C_A}{t}$
Batch	$\frac{dC_A}{dt} = -r_A$
PFR	$\frac{dC_A}{d\tau} = -r_A$ for incompressible fluid.
RFR	$\frac{dC_A}{d\tau} = -r_A$ for incompressible fluid.
Semi-Batch	$\frac{dC_A}{dt} = (-r_A) - \frac{C_A \times q_{B0}}{V}$
Sidestream reactor	$\frac{dC_A}{d\tau} = (-r_A) - \frac{C_A \times q_{B0}}{V_p}$ for incompressible fluid.

3.7 Basic product separation

The product data used in detailed kinetic modeling of industrial processes is not often an effluent from the reactor, but rather a set of product characterizations after a post-separation process (e.g, gasoline, kerosene, diesel and bottom), as is favored in practical industries. It is necessary to consider the complexity of process configurations due to the real fractionators' operation. However, a rigorous fractionators' model involves multiple complex factors which lead to increased computational burden. Therefore, it is necessary to balance the complex calculation of this separation and the detailed kinetic modeling. As a result, a basic separation model calculation is employed in this thesis instead of a rigorous fractionator model.

The concept of this basic separation model¹⁶ is derived from an approximated linear relationship between the logarithm of the ratio of a component in distillate to bottoms with its boiling point temperature (Figure 3.17). So, based on an effective cut point and the component's normal boiling point (provided by a structure property correlation module), the split fraction of a component can be

calculated. By using this method for all components in the model, a set of separation fractions of each component can be obtained. In Figure 3.17, the model interpolates a value of $\ln(D_i/B_i)$ for a given TB_i . Usually, the model is set with three points representing the two straight lines around the effective cut point where $\ln(D_i/B_i)$ is zero (meaning a 50% split of the component between top and bottom streams). However, to protect from numerical singularities in the model, two more points are defined at the border line.

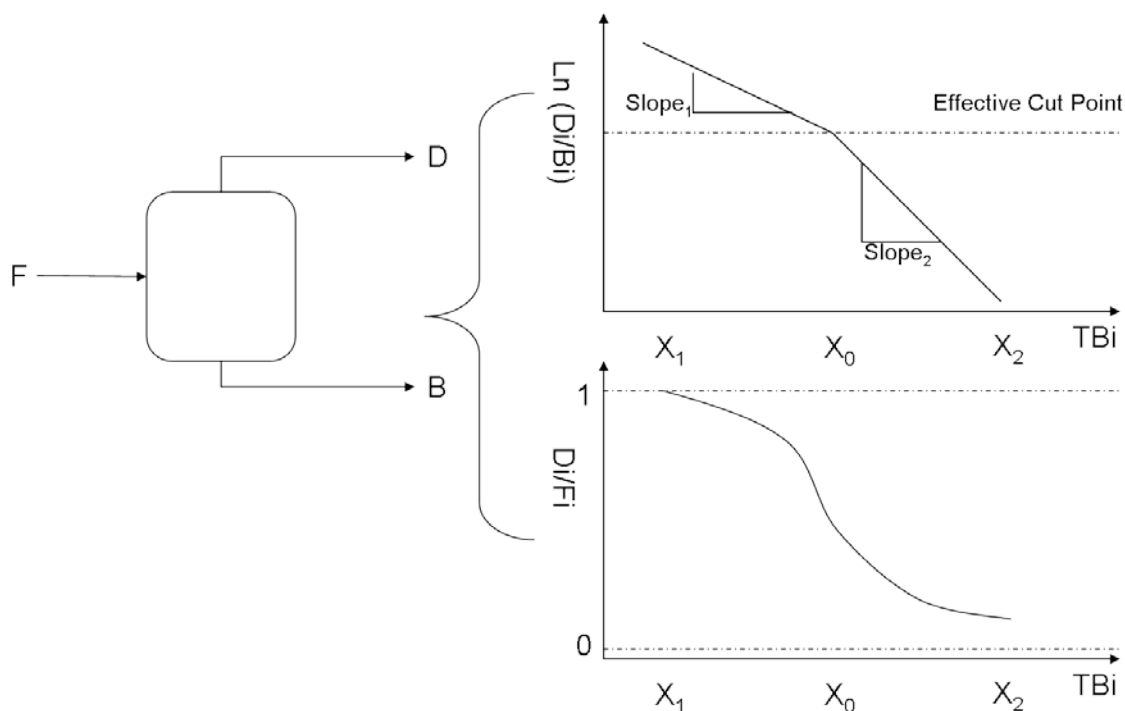


Figure 3.17 Principles of basic separation model

3.8 Model analysis

To evaluate the fitness of the detailed kinetic modeling, a statistical package was implemented in KME. Now, statistical analysis parameters such as stand deviation, Q test, confidence limits, etc. can be easily obtained. KME can evaluate the sensitivity analysis of each reaction in a model by calculating the differences when a reaction is turned on or off. This simple approach can be used to evaluate the sensitivity of kinetics parameters, and as a result, model reduction can be implemented.

3.8.3 Statistical analysis

When a model is calibrated, the statistical analysis of the parameter estimation can show the fitness of the model and thus give a summary result that shows whether the model is well tuned. In this thesis, a set of critical statistical items are evaluated in KME numerically such as: the total sum of squares of variance, the standard deviation of each parameter, the confidence limit of each parameter and so on.

The total sum of squares of variance can be evaluated as E.q. 3.11

$$S.S = \sum_{i=1}^p \sum_{j=1}^n (yCalc_{ij} - yObs_{ij}) \quad (3.11)$$

$yCalc_{ij}$ and $yExp_{ij}$ are the model calculated values and observed experiment data.

j is the index of experiment and n is the total number of experiments in this model.

i is the component index and p is the total number of components in this model.

To calculate the confidence limit and standard deviation of each parameter, the three-dimensional Jacobian matrix of the model must be calculated for every parameter.

A Jacobian matrix is the partial derivatives of predicted value vector (Y) with respect to the parameter vector (b) evaluated at all n experiment points. This matrix is three dimensional as there are n experiments by v parameters by p components. Each element can be represented as E.q.3.12:

$$J_{ikj} = \frac{\partial Y_{ij}}{\partial b_k} = \frac{(Y_{ij} |_{b_k + \Delta b_k} - Y_{ij} |_{b_k})}{\Delta b_k} \quad (3.12)$$

i- Experiment index

j- Component index

k- Parameter index

After obtaining the Jacobian matrix of the parameters in the model, the next step is to calculate the var-cov matrix. The first step is to define a set of sub matrices that encompass the whole three dimension Jacobian matrix of the model (E.q. 3.13):

For each component, JJ_j represents the two dimensional matrix of n experiments by p component numbers.

$$JJ_j = \begin{bmatrix} \frac{\partial Y_1}{\partial b_1} & \dots & \frac{\partial Y_1}{\partial b_v} \\ \dots & \dots & \dots \\ \frac{\partial Y_n}{\partial b_1} & \dots & \frac{\partial Y_n}{\partial b_v} \end{bmatrix} \quad (3.13)$$

Next, the var-cov matrix A is defined as E.q.3.14:

$$A = \left(\sum_{j=1}^p w_j^2 \times JJ_j^T \times JJ_j \right)^{-1} \quad (3.14)$$

w_j is the weight factor of this matrix, which can be evaluated by E.q.3.15

$$w_j = \frac{1/\delta_j^2}{\frac{1}{\sum_{k=1}^v n_i} \left(\sum_{k=1}^v \sum_{i=1}^{n_i} \delta_{ki} \right)} \quad (3.15)$$

The squared standard deviation of the parameter can be attained as E.q.3.16:

$$V(b_k) = \delta^2 a_{kk} \quad (3.16)$$

$a_{kk} - \text{diag log elements of var-cov matrix } A$

δ is the overall standard deviation of the whole model, which can be estimated as given in E.q.3.17, where DOF is the degree of freedom of the model.

$$s = \frac{\sum_{j=1}^p \sum_{i=1}^n \left(\frac{yCALC_{ij} - yOBS_{ij}}{w_j} \right)^2}{DOF} \quad (3.17)$$

Using a t distribution with $n \times p - v$ degrees as E.q.3.18:

$$t = \frac{b_k - \beta_k}{s \sqrt{a_{kk}}} \sim t(n \times p - v) \quad (3.18)$$

b_k is the optimized parameter value.

β_k is the real parameter value.

So the confidence limits can be attained by E.q.3.19:

$$b_k - t_{(1-\alpha/2, n \times p - k)} s \sqrt{a_{kk}} < \beta_k < b_k + t_{(1-\alpha/2, n \times p - k)} s \sqrt{a_{kk}} \quad (3.19)$$

3.8.4 Sensitivity analysis and model reduction

ARM as described in the previous section is a model reduction approach derived from insightful chemical understanding. Mathematical reduction strategies can also be exploited based on the sensitivity analysis of the tuned model.

To understand these methods, it is important to differentiate the types of the species in the model: important, necessary, and redundant. Important species are those species of direct importance to the user. Necessary species are those species which the model must retain in order to correctly model the important species. Redundant species are those that can be eliminated from the model without largely affecting the prediction of the important species. The user must define the set of important species and an acceptable level of error in those species. The model reduction algorithms then determine the necessary and redundant species and thus provide the information necessary in reducing the model reaction network.

Turanyi¹⁷ proposed an iterative scheme involving the sensitivity analysis on the rate of production of important species with respect to concentrations of all other species as a method of model reduction. He defined the vector B to have elements shown in E.q. 3.20

$$B_i = \sum_{n=1}^N \left(\frac{\partial \ln r_n}{\partial \ln c_i} \right)^2 \quad (3.20)$$

Where, r_n is the reaction sets involved in B_i and c_i is the concentration of species involved in B_i . The sum was carried out over the set of important and necessary species. Species possessing the highest B_i values were included in the set of necessary species, and the process was repeated until the vector B converged. Redundant species are those not considered necessary or important after convergence occurs. A

generalized species-based algorithm for model reduction building upon the above method has been developed. It is quoted as following description:

1. Start with an empty set of necessary species. Select the set of important species, a set of times of interest, and a fractional cutoff (between 0 and 1) to be used later. Solve the full model for the concentrations of each species at each of time of interest.
2. Calculate the sensitivity of the rates of production of important and necessary species to changes in the concentration of all other species in the reaction network. Call this measure of error Φ_j for species j.
3. Determine the range of Φ_j and compute a cutoff value Φ_c that is the fractional cutoff (as selected in step 1) of the Φ_j range. Add all species with $\Phi_j > \Phi_c$ to the necessary species set.
4. Repeat step 2 and 3 until no new necessary species is discovered.
5. Construct the reduced model using all reactions that consume the important and necessary species.

Similarly, a reaction-based model reduction algorithm is also developed and quoted as follows:

1. Start with an empty set of necessary species. Select the set of important species, a set of times of interest, and a fractional cutoff (between 0 and 1) to be used later. Solve the full model for the concentrations of each species at each of time of interest.
2. Calculate the sensitivity of the rates of production of important and necessary species to changes in the rate parameters of each reaction. Call this measure of error Φ_j for reaction j.
3. Determine the range of Φ_j and compute a cutoff value Φ_c that is the fractional cutoff (as selected in step 1) of the Φ_j range. Add all species that are consumed by reactions with $\Phi_j > \Phi_c$ to the necessary species set.
4. Repeat step 2 and 3 until no new necessary species is discovered.
5. Construct the reduced model using all reactions that have a Φ_j value greater than a given cutoff.

Different species-based or reaction-based model reduction techniques can be applied by defining an appropriate function to determine the sensitivity of the important and necessary species to each of the other species or reactions as required in step 2 of both algorithms. Two such functions are sensitivity analysis and the "on/off" method. Using sensitivity analysis, the function for quantifying the effect one reaction has on the set of important and necessary species can be shown in E.q.3.21:

$$\Phi_j = \sum_1^{\text{times species}} \left(\frac{\partial r_i}{\partial k_j} \right)^2 \quad (3.21)$$

where the species in the inner sum are the members of important and necessary species. Thus, Φ_j carries a measure of the instantaneous sensitivity of the rate of production of the important and necessary species on the kinetic parameter of reaction_j at a specified set of times. The "on/off" method sets the kinetic parameter or concentration being analyzed to zero and examines the effect this change has on the predictions made by the model. The function to be analyzed is in E.q.3.22:

$$\Phi_j = \sum_1^{\text{times species}} (r_i^* - r_i)^2 \quad (3.22)$$

where a superscript * represents the reduced model and the sum over the species is carried over the important and necessary ones only. Both the species-based and reaction-based cases use the concentration profiles of the full model as a basis when computing r_i^* so the reduced model does not have to be solved.

Families of reduced models can thus be created using this methodology by changing the fractional cutoff only. Generally, since the implications of a particular choice of the cutoff fraction are not apparent a priori, a range of cutoff fractions should be explored to locate the optimal one. A plateau in overall error is a general feature of a graph of overall error vs. number of species or number of reactions in the reduced model, and the smallest model on this plateau marks an optimum reduced model if no target accuracy is available.

Baynes (1998)¹⁸ has applied these algorithms on an n-hexane acid cracking mechanistic model with 597 species and 2932 reactions and reduced the model to 390 species and 1342 reactions, which correctly models the concentrations of the specified important species and only takes 33% of the time to solve compared with the full one.

In this thesis, the automation algorithm in KME is adapted from the reaction-based model reduction method rather than the species-based one. If an important species undergoes both fast and slow reactions, reaction-based reduction can remove the slower reactions, but species-based reduction cannot. The trade-off is that the reaction-based model reduction takes longer to complete because it analyzes more possibilities. In summary, the sensitivity analysis in KME is performed as the “on/off” method demonstrated above, after which KME can provide a model reduction strategy.

This algorithm is very useful for the detailed kinetic modeling of a complex system. After tuning and the user’s selection of the important species in the system, KME can build a simplified mathematical model containing much less computational burden than the full model. This benefit comes at the low cost of removing species that are not of interest in the model. Such a reduced model is often required for applications which solution time is critical such as: advanced control, real time optimization. In addition, the “On/off” method provides a great help for reconciling complex models. For models containing thousand of species and equations, it can give an opportunity to divide the model into several sub-blocks which can be studied individually. A comprehensive understanding of the full complex model can be obtained from the understanding of each block.

3.9 Summary and discussion

This chapter focused on the methodologies of handling the reaction equation generation and solution in the development of detailed kinetic models for the complex system.

To address the conflict between the complexity of species and equations and the impractical computational burden for a complex process chemistry, a hybrid statistical approach was developed in this thesis (ARM). The essential part of ARM is that it will turn a multiple dimensional equations to N one dimensional equations. This method can turn tens of thousands of ODE equations into a more computationally practical set of ODE equations while maintaining the full molecular details of the model.

Besides the ARM approach for the reduction of the computational complexity of a detailed model, other the methodologies for reducing the number of parameters were addressed. The LFER concept of organizing kinetic data was reviewed from classical transition state theory and the concept of reaction families was introduced along the way. Some representative results developed by the Klein group were shown in the thesis. Based on the accuracy of results, an automated algorithm was developed in KME to employ the LFER and reaction family concepts via the Evans-Polanyi principle. The information of reactions and reaction families can be imported into KME from INGen via an automated algorithm.

To deal with the heterogeneous systems, e.g., all the catalyzed processes, it was determined that the LHHW formalism is still the best rate law in most of the cases. The LHHW formalism has been extended to handle multi-species and multi-site cases found in the complex process chemistries. Automation of LHHW rate laws was developed for KME. Surface reaction control was chosen as the default, but it is easy to alter to other rate-determining steps using the KME LHHW tool.

The complexity from actual process configurations was also considered in this thesis. A two fold deactivation process model was developed for KME in this thesis, and it can satisfy the simulation and tuning of both laboratory level and industrial plant data. In addition, the balance equations for multiple reactor types were integrated into KME as a product of this thesis. Finally, a basic separation model was developed for KME to supplement practical industrial demand.

The final part of this chapter concentrated on using analytical methods to provide auxiliary tools for model tuning in KME. A statistical package was developed, and model sensitivity analysis and reduction algorithm were addressed.

The above aspects were used to solve the dilemma of the complexity of real feedstocks and process chemistries as well as the complexity of the configurations and the data obtained from practical processes. KME now provides a more robust user friendly environment for model tuning and solution. Unlike the initial version of KME mentioned in Chapter1, the new KME enhanced by the contributions listed in this chapter, shows strong capability to handle modeling the complex systems found in the real word.

Chapter2 and this chapter discussed how to transform the analytical information of feedstocks to digital molecular information and the techniques used to transform that information to a kinetic model and its solution. In those discussions, the acquisition of property data proved a critical part for real world applications. The next chapter will address how to obtain property data based on structure correlations.

Reference

- ¹ Craig, Bennett, "User-controlled kinetic network generation with InGen", Doctoral Dissertation, Rutgers University(2009)
- ² Freudenberg, K. and Neish, A.C "Constitution and Biosynthesis of Lignin", Springer-Verlag, New York (1968)
- ³ Hammett, L.P., J. Am. Chem. Soc. 59, 96 (1937)
- ⁴ Evans, M.G. and Polanyi, M. *Inertia and Driving Force of Chemical Reactions* (1937)
- ⁵ Mochida, I. and Yoneda, Y. "Linear Free Energy Relationships in Heterogeneous Catalysis I. Dealkylation of Alkylbenzenes on Cracking Catalysts.," *J. Catal.*, 7, 386-392. (1967a)
- ⁶ Mochida, I. and Yoneda, Y. "Linear Free Energy Relationships in Heterogeneous Catalysis II. Dealkylation and Isomerization Reactions on Various Solid Acid Catalysts," *J. Catal.*, 7, 393-396. (1967b)
- ⁷ Mochida, I. and Yoneda, Y. "Linear Free Energy Relationships in Heterogeneous Catalysis III. Temperature Effects in Dealkylation of Alkylbenzenes on the Cracking Catalysts," *J. Catal.*, 8, 223-230. (1967c)
- ⁸ Neurock, M. and Klein, M.T. "Linear Free Energy Relationships in Kinetic Analyses: Applications of Quantum Chemistry," *Polyc. Arom. Compounds*, 3, 231-246.(1993a)
- ⁹ Neurock, M. and Klein, M.T. "When you can't measure— model," *ChemTech*, 23, 26-32. (1993b)
- ¹⁰ Neurock, M.T. *A Computational Chemical Reaction Engineering Analysis of Complex Heavy Hydrocarbon Reaction Systems*, Ph.D. Thesis, University of Delaware (1992)
- ¹¹ Korre, S. "Quantitative Structure/Reactivity Correlations as a Reaction Engineering Tool: Applications to Hydrocracking of Polynuclear Aromatics," Ph.D. Thesis, University of Delaware (1995)
- ¹² Broadbelt, L. J., Stark, S.M., and Klein, M. T., *Ind. And Eng. Chem. Res.*, 33, pp. 790-799 (1994)
- ¹³ La Vopa, V., Satterfield, C.N. *J. Catal.* 110, 375. (1988)
- ¹⁴ Gang, Hou, Integrated chemical engineering tools for the building, solution, and delivery of detailed kinetic models and their industrial applications, Doctoral Dissertation, University of Delaware (2001)
- ¹⁵ Froment, G.F., Bischoff, K.B., "Chemical Reactor Analysis and Design"
- ¹⁶ AspenTech Corp.
- ¹⁷ Turanyi, T., "Reduction of Large Reaction Mechanisms", *New J. Chem.*, 14, 795-803.(1990)
- ¹⁸ Baynes, B., BS Thesis, University of Delaware (1997)

Chapter 4 Structure property correlation module

4.1 Introduction

The third proposed contribution towards this thesis is the provision of properties estimation supporting the detailed kinetic modeling of complex process chemistries. As stated in Chapter 2 and Chapter 3, both composition and reaction modeling require the calculation of intermediate and end-use properties. The role of the structure property correlation module (SPCM), as an integrated approach of this thesis, is illustrated in Figure 4.1.

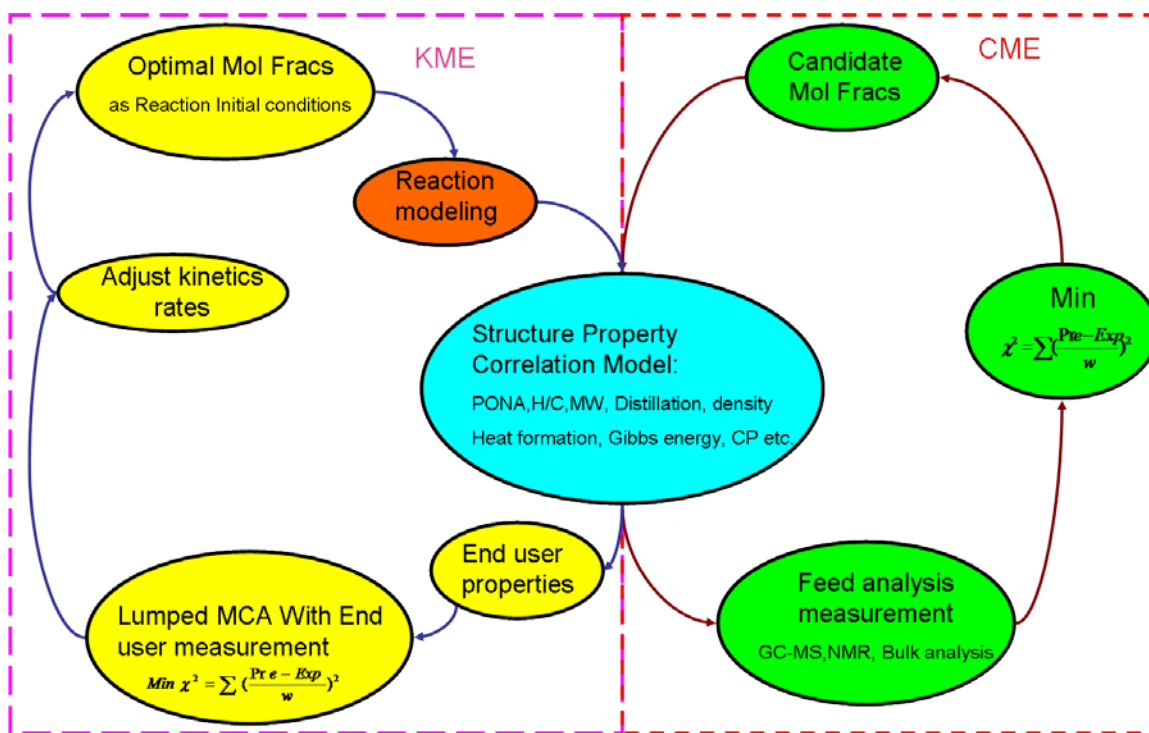


Figure 4.1 The role of structure property correlation in KMT

For most heavy complex feedstocks, CME will use an optimization loop to get the optimal MCA0. SPCM will evaluate the properties corresponding to the analytical measurements and determine the value of objective function during optimization. During reaction kinetics modeling, KME will evaluate

the reaction activity properties (such as enthalpy, and entropy) by structure property correlation and calculate the end user properties (mixed density, simulated distillation) from the MCA of the products. Since the molecules in the MCA0/MCA are already represented by BE matrices. SPCM will use those BE matrices as the basis for the 2D structural information for each composition in the feedstock. Thus, SPCM acts a fundamental data pool in facilitating the two previous contributions in the detailed kinetics.

The calculation of pure component properties is one important part in SPCM. It provides the basis to calculate the bulk properties which are used by CME as the objective function for quantitative information optimization and by KME as the commercially relevant properties of the final products. In addition, the structural and thermodynamic properties of each component are employed as reaction indices for LFER parameter estimation. Therefore, the evaluation of pure component properties will be elaborated upon in this chapter.

As a supplemental management tool, a pure component property database has been developed for this thesis. Such a database not only allows a user to correct data, but also provides a convenient way to manage and facilitate the use of properties for a huge number of species.

Based on the pure component property estimation, an algorithm for the mixed property estimation was developed based on customizable mixing rules. In addition, the calculation of the thermodynamic properties of reaction is addressed and will be illustrated in the last part of this chapter.

4.2 Pure component property estimation

The property estimation of a pure component is classified into two aspects: structural properties and thermodynamic properties. The former is evaluated explicitly from a BE matrix. The later is obtained from group contribution methods and semi-empirical quantum chemistry software.

4.2.1 Structural properties' estimation

Since a BE matrix can provide the full 2D structure of a molecule, SPCM will evaluate structural properties in a straightforward process. The structural properties provided by SPCM are shown in Table 4.1 and can be used to calculate bulk properties based on a mixing rule.

Table 4.1 Pure component structural properties caculated in SPCM

Molecular Weight	Total Naphthenic Ring Number
Total Hydrogen Number	Five Membered Naphthenic Ring Number
Total Carbon Number	Six Membered Naphthenic Ring Number
Total Sulfur Number	Total Aromatic Ring Number
Total Nitrogen Number	Thiophenic Ring Number
Aromatic Hydrogen Number	The Carbon Number on Aromatic Ring
Alpha Hydrogen Number	The Carbon Number on Naphthenic Ring
The Number of Side Chains	

However, there are crucial thermodynamic properties that can not be obtained directly from a BE matrix, but must be evaluated by empirical or semi-empirical methods. Those thermodynamic properties are very important to model a complex system. The physical properties (e.g., Boiling Point (BP), Density etc) are the basis to the bulk properties estimation and other thermodynamics properties (e.g. the heat of formation, Gibbs energy of formation, the entropy of formation, heat capacity etc.) can be used as reaction indecies in support of the reaction modeling using an LFER approach. An efficient method to calculate those thermodynamic properties is the general fragmental group contribution method.

4.2.2 Thermodynamic property estimation by group contribution

The general fragmental group contribution method is a classic methodology wherein the property of a compound is a function of its structural group parameters (as determined by summing the frequency of each group occurring in the molecule and multiplying by its contribution). This method provides the advantage of obtaining quick estimates without requiring substantial computational resources. Since a BE matrix can be scanned for any pre-defined structural group, SPCM can make the use of any existing

group contribution method. In this thesis, two group contribution methods were employed: Joback¹ and Gani².

4.2.2.1 Selected properties calculated by Joback method

The Joback and Reid¹ method identified 41 structural groups via the regression of 384 pure chemical compounds. These groups can then provide the set of thermodynamic properties listed in Table 4.2.

Table 4.2 Thermodynamics properties in Joback method

Critical Pressure	Heat Capacity
Critical Temperature	Standard Enthalpy of Formation at 298 K
Critical Volume	Standard Gibbs Energy of Formation at 298 K
Boiling Point	Standard Enthalpy of Fusion
Melting Point	Standard Enthalpy of Vaporization at 298 K
Viscosity	

To elaborate on how the Joback method was implemented in SPCM, Table 4.3 shows the details of the 41 structural groups. In this table, the rows indicate the functional groups and the columns represent the contributions towards each of the properties in Table 4.2 .

Table 4.3 The structural groups in Jack method¹.

Group	Tc	Pc	Vc	Tb	Tm	Hform	Gform	CPa	CPb	CPc	CPd	Hfusion	Hvap	Visa	Visb
-CH3	0.01	0.00	65.00	23.58	-5.10	-76.45	-43.96	19.50	-0.01	0.00	0.00	0.91	2.37	548.29	-1.72
-CH2-	0.02	0.00	56.00	22.88	11.27	-20.64	8.42	-0.91	0.10	0.00	0.00	2.59	2.23	94.16	-0.20
>CH-	0.02	0.00	41.00	21.74	12.64	29.89	58.36	-23.00	0.20	0.00	0.00	0.75	1.69	-322.15	1.19
>C<	0.01	0.00	27.00	18.25	46.43	82.23	116.02	-66.20	0.43	0.00	0.00	-1.46	0.64	-573.56	2.31
0	0.01	0.00	56.00	18.18	-4.32	-9.63	3.77	23.60	-0.04	0.00	0.00	-0.47	1.72	495.01	-1.54
=CH-	0.01	0.00	46.00	24.96	8.73	37.97	48.53	-8.00	0.11	0.00	0.00	2.69	2.21	82.28	-0.24
=C<	0.01	0.00	38.00	24.14	11.14	83.99	92.36	-28.10	0.21	0.00	0.00	3.06	2.14	NA	NA
=C=	0.00	0.00	36.00	26.15	17.78	142.14	136.70	27.40	-0.06	0.00	0.00	4.72	2.66	NA	NA
≡CH	0.00	0.00	46.00	9.20	-11.18	79.30	77.71	24.50	-0.03	0.00	0.00	2.32	1.16	NA	NA
≡C-	0.00	0.00	37.00	27.38	64.32	115.51	109.82	7.87	0.02	0.00	0.00	4.15	3.30	NA	NA
-CH2-	0.01	0.00	48.00	27.15	7.75	-26.80	-3.68	-6.03	0.09	0.00	0.00	0.49	2.40	307.53	-0.80
>CH-	0.01	0.00	38.00	21.78	19.88	8.67	40.99	-20.50	0.16	0.00	0.00	3.24	1.94	-394.29	1.25
>C<	0.00	0.01	27.00	21.32	60.15	79.72	87.88	-90.90	0.56	0.00	0.00	-1.37	0.64	0.00	0.00
=CH-	0.01	0.00	41.00	26.73	8.13	2.09	11.30	-2.14	0.06	0.00	0.00	1.10	2.54	259.65	-0.70
=C<	0.01	0.00	32.00	31.01	37.02	46.43	54.05	-8.25	0.10	0.00	0.00	2.39	3.06	-245.74	0.91
-F	0.01	-0.01	27.00	-0.03	-15.78	-251.92	-247.19	26.50	-0.09	0.00	0.00	1.40	-0.67	NA	NA
-Cl	0.01	0.00	58.00	38.13	13.55	-71.55	-64.31	33.30	-0.10	0.00	0.00	2.52	4.53	625.45	-1.81
-Br	0.01	0.01	71.00	66.86	43.43	-29.48	-38.06	28.60	-0.06	0.00	0.00	3.60	6.58	738.91	-2.04
-I	0.01	0.00	97.00	93.84	41.69	21.06	5.74	32.10	-0.06	0.00	0.00	2.72	9.52	809.55	-2.22
-OH(alcohol)	0.07	0.01	28.00	92.88	44.45	-208.04	-189.20	25.70	-0.07	0.00	0.00	2.41	16.83	2173.72	-5.06
-OH(phenol)	0.02	0.02	-25.00	76.34	82.83	-221.65	-197.37	-2.81	0.11	0.00	0.00	4.49	12.50	3018.17	-7.31
-O-(nonring)	0.02	0.00	18.00	22.42	22.23	-132.22	-105.00	25.50	-0.06	0.00	0.00	1.19	2.41	122.09	-0.39
-O-(ring)	0.01	0.00	13.00	31.22	23.05	-138.16	-98.22	12.20	-0.01	0.00	0.00	5.88	4.68	440.24	-0.95
>C=O(nonring)	0.04	0.00	62.00	76.75	61.20	-133.22	-120.50	6.45	0.07	0.00	0.00	4.19	8.97	340.35	-0.35
>C=O(ring)	0.03	0.00	55.00	94.97	75.97	-164.50	-126.27	30.40	-0.08	0.00	0.00	NA	6.65	NA	NA
O=CH-(aldehyde)	0.04	0.00	82.00	72.24	36.90	-162.03	-143.48	30.90	-0.03	0.00	0.00	3.20	9.09	740.92	-1.71
-COOH(acid)	0.08	0.01	89.00	169.09	155.50	-426.72	-387.87	24.10	0.04	0.00	0.00	11.05	19.54	1317.23	-2.58
-COO-(ester)	0.05	0.00	82.00	81.10	53.60	-337.92	-301.95	24.50	0.04	0.00	0.00	6.96	9.63	483.88	-0.97
O(other)	0.01	0.01	36.00	-10.50	2.08	-247.61	-250.83	6.82	0.02	0.00	0.00	3.62	5.91	675.24	-1.34
-NH2	0.02	0.01	38.00	73.23	66.89	-22.02	14.07	26.90	-0.04	0.00	0.00	3.52	10.79	NA	NA
>NH(non-ring)	0.03	0.01	35.00	50.17	52.66	53.47	89.39	-1.21	0.08	0.00	0.00	5.10	6.44	NA	NA
>NH(ring)	0.01	0.01	29.00	52.82	101.51	31.65	75.61	11.80	-0.02	0.00	0.00	7.49	6.93	NA	NA
>N-(nonring)	0.02	0.01	9.00	11.74	48.84	123.34	163.16	-31.10	0.23	0.00	0.00	4.70	1.90	NA	NA
-N=(nonring)	0.03	-0.01	NA	74.60	NA	23.61	NA	NA	NA	NA	NA	NA	3.34	NA	NA
-N=(ring)	0.01	0.01	34.00	57.55	68.40	93.70	119.66	5.69	0.00	0.00	0.00	3.65	6.53	NA	NA
=NH	NA	NA	NA	83.08	68.91	93.70	119.66	5.69	0.00	0.00	-16.88	0.00	12.17	NA	NA
=CN	0.05	-0.01	91.00	125.66	59.89	88.43	89.22	36.50	-0.07	0.00	0.00	2.41	12.85	NA	NA
-NO2	0.04	0.01	91.00	152.54	127.24	-66.57	-16.83	25.90	0.00	0.00	0.00	9.68	16.74	NA	NA
-SH	0.00	0.01	63.00	63.56	20.09	-17.33	-22.99	35.30	-0.08	0.00	0.00	2.36	6.88	NA	NA
-S-(nonring)	0.01	0.00	54.00	68.78	34.40	41.87	33.12	19.60	-0.01	0.00	0.00	4.13	6.82	NA	NA
-S-(ring)	0.00	0.01	38.00	52.10	79.93	39.10	27.76	16.70	0.00	0.00	0.00	1.56	5.98	NA	NA

The calculations of the properties in Table 4.2 can be illustrated in E.q.4.1¹.

$$\text{Boiling Point : } T_b (K) = 198.2 + \sum N_i \cdot T_{bi}$$

$$\text{Melting Point : } T_m (K) = 122.5 + \sum N_i \cdot T_{mi}$$

$$\text{Critical Temperature : } T_c (K) = \frac{T_b}{(0.584 + 0.965 \cdot \sum N_i \cdot T_{ci} - (\sum N_i \cdot T_{ci})^2)}$$

$$\text{Critical Pressure : } P_c (\text{Bar}) = \frac{1}{(0.113 + 0.0032 \cdot \text{TotalAtoms} - \sum N_i \cdot P_{ci})^2}$$

$$\text{Critical Volume : } V_c (\text{cm}^3 / \text{mol}) = 17.5 + \sum N_i \cdot V_{ci}$$

$$\text{Standard Enthalpy of Formation : } \Delta H_f^0 (KJ / \text{mol}) = 68.29 + \sum N_i \cdot H_{fi}$$

$$\text{Standard Gibbs energy of Formation : } \Delta G_f^0 (KJ / \text{mol}) = 53.88 + \sum N_i \cdot G_{fi}$$

$$\text{Standard Enthalpy of Vaporization : } \Delta H_{vap}^0 (KJ / \text{mol}) = 0.88 + \sum N_i \cdot H_{vapi}$$

$$\text{Standard Enthalpy of Fusion : } \Delta H_{fus}^0 (KJ / \text{mol}) = 15.30 + \sum N_i \cdot H_{fusi}$$

$$\text{Heat Capacity : } Cp (KJ / K \cdot \text{mol}) = Cp_a + Cp_b \cdot T + Cp_c \cdot T^2 + Cp_d \cdot T^3$$

$$Cp_a = \sum N_i \cdot C_{pai} - 37.93, Cp_b = \sum N_i \cdot C_{pbi} + 0.210$$

$$Cp_c = \sum N_i \cdot C_{pci} - 3.91 \times 10^{-4}, Cp_d = \sum N_i \cdot C_{pdi} + 2.06 \times 10^{-7}$$

$$\text{Viscosity : } \eta, \log(\eta/\text{MW}) = \frac{(\sum N_i \cdot Vis_{bi} - 597.82)}{T} + (\sum N_i \cdot Vis_{bi} - 4.294) \quad (4.1)$$

In E.q.4.1, N_i is the occurrence of the i th group in a specific component. The values T_{bi} is the property contribution of the i th group.

The Joback method was later expanded to 87 groups for boiling point prediction by Stein, who included various CH groups for the paraffin, naphthenic ring and aromatic parts of a component. The key factor in employing this method is the identification of the structural groups in Table 4.3 given the BE matrix of a component. A search algorithm was developed in this thesis to determine those groups and subsequently calculate the properties by E.q.4.1.

Unfortunately, the Joback method is limited to lighter complex feedstocks, and does not work well with heavy feedstocks. To get more accurate results for heavier complex feedstocks, more isomeric details, aggregate ring structures and the interactions between multiple groups needed to be considered.

4.2.2.2 Selected properties calculated by Rafiqul Gani method

Recently, Rafiqul Gani² and his coworkers developed a new approach for property estimation that used a total of 370 groups based on studying the data set of over 2000 pure compounds ranging from C3 to C60. Their method stressed the addition of isomers and multi-functional components and, thus, will be appropriate for heavy feedstocks. The properties provided by this method are listed in Table 4.4.

Table 4.4 Thermodynamics properties in Rafiqul Gani method

Critical Pressure	
Critical Temperature	Standard Enthalpy of Formation at 298 K
Critical Volume	Standard Gibbs Energy of Formation at 298 K
Boiling Point	Standard Enthalpy of Fusion
Melting Point	Standard Enthalpy of Vaporization at 298 K

In this new method, the molecular structure of a compound is considered to be a collection of three types of groups: first-order groups, second-order groups and third-order groups. The first-order groups (Table 4.5) are the basic building blocks intended to describe a wide variety of organic compounds. The second level groups (Table 4.6) permit a better description of polyfunctional compounds and differentiation among isomers. The third-order (Table 4.7) groups allow for detailed representations of systems of fused aromatic rings, systems of fused aromatic and nonaromatic rings, and systems of nonfused rings joined by chains of various other functional groups.

Table 4.5 The first order groups and their contributions in Gani method²

	Group	Example	$T_{m(l)}$	$T_{S(l)}$	$T_{C(l)}$	$P_{C(l)}$	$V_{C(l)}$	$G_{(l)}$	$H_{(l)}$	$H_{S(l)}$	$H_{m(l)}$
1	CH ₃	<i>n</i> -Tetracontane (2)	0.6953	0.8491	1.7506	0.018615	68.35	2.878	-42.479	0.217	1.660
2	CH ₂	<i>n</i> -Tetracontane (38)	0.2515	0.7141	1.3327	0.013547	56.28	8.064	-20.829	4.910	2.639
3	CH	2-Methylpentane (1)	-0.3730	0.2925	0.5960	0.007259	37.50	8.254	-7.122	7.962	0.134
4	C	2,2-Dimethylbutane (1)	0.0256	-0.0671	0.0306	0.001219	16.01	16.413	8.928	10.730	-1.232
5	CH ₂ =CH	1-Hexene (1)	1.1728	1.5596	3.2295	0.025745	111.43	95.738	57.509	4.031	1.268
6	CH=CH	2-Hexene (1)	0.9460	1.5597	3.0741	0.023003	98.43	92.656	69.664	9.456	4.441
7	CH ₂ =C	2-Methyl-1-butene (1)	0.7662	1.3621	2.7717	0.021137	91.40	85.107	61.625	8.602	2.451
8	CH=C	2-Methyl-2-butene (1)	0.1732	1.2971	2.5666	0.019609	83.89	88.691	81.835	14.095	3.032
9	C=C	2,3-Dimethyl-2-butene (1)	0.3928	1.2739	2.6391	0.014114	90.66	93.119	95.710	19.910	2.616
10	CH ₂ =C=CH	1,2-Butadiene (1)	1.7036	2.6840	5.4330	0.035483	143.57	229.906	198.840	11.310	7.076
11	CH ₂ =C=C	3-Methyl-1,2-butadiene (1)	1.5453	2.4014	4.8219	0.029678	146.36	226.710	208.490	*****	7.435
12	CH=C=CH	2,3-Pentadiene (1)	1.2850	2.5400	*****	*****	*****	*****	*****	*****	6.000
13	CH≡C	1-Pentyne (1)	2.2276	1.7618	3.7897	0.014010	84.60	230.029	224.902	6.144	-1.548
14	C≡C	3-Decyne (1)	2.0516	1.6767	4.5870	0.010888	74.66	216.013	228.282	12.540	6.128
15	aCH	Benzene (6)	0.5860	0.8365	2.0337	0.007260	42.39	26.732	12.861	3.683	1.948
16	aC fused with aromatic ring	Naphthalene (2)	1.8955	1.7324	5.4979	0.003564	35.71	20.379	20.187	6.631	0.845
17	aC fused with non-aromatic subring	Indane (2)	1.2065	1.1995	3.1058	0.006512	34.65	33.912	30.768	6.152	1.095
18	aC except as above	Benzophenone (1)	0.9176	1.5468	4.5344	0.012859	26.47	23.331	24.701	6.824	-0.531
19	aN in aromatic ring	Pyridine (1)	2.0438	1.3977	4.0954	-0.003339	36.47	89.902	70.862	9.420	2.555
20	aC-CH ₃	Toluene (1)	1.0068	1.5653	3.4611	0.020907	97.33	24.919	-19.258	8.279	2.969
21	aC-CH ₂	Ethylbenzene (1)	0.1065	1.4925	2.9003	0.018082	87.19	31.663	4.380	11.981	0.948
22	aC-CH	Cumene (1)	-0.5197	0.8665	1.9512	0.011795	73.51	30.393	18.440	13.519	-1.037
23	aC-C	<i>tert</i> -Butylbenzene (1)	-0.1041	0.5229	0.8576	0.011298	67.20	40.127	35.297	16.912	-2.856
24	aC-CH=CH ₂	Styrene (1)	1.2832	2.4308	5.7861	0.030637	134.69	114.531	77.863	*****	4.013
25	aC-CH=CH	1-Propenylbenzene (1)	1.7744	2.9262	6.5062	0.026282	128.84	111.216	88.084	*****	8.274
26	aC-C=CH ₂	α -Methylstyrene (1)	1.2612	2.1472	4.9967	0.026371	110.74	115.728	90.927	*****	3.324
27	aC-C=CH	Phenylacetylene (1)	1.7495	2.3057	6.4572	0.019507	112.08	263.205	257.448	*****	2.514
28	aC-C≡C	1-Phenyl-1-propyne (1)	*****	2.7341	*****	*****	*****	*****	*****	*****	*****
29	OH	1,4-Butanediol (2)	2.7888	2.5670	5.2188	-0.005401	30.61	-144.051	-178.360	24.214	4.786
30	aC-OH	Phenol (1)	5.1473	3.3205	9.3472	-0.008788	50.77	-131.327	-164.191	34.099	8.427
31	COOH	1,5-Pentanedioic acid (2)	7.4042	5.1108	14.6038	0.009885	90.66	-337.090	-389.931	17.002	10.692
32	aC-COOH	Benzoic acid (1)	12.4296	6.0677	15.4515	0.017100	119.10	-312.422	-361.249	*****	14.649
33	CH ₃ CO	2-Butanone (1)	2.9588	3.1178	7.0058	0.025227	127.99	-120.667	-180.604	15.195	8.062
34	CH ₂ CO	3-Pentanone (1)	2.5232	2.6761	5.7157	0.019619	112.79	-120.425	-163.090	19.392	8.826
35	CHCO	2,4-Dimethyl-3-pentanone (1)	1.1565	2.1748	4.4743	0.012487	97.16	-116.799	-139.909	20.350	7.205
36	CCO	2,2,4,4-Tetramethyl-3-pentanone (1)	1.0638	1.7287	*****	*****	*****	*****	*****	*****	*****
37	aC-CO	Acetophenone (1)	2.9157	3.4650	9.4806	0.011007	90.69	-91.812	-106.965	25.036	4.852
38	CHO	1-Hexanal (1)	3.0186	2.5388	5.8013	0.010204	71.08	-100.882	-130.816	12.370	11.325
39	aC-CHO	Benzaldehyde (1)	2.4744	3.5172	9.4795	0.019633	122.91	-80.222	-107.159	*****	7.273
40	CH ₂ COO	Butyl acetate (1)	2.1657	3.1228	6.3179	0.033812	148.91	-306.733	-387.458	19.342	7.910
41	CH ₂ COO	Methyl Butyrate (1)	1.6329	2.9850	5.9619	0.026983	132.89	-298.332	-364.204	21.100	9.479
42	CHCOO	Ethyl isobutyrate (1)	1.0668	2.2869	4.7558	0.021990	125.52	-301.414	-352.057	24.937	9.317
43	CCOO	Ethyl 2,2-dimethylpropionate (1)	0.3983	1.6918	*****	*****	*****	*****	*****	23.739	*****
44	HCOO	Propyl formate (1)	2.0223	2.5972	5.6064	0.015249	93.29	-276.878	-327.678	15.422	8.115

Group	Example	T_{mL}	T_{SL}	T_{CL}	P_{CL}	V_{CL}	G_{CL}	H_{CL}	H_{VL}	H_{BL}
94	aC-CONH ₂	Benzamide	12.8071	8.3775	*****	*****	*****	*****	*****	16.811
95	aC-NH(CO)H	N-phenylformamide (1)	5.6631	7.3497	19.8979	0.023447	162.08	-44.595	-125.052	8.658
96	aC-N(CO)H	N-methyl-N-phenylmethanamide (1)	3.3602	5.1373	*****	*****	*****	*****	*****	*****
97	aC-CONH	N-methylbenzamide (1)	6.5160	7.5850	*****	*****	*****	*****	*****	10.959
98	aC-NHCO	N-(2-methylphenyl)acetamide (1)	9.8204	7.4955	*****	*****	*****	*****	*****	4.370
99	aC-NCO	Phenylmethylacetamide (1)	7.2552	*****	*****	*****	*****	*****	*****	*****
100	NHCONH	N,N'-dimethylurea (1)	9.3110	8.9406	*****	*****	*****	*****	*****	9.862
101	NH ₂ CONH	Methylurea (1)	14.2020	16.3539	*****	*****	*****	*****	*****	12.845
102	NH ₂ CON	N,N'-dimethylurea (1)	13.0856	2.0796	*****	*****	*****	*****	*****	10.958
103	NHCON	Trimethylurea (1)	8.4447	7.1529	*****	*****	*****	*****	*****	12.098
104	NCON	Tetramethylurea (1)	3.5041	4.1459	*****	*****	*****	*****	*****	9.557
105	aC-NHCONH ₂	Phenylurea (1)	13.4695	5.7604	*****	*****	*****	*****	*****	16.703
106	aC-NHCONH	N,N'-diphenylurea	23.2570	1.1633	*****	*****	*****	*****	*****	18.460
107	NHCO except as above	N-chloroacetamide (1)	3.0882	*****	*****	*****	*****	*****	*****	*****
108	CH ₂ Cl	1-Chlorobutane (1)	1.9253	2.6364	6.2561	0.021419	112.12	-19.484	-65.056	11.754
109	CHCl	2-Chloropropane (1)	1.0224	2.0246	4.3756	0.015640	100.78	-31.933	-65.127	12.048
110	CCl	2-Chloro-2-methylpropane (1)	1.8424	1.7049	3.7063	0.009187	87.01	-37.848	-62.881	16.597
111	CHCl ₂	1,1-Dichloroethane (1)	2.5196	3.3420	7.8956	0.028236	159.79	-24.214	-80.812	17.251
112	CCl ₂	2,2-Dichloropropane (1)	3.6491	2.9609	*****	*****	*****	*****	*****	20.473
113	CCl ₃	1,1,1-Trichloroethane	4.4493	3.9093	8.8073	0.036746	204.71	-44.122	-105.369	20.550
114	CH ₂ F	1-Fluorobutane (1)	1.5597	1.5022	3.3179	0.023315	87.71	-180.212	-227.469	8.238
115	CHF	2-Fluorobutane (1)	1.1289	1.3738	2.6702	0.020040	78.08	-228.239	-261.901	*****
116	CF	2-Fluoro-2-methylpropane (1)	2.5398	1.0084	2.1633	-0.010120	*****	*****	*****	6.739
117	CHF ₂	1,1-Difluoroethane (1)	2.1689	2.2238	3.5702	0.031524	102.71	-411.239	-463.901	*****
118	CF ₂	Perfluorohexane (4)	0.1312	0.5142	0.8543	0.018572	95.09	*****	*****	1.621
119	CF ₃	Hexafluoroethane (2)	1.4828	1.1916	1.7737	0.048565	108.85	-615.333	-673.875	7.352
120	CCl ₂ F	Tetrachloro-1,2-difluoroethane (2)	3.2035	2.5053	5.1653	0.037948	171.04	-249.020	-306.765	8.630
121	HCClF	1-Chloro-1,2,2,2-tetrafluoroethane (1)	*****	2.0542	*****	*****	*****	*****	*****	*****
122	CClF ₂	1,2-Dichlorotetrafluoroethane (2)	1.7510	1.7227	3.0593	0.041641	146.01	-396.814	-458.074	8.086
123	aC-Cl	Chlorobenzene (1)	1.7134	2.0669	5.7046	0.016033	92.67	1.985	-17.002	11.224
124	aC-F	Hexafluorobenzene (6)	0.9782	0.7945	1.5491	0.014037	54.36	-141.306	-160.965	3.965
125	aC-I	Iodobenzene (1)	2.1905	3.7739	12.4470	0.014403	131.08	91.505	95.048	*****
126	aC-Br	Bromobenzene (1)	2.4741	2.8414	8.4199	0.010199	104.12	42.977	38.917	14.393
127	I ⁻ except as above	Iodoethane (1)	1.9444	3.1778	8.5775	-0.004637	104.28	43.910	47.632	14.171
128	Br ⁻ except as above	Bromoethane (1)	1.7641	2.4231	4.5036	-0.001460	77.99	5.528	-1.703	9.888
129	F ⁻ except as above	Benzyl fluoride (1)	1.2308	0.8504	0.8976	0.012034	24.62	-182.973	-201.968	*****
130	Cl ⁻ except as above	Ethyl chloroacetate (1)	1.5454	1.5147	4.0947	0.007923	57.77	-29.876	-46.963	*****
131	CHNOH	Propionaldehyde oxime (1)	3.9813	4.5721	*****	*****	*****	*****	*****	*****
132	CNOH	Diethyl ketoxime (1)	3.5484	4.0142	*****	*****	*****	*****	*****	*****
133	aC-CHNOH	Phenyl oxime (1)	10.5579	*****	*****	*****	*****	*****	*****	*****
134	OCH ₂ CH ₂ OH	2-Ethoxyethanol (1)	2.3651	4.8721	10.4579	0.025986	159.33	-233.335	-343.903	31.493
135	OCHCH ₂ OH	2-Ethoxy-1-propanol (1)	*****	4.2329	*****	*****	*****	*****	*****	*****
136	OCH ₂ CHOH	1-Methoxy-2-propanol (1)	1.5791	3.6653	*****	0.018783	147.66	-239.423	-333.385	*****
137	-O-OH	tert-Butylhydroperoxide (1)	4.8181	3.1669	5.8307	-0.002815	58.01	-75.568	-125.111	*****
138	CH ₂ SH	Ethanthiol (1)	2.2992	3.1974	7.7300	0.017299	105.68	27.469	-8.021	16.815
139	CHSH	2-Propanethiol (1)	0.9704	2.5910	5.8527	0.008968	109.36	27.030	3.510	17.098
140	CSH	2-Methyl-2-propanethiol (1)	4.2329	2.0902	4.6431	0.005118	94.01	27.338	12.589	18.397
141	aC-SH	Benzenethiol (1)	2.8464	3.2675	9.5115	0.010086	95.08	48.905	41.648	17.413
142	-SH (except as above)	Cyclohexanethiol (1)	0.9600	2.3323	7.7987	0.006399	57.89	15.818	11.339	9.813
143	CH ₂ S	Dimethylsulfide (1)	1.7150	2.9892	6.9733	0.018013	122.03	35.845	-3.337	14.296
144	CH ₂ S	Diethylsulfide (1)	1.0063	2.6524	6.4871	0.015254	106.60	42.684	21.492	16.965
145	CHS	Diisopropylsulfide (1)	0.7892	2.0965	*****	*****	*****	*****	*****	19.038
146	CS	di-tert-Butylsulfide (1)	1.1170	1.6412	*****	*****	*****	*****	*****	19.996
147	aC-S-	Phenyl methyl sulfide (1)	0.9646	2.9731	*****	*****	*****	*****	*****	*****
148	SO	Dimethyl sulfoxide (1)	5.3663	6.2796	19.8953	-0.005534	82.36	-52.231	-71.050	*****
149	SO ₂	Dimethyl sulfone (1)	7.0778	7.0976	17.2586	-0.000784	89.95	-257.608	-305.498	*****
150	SO ₃ (sulfite)	*****	3.9199	8.6910	0.004240	115.80	*****	-430.833	*****	*****
151	SO ₃ (sulfonate)	Dimethyl sulfonate (1)	5.8426	6.7785	*****	*****	*****	*****	*****	*****
152	SO ₄ (sulfate)	Dimethyl sulfate (1)	3.6976	5.5627	18.9366	-0.027208	144.58	-519.853	-621.412	*****
153	aC-SO	Phenyl methyl sulfoxide (1)	3.9911	6.1185	*****	*****	*****	*****	*****	*****
154	aC-SO ₂	Diphenyl sulfone (1)	5.2948	8.4333	*****	*****	135.47	-314.643	-370.493	*****
155	PH (phosphine)	Dimethylphosphine (1)	*****	2.0536	*****	*****	*****	*****	*****	3.281
156	P (phosphine)	Trimethylphosphine (1)	*****	1.0984	*****	*****	*****	*****	*****	*****
157	PO ₂ (phosphite)	Triethylphosphite (1)	1.0306	2.7900	*****	*****	*****	*****	*****	*****
158	PHO ₂ (phosphonate)	Dimethylphosphonate (1)	*****	5.6433	*****	*****	*****	*****	*****	*****
159	PO ₃ (phosphonate)	Trimethylphosphonate (1)	*****	4.5468	*****	*****	*****	*****	*****	*****
160	PHO ₄ (phosphate)	Diethylphosphate (1)	2.7461	5.1567	*****	*****	*****	*****	*****	*****
161	PO ₄ (phosphate)	Trimethylphosphate (1)	2.0330	3.7657	16.9914	-0.029036	85.59	*****	-1060.325	*****
162	aC-PO ₄	Triphenylphosphate (1)	-1.7840	2.3522	*****	*****	*****	*****	-1005.161	*****
163	aC-P	Triphenylphosphine (1)	0.2337	2.9272	38.6148	-0.126108	-142.79	*****	72.339	*****
164	CO ₃ (carbonate)	Diethylcarbonate (1)	3.6593	2.8847	6.6804	0.007235	93.56	-447.186	-516.282	21.613
165	C ₂ H ₂ O	Ethyl oxirane (1)	1.3135	2.8451	6.6418	0.021238	125.43	11.149	-52.241	*****
166	C ₂ H ₂ O	2,2-Dimethyl oxirane (1)	*****	2.6124	6.0159	0.010678	194.36	1.890	-51.390	*****
167	C ₂ O	Trimethyl oxirane (1)	*****	2.2036	*****	*****	*****	*****	*****	*****
168	CH ₂ (cyclic)	Cyclopentane (5)	0.5699	0.8234	1.8815	0.009884	49.24	13.287	-18.575	3.341
169	CH (cyclic)	Methylcyclopentane (1)	0.0335	0.5946	1.1020	0.007596	44.95	6.107	-12.464	6.416
170	C (cyclic)	1,1-Dimethylcyclohexane (1)	0.1695	0.0386	-0.2399	0.003268	33.32	-0.193	-2.098	7.017
171	CH=CH (cyclic)	Cyclobutene (1)	1.1936	1.5985	3.6426	0.013815	83.91	86.493	59.841	7.767
172	CH=C (cyclic)	1-Methylcyclopentene (1)	0.4344	1.2529	3.5475	0.010576	70.98	67.056	64.295	7.171
173	C=C (cyclic)	1,2-Dimethylcyclopentene (1)	0.3048	1.1975	*****	*****	*****	*****	*****	*****
174	CH ₂ =C (cyclic)	Methylene cyclohexane (1)	0.2220	1.5109	4.4913	0.019101	83.96	*****	*****	5.351
175	NH (cyclic)	Cyclopentimine (1)	3.4814	2.1634	5.9726	-0.003678	51.80	72.540	23.138	13.700
176	N (cyclic)	N-methylpyrrolidine (1)	0.6040	1.6541	4.3905	-0.001179	31.41	83.779	65.622	*****
177	CH=N (cyclic)	Imidazole (1)	5.5779	6.5230	*****	*****	*****	*****	*****	3.993
178	C=N (cyclic)	2-Methyl-1H-imidazole (1)	6.6382	6.6710	*****	*****	*****	*****	*****	*****
179	O (cyclic)	Tetrahydropyran (1)	1.3828	1.0245	2.7409	-0.000387	17.69	-114.062	-137.353	6.877
180	CO (cyclic)	Cyclobutanone (1)	3.2119	2.8793	12.6396	-0.000207	57.38	-156.672	-180.166	17.124
181	S (cyclic)	2-Methyl-thiophene (1)	1.6023	2.3256	5.5523	0.001540	45.45	12.020	15.453	12.262
182	SO ₂ (cyclic)	Cyclobutadiene sulfone (1)	6.1006	*****	24.3995	0.002487	96.66	-241.601	-283.839	*****

Table 4.6 The second order group and their contributions in Gani method²

Group	Example	T_{n2j}	T_{n2j}	T_{c2j}	P_{c2j}	V_{c2j}	G_{c2j}	H_{c2j}	H_{c2j}	H_{ln2j}
1	(CH ₃) ₂ CH	2-Methylpentane (1)	0.1175	-0.0035	-0.0471	0.000473	1.71	-0.418	-0.419	0.396
2	(CH ₃) ₃ C	2,2,4,4-Tetramethylpentane (2)	-0.1214	0.0072	-0.1778	0.000340	3.14	-2.776	-1.967	0.554
3	CH(CH ₃)CH(CH ₃)	2,3,4-Trimethylpentane (2)	0.2390	0.3160	0.5602	-0.003207	-3.75	6.996	6.065	0.532
4	CH(CH ₃)C(CH ₃) ₂	2,2,3,4,4-Pentamethylpentane (2)	-0.3276	0.3976	0.8994	-0.008733	-10.06	8.938	8.078	0.623
5	C(CH ₃) ₂ C(CH ₃) ₂	2,2,3,3,4,4-Hexamethylpentane (2)	3.3297	0.4487	1.5535	-0.016852	-8.70	10.735	10.535	5.086
6	CH ₂ =CH _n -CH _m =CH _p (<i>k, m, n, p</i> in 0..2)	1,3-Butadiene (1)	0.7451	0.1097	0.4214	0.000792	-7.88	-6.562	-11.786	1.632
7	CH ₂ -CH _n =CH _m (<i>m, n</i> in 0..2)	2-Methyl-2-butene (3)	0.0524	0.0369	-0.0172	-0.000101	0.50	-0.120	-0.048	0.064
8	CH ₂ -CH _n =CH _m (<i>m, n</i> in 0..2)	1,4-Pentadiene (2)	-0.1077	-0.0537	0.0262	0.000815	0.14	1.006	1.449	-0.060
9	CH ₂ -CH _n =CH _m (<i>m, n</i> in 0..2; <i>p</i> in 0..1)	3-Methyl-1-butene (1)	-0.2485	-0.0093	-0.1526	-0.000163	-2.67	3.857	3.964	0.004
10	CHCHO or CCHO	2-Methylbutyraldehyde (1)	0.5715	-0.1286	-1.0434	0.005789	10.36	-0.525	1.514	-0.550
11	CH ₃ COCH ₂	2-Pentanone (1)	-0.0968	-0.0215	-0.0338	-0.000111	-4.08	-1.543	0.033	-0.403
12	CH ₃ COCH or CH ₃ COC	3-Methyl-2-pentanone (1)	-0.6024	-0.0803	-0.3658	-0.001892	3.02	2.202	4.994	0.723
13	CHCOOH or CCOOH	2-Methyl butanoic acid (1)	-3.1734	-0.3203	-4.7275	0.006916	10.56	3.920	1.121	7.422
14	CH ₃ COCH or CH ₃ COOC	Isopropyl acetate (1)	0.2114	-0.2066	-0.5537	-0.000569	4.28	-11.779	-12.295	-1.871
15	CO-O-CO	Propanoic anhydride (1)	-1.2441	-0.0500	-0.3576	0.001812	2.98	-16.075	-14.140	****
16	CHOH	2-Butanol (1)	-0.3489	-0.2825	-0.6768	0.000246	-3.04	-5.614	-4.422	-0.206
17	COH	2-Methyl-2-butanol (1)	0.3695	-0.5325	-1.5224	0.003224	13.98	-25.382	-25.929	-1.579
18	CH ₃ COCH ₂ OH (<i>n</i> in 0..2)	3-Hydroxy-2-butanone (1)	0.9886	-0.2987	-0.3940	-0.002912	5.17	6.621	8.244	****
19	NCCOH or NCCOH	2-Hydroxypropionitrile (1)	-1.1810	0.2981	0.3414	-0.000516	0.68	4.833	0.000	****
20	OH-CH ₂ -COO (<i>n</i> in 0..2)	Ethyl lactate (1)	-0.1526	-0.2310	****	****	****	****	****	****
21	CH _n (OH)CH _m (OH) (<i>m, n</i> in 0..2)	Ethylene glycol (1)	-0.0414	0.8854	1.9395	-0.004712	7.54	-1.051	-0.592	-6.611
22	CH _n (OH)CH _m (-) (<i>m, n, p</i> in 0..2)	2-Amino-1-butanol (1)	-0.5941	0.5082	1.2342	0.002581	5.58	-1.506	-0.959	****
23	CH _n (NH ₂)CH _m (NH ₂) (<i>m, n</i> in 0..2)	Ethylenediamine (1)	0.3258	-0.0064	-3.3555	0.000726	20.82	0.344	-1.443	2.384
24	CH _n (NH ₂)CH _m (NH ₂) (<i>m, n</i> in 1..2)	Diethylenetriamine (1)	-1.8403	0.2318	-1.1598	0.000157	-26.31	3.848	3.608	****
25	H ₂ NCOCH ₂ CH _n CONH ₂ (<i>m, n</i> in 1..2)	Butanediamide (1)	11.5351	****	****	****	****	****	****	****
26	CH _n (NH ₂)-COOH (<i>m, n</i> in 0..2)	L-Alanine (1)	12.3481	****	62.4740	-0.002696	17.78	3.145	6.598	****
27	HOOC-CH _n -COOH (<i>n</i> in 1..2)	Malonic acid (1)	0.9327	-0.1222	1.9595	-0.001479	12.46	-5.217	-6.058	****
28	HOOC-CH _n -CH _m -COOH (<i>n, m</i> in 1..2)	Succinic acid (1)	7.5057	****	0.7686	0.000090	15.17	-4.281	-6.929	****
29	HO-CH ₂ -COOH (<i>n</i> in 1..2)	2-Hydroxyisobutyric acid (1)	-0.4531	-0.4625	****	****	****	****	****	****
30	NH ₂ -CH _n -CH _m -COOH (<i>n, m</i> in 1..2)	β-Alanine (1)	14.1593	****	****	****	****	****	****	****
31	CH ₃ -O-CH ₂ -COOH (<i>n</i> in 1..2)	Methoxyacetic acid (1)	-2.3026	0.9198	0.4750	-0.001445	7.91	-2.678	-1.727	****
32	HS-CH ₂ -COOH	2-Mercaptopropionic acid (1)	-2.1535	****	****	****	****	****	****	****
33	HS-CH ₂ -CH _n -COOH (<i>n, m</i> in 1..2)	β-Thiolactic acid (1)	-2.7514	****	-0.2697	0.000655	20.43	-7.376	7.292	****
34	NC-CH ₂ -CH _n -CN (<i>n, m</i> in 1..2)	1,2-Dicyanoothane (1)	4.0747	1.8957	1.9699	0.002330	24.82	18.974	5.661	****
35	OH-CH ₂ -CH _n -CN (<i>n, m</i> in 1..2)	3-Hydroxypropanenitrile (1)	-0.9493	1.3434	0.2311	-0.001022	14.54	0.558	-3.906	****
36	HS-CH ₂ -CH _n -SH (<i>n, m</i> in 1..2)	1,2-Ethanedithiol (1)	0.2232	0.1815	2.1272	0.001321	-10.31	6.728	0.794	-0.683
37	COO-CH _n -CH _m -OOC (<i>n, m</i> in 1..2)	Ethylene glycol diacetate (1)	-0.5946	0.3401	1.5418	-0.003385	-2.33	1.306	4.025	1.203
38	OOC-CH _n -CH _m -COO (<i>n, m</i> in 1..2)	Dimethylsuccinate (1)	2.5962	0.5794	****	****	****	****	****	2.303
39	NC-CH ₂ -COO (<i>n</i> in 1..2)	Methylcyanoacetate (1)	-0.2509	1.2171	2.7051	-0.001999	-0.73	****	****	****
40	COCH ₂ COO (<i>n</i> in 1..2)	Methylacetoacetate (1)	0.6304	0.2427	0.7502	-0.000231	1.69	10.556	-7.261	****
41	CH _n -O-CH ₂ =CH _p (<i>m, n, p</i> in 0..3)	Ethyl vinyl ether (1)	-0.0811	0.1399	0.2900	-0.000432	-4.54	-10.098	-9.411	0.372
42	CH _n =CH _m -F (<i>m, n</i> in 0..2)	1-Fluoro-1-propene (1)	-0.2568	0.0591	****	****	****	****	****	****
43	CH _n =CH _m -Br (<i>m, n</i> in 0..2)	1-Bromo-1-propene (1)	-0.4329	-0.3192	****	-0.010021	2.63	14.470	17.014	****
44	CH _n =CH _m -I (<i>m, n</i> in 0..2)	1-Iodo-1-propene (1)	****	-0.3486	****	****	****	****	****	****
45	CH _n =CH _m -Cl (<i>m, n</i> in 0..2)	1-Chloro-2-methylpropene (1)	0.0446	-0.0268	-0.0188	0.000152	2.80	8.207	9.715	****
46	CH _n =CH _m -CN (<i>n, m</i> in 0..2)	Acrylonitrile (1)	0.1027	0.0653	-1.1249	0.000893	3.82	-8.304	-16.903	****

Table 4.7 The third order group and their contributions in Gani method²

Group	Example	T_{m3}	T_{i3}	T_{c3}	P_{c3}	V_{c3}	G_{i3}	H_{i3}	H_{m3}
1	HOOC-(CH ₂) _n -COOH ($m > 2, n$ in 0..2)	-1.5257	1.6498	-1.6986	0.001544	-3.72	-4.708	-6.572	-7.583
2	NH ₂ -(CH ₂) _n -COOH ($m > 2, n$ in 0..2)	11.2277	*****	*****	*****	*****	*****	*****	*****
3	NH ₂ -(CH ₂) _n -OH ($m > 2, n$ in 0..2)	0.7732	1.0750	0.4950	0.000728	-23.74	3.079	4.171	-4.840
4	OH-(CH ₂) _n -OH ($m > 2, n$ in 0..2)	0.6674	0.7193	0.1725	-0.000327	-0.84	7.536	5.411	-0.272
5	OH-(CH ₂) _k -O-(CH ₂) _n -OH ($m, k > 0; p, n$ in 0..2)	-0.1073	1.1867	6.6872	0.001937	1.44	-8.397	-8.651	1.661
6	OH-(CH ₂) _k -S-(CH ₂) _n -OH ($m, k > 0; p, n$ in 0..2)	-1.3891	*****	2.6769	0.003792	-1.62	10.194	8.164	-3.479
7	OH-(CH ₂) _n -NH ₂ -(CH ₂) _n -OH ($m, k > 0; p, n, x$ in 0..2)	-0.0781	0.2991	*****	0.003254	-0.69	1.662	1.753	0.301
8	CH ₃ -O-(CH ₂) _n -OH ($m > 2; n, p$ in 0..2)	*****	-0.4605	*****	*****	*****	*****	*****	*****
9	NH ₂ -(CH ₂) _n -NH ₂ ($m > 2; n$ in 0..2)	-0.0604	0.0060	-4.3195	0.006734	6.69	4.100	0.371	5.666
10	NH ₂ -(CH ₂) _n -NH ₂ ($m > 2; k$ in 0..1; n in 0..2)	-1.1888	-0.1819	*****	*****	*****	*****	*****	*****
11	SH-(CH ₂) _n -SH ($m > 2; n$ in 0..2)	0.6669	0.4516	*****	*****	*****	*****	*****	*****
12	NC-(CH ₂) _n -CN ($m > 2; n$ in 0..2)	-0.3798	1.3440	0.0834	-0.011090	-36.89	-7.035	7.782	-0.607
13	COO-(CH ₂) _n -OOC ($m > 2; n$ in 0..2)	-2.6542	*****	*****	*****	*****	*****	*****	*****
14	aC-(CH ₂) _n -CH ₂ -aC (fused rings) (n, m in 0..1)	0.2479	-0.3741	-0.0185	0.000851	-8.87	-1.601	2.689	-2.703
15	aC-aC (different rings)	1.1395	-0.4961	6.1894	-0.040100	-26.26	-4.459	-4.558	-0.385
16	aC-CH ₂ -aC (different rings) (n in 0..1)	0.0570	-0.4574	-0.2474	-0.005826	-8.55	-5.267	-5.914	-0.442
17	aC-CH ₂ -aC (fused rings) (n in 0..1)	-0.5640	-0.1736	0.5060	-0.003746	-11.56	-4.203	-4.863	-0.143
18	aC-(CH ₂) _n -aC (different rings) ($m > 1; n$ in 0..2)	1.9902	0.3138	3.0321	0.003007	9.73	1.318	0.084	5.377
19	aC-(CH ₂) _n -CH ₂ -aC (different rings) ($m > 0; n$ in 0..2)	*****	0.5928	*****	*****	*****	*****	*****	*****
20	CH ₂ -CH ₂ -CH ₂ -aC (different rings)	0.5460	0.4387	2.1761	0.002745	7.72	-67.517	-66.870	*****
21	CH ₂ -CH ₂ -CH ₂ -CH ₂ -aC (different rings) ($m > 0; n$ in 0..2)	0.4497	0.5632	*****	*****	*****	*****	*****	*****
22	CH multiring	0.6647	0.1415	0.4963	-0.000985	-3.33	*****	*****	0.223
23	C multiring	0.0792	*****	*****	*****	*****	*****	*****	*****
24	aC-CH ₂ -aC (different rings) (m in 0..2)	0.6457	0.2391	0.1174	-0.002673	-4.67	-0.729	0.866	-0.958
25	aC-(CH ₂) _n -CH ₂ -aC (different rings) (m, n in 0..2)	0.9608	0.7192	0.7039	-0.004661	14.31	-0.702	-2.291	3.275
26	(CH ₂) _n -C(CH ₂) _n -CH=CH-(C=CH ₂) _n (different rings)	16.2235	*****	*****	*****	*****	*****	*****	*****
27	(CH ₂) _n -C(CH ₂) _n -CH ₂ -C(CH ₂) _n (different rings)	16.8558	*****	*****	*****	*****	*****	*****	*****
28	aC-CO-aC (different rings)	-1.0394	1.0171	-0.2678	-0.001837	-7.05	11.125	7.108	-4.091
29	aC-CH ₂ -CO-aC (different rings) (m in 0..2)	-0.4486	0.9674	*****	*****	*****	*****	*****	*****
30	aC-CO-(C=CH ₂) _n -aC (different rings) (n in 0..1)	-0.1376	0.1126	*****	*****	*****	*****	*****	*****
31	aC-CO-CO-aC (different rings)	0.4361	0.9317	*****	*****	*****	*****	*****	-3.687
32	aC-CO ₂ -aC (fused rings)	3.6847	0.5031	*****	*****	*****	*****	*****	2.047
33	aC-CO-(CH ₂) _n -CO-aC (different rings) ($m > 0; n$ in 0..2)	4.9038	*****	*****	*****	*****	*****	*****	7.327
34	aC-CO-CH ₂ -aC (different rings) (n in 0..1)	-7.0038	*****	*****	*****	*****	*****	*****	*****
35	aC-CO-NH ₂ -aC (different rings) (n in 0..1)	5.9653	*****	*****	*****	*****	*****	*****	2.510
36	aC-NH ₂ -CONH ₂ -aC (different rings) (n, m in 0..1)	1.5629	*****	*****	*****	*****	*****	*****	0.018
37	aC-CO-N ₃ -aC (different rings)	-9.1856	*****	*****	*****	*****	*****	*****	*****
38	aC-S ₂ -aC (fused rings)	0.2612	0.2242	3.5541	0.004600	12.60	8.333	9.212	-0.784
39	aC-S-aC (different rings)	-1.8403	0.0185	*****	*****	*****	*****	*****	*****
40	aC-PO ₃ -aC (different rings) (n in 0..4)	0.0393	*****	*****	*****	*****	*****	*****	*****
41	aC-SO ₂ -aC (different rings) (n in 1..4)	0.9514	-0.0850	*****	*****	*****	*****	*****	-2.485
42	aC-NH ₂ -aC (fused rings) (n in 0..1)	3.4983	1.1457	3.5541	0.017201	0.44	-2.221	-16.080	0.196
43	aC-NH ₂ -aC (different rings)	-0.3048	0.5768	0.9519	0.008484	1.42	-0.596	-1.994	1.934
44	aC-(C=N) ₂ -aC (different rings)	-1.3060	-0.5335	*****	*****	*****	*****	*****	*****
45	aC-(N=CH ₂) _n -aC (fused rings) (n in 0..1)	-4.9289	-5.2736	*****	*****	*****	*****	*****	-0.599
46	aC-(CH ₂ =N) ₂ -aC (fused rings) (n in 0..1)	-10.1007	*****	*****	*****	*****	*****	*****	*****
47	aC-O-CH ₂ -aC (different rings) (n in 0..2)	1.0834	6.6571	*****	*****	*****	*****	*****	*****
48	aC-O-aC (different rings)	-0.4803	-0.8252	-0.9785	0.001162	-2.63	2.668	-5.074	1.193
49	aC-CH ₂ -O-CH ₂ -aC (different rings) (n, m in 0..2)	-3.2676	6.2790	-1.4002	-0.004716	28.42	-4.229	-2.303	-3.971
50	aC-O ₂ -aC (fused rings)	-0.3545	-0.6848	*****	*****	*****	*****	*****	-1.153
51	AROMFUSED[2]	0.2825	0.0441	-1.0095	-0.001332	-6.88	1.993	1.904	0.694
52	AROMFUSED[2]s ¹	-1.2836	-0.1666	0.1605	-0.002030	-3.17	-2.940	-2.274	-3.699
53	AROMFUSED[2]s ²	0.3378	-0.2692	-0.6765	-0.002436	-3.85	-1.873	-1.316	2.037
54	AROMFUSED[2]s ² s ³	1.8941	-0.2807	*****	*****	*****	*****	*****	2.150
55	AROMFUSED[2]s ¹ s ⁴	-2.7585	-0.3294	*****	*****	*****	*****	*****	*****
56	AROMFUSED[2]s ² s ²	-3.0362	-0.2931	*****	*****	*****	*****	*****	*****
57	AROMFUSED[2]s ² s ³	-3.2228	-0.3360	*****	*****	*****	*****	*****	*****
58	AROMFUSED[3]	1.6600	0.0402	-1.0430	0.004695	35.21	3.896	5.819	1.176
59	AROMFUSED[4a]	7.0402	1.0466	3.3011	0.015244	-6.96	13.843	11.387	5.027
60	AROMFUSED[4a]s ¹	-6.3463	-7.8521	*****	*****	*****	*****	*****	*****
61	AROMFUSED[4a]s ¹ s ⁴	6.8373	*****	*****	*****	*****	*****	*****	*****
62	AROMFUSED[4p]	-1.5856	6.9126	2.8885	0.007280	-24.02	-16.040	-19.089	-3.417
63	AROMFUSED[4p]s ¹ s ⁴	2.0821	*****	*****	*****	*****	*****	*****	*****
64	PYRIDINE.FUSED[2]	-4.4725	-0.9432	1.1251	-0.005369	63.29	8.688	13.586	-4.967
65	PYRIDINE.FUSED[2-iso]	-2.5898	-0.5844	3.9241	-0.011207	-2.71	-5.112	-0.314	-2.587
66	PYRIDINE.FUSED[4]	1.0358	6.1733	7.7134	-0.001275	-12.04	20.073	15.786	-1.365

The general expression of Gani model is shown in E.q. 4.2.

$$f(X) = \sum_i N_i C_i + \sum_j M_j D_j + \sum_k O_k E_k \quad (4.2)$$

N_i , M_j and O_k are the occurrences of each group within the three group orders respectively. C_i , D_j and E_k are the property contributions of each group within ith respective group order.

The property contributions of E.q.4.2 are shown in Table 4.8.

Table 4.8 The additional parameters in E.q.4.2²

Property	f(X): the LHS of E.q.4.2
Normal melting point(T_m),K	$\exp(T_m/147.450)$
Normal boiling point(T_b),K	$\exp(T_b/222.543)$
Critical temperature(T_c),K	$\exp(T_c/231.239)$
Critical pressure(P_c),Bar	$(P_c-5.9827)^{-0.5}-0.108998$
Critical volume(V_c),cm ³ /mol	$V_c-7.95$
Standard Gibbs energy at 298K(G_f),kJ/mol	$G_f+34.967$
Standard enthalpy of formation at 298K(H_f),kJ/mol	$H_f-5.549$
Standard enthalpy of vaporization at 298K(H_v),kJ/mol	$H_v-11.733$
Standard enthalpy of fusion(H_{fus}),kJ/mol	$H_{fus}+2.806$

The Gani method considered not only 182 individual first order functional group structures, but also the inter-group effects and ring configurations of 122 second order group structures and 66 third order structural groups. Because of overlapping structural groups, it is necessary to prioritize their identification within the SPCM code. The priority used in this thesis is: heteroatom ring class > aromatic ring class > naphthenic ring class > non ring class. An algorithm to identify those groups from a BE matrix was developed for this thesis, whereby following the properties listed in Table 4.8 to be calculated by E.q.4.2. As a byproduct, this algorithm also allow for the calculation of the frequency of functional groups from compositional analytic data (e.g., methyl, ethyl, phenol, etc.).

Although these two group contribution methods provide most of the properties a given model may need, the density of a component is not included. Density plays an important role in bulk property estimation. The calculation of density for a pure component can be indirectly derived from the properties obtained from the above group contribution methods.

4.2.2.3 Density calculation for a pure component

There are two methods to calculate the density of one pure component. One is to first calculate the molecular volume by the equation of state (EOS), and then get the value of the density by dividing the molecular volume by the molecular weight. The second is to use a correlation model based on critical properties. This thesis explored both methods.

The first method was employed by using two cubic equations of state: Peng-Robinson (PR) and Soave modification of Redlich-Kwong (SRK). The critical properties in the EOS model are found by using the above group contribution method as well as other published correlations and parameters such as the acentric factor.

The second method applied in this thesis follows that of the Yen-woods correlation model based on critical properties shown in E.q.4.3.

$$\rho = (1 + \sum_{i=1}^4 k_i (1 - T_r)^{i/3}) \cdot MW / V_c$$

$$k_1 = 17.4425 - 214.578 Z_c + 989.625 Z_c^2 - 1522.06 Z_c^3$$

$$k_2 = -3.28257 - 13.6377 Z_c + 107.4844 Z_c^2 - 384.211 Z_c^3 \quad (Z_c < 0.26) \quad (4.3)$$

$$k_2 = 60.2091 - 402.063 Z_c + 501.0 Z_c^2 + 641.0 Z_c^3 \quad (Z_c \geq 0.26)$$

$$Z_c = \frac{P_c V_c}{RT_c}, T_r = \frac{T}{T_c}$$

The methods to estimate the density of a pure component are not as accurate as the other properties estimated using the above group contribution methods. Therefore, future work should be done to modify or replace the density calculation methods.

4.2.2.4 Summary of thermodynamic property calculation

Overall, group contribution methods perform well for the calculation of most of the thermodynamic properties required in composition and reaction modeling. However, the accuracy of some thermodynamic properties for pure components (e.g., the heat of formation) may not meet the standards required when reaction modeling with an LFER approximation because of the sensitivity of some reaction indices in LFER. In addition, the group contribution method can not be used to calculate the properties of immediate species (e.g., radical, ions), which are important when modeling at a mechanistic level. Therefore, quantum chemistry package has been applied for the calculation of those properties.

4.2.3 Selected properties calculated by quantum chemistry

The semi-empirical molecular orbital calculation from quantum chemistry can provide a high degree of accuracy for some properties, such as heat of formation, deriving from the molecular geometry. It is used MOPAC package as quantum chemistry calculation package in this thesis. To apply MOPAC in SPCM, it needs to convert the 2D structure of BE matrix to 3D structure first

4.2.3.1 The molecule's conversion from 2D BE matrix to 3D structure

The 3D structure of a molecule provides a preliminary basis to launch the MOPAC calculation. The 3D representation of a molecule is determined by its 3D geometry. The geometry can be defined in terms of either internal or Cartesian coordinates. Both of these two coordinates can be used as input parameters for quantum chemistry calculation. In thesis, it will illustrate the 3D geometry of a molecule by internal coordinates since it is a traditional default input format for MOPAC³. In this thesis, it applied two 2D-3D conversion methods. It has been developed an improved algorithm based on Broadbelt⁴ for pure hydrocarbon mixtures. Moreover, an external package-OpenBabel⁵ has been employed to get the 3D coordinate for a more general complex feedstock.

An internal coordinate of a component can be expressed as following: For any one *current* atom *i* in this component, there is an interatomic distance (*bond length*) from an already-defined *parent* atom *j*; there is an interatomic angle in degrees (*bond angle*) between atoms *i* and *j* and an already defined *angle atom* *k*, (*k* and *j* must be different atoms), and, finally, a torsional angle (*dihedral angle*) in degrees between atoms *i*, *j*, *k*, and an already defined *dihedral* atom *l* (*l* cannot be the same as *k* or *j*). An internal coordinate example of ethane is shown in Table 4.9.

Table 4.9 Internal coordinate representation of ethane

	Bond length		Bond angle		Dihedral angle		Parent atom	Angle atom	Dihedral atom
	1	2	3	4	5	6	7	8	9
C	0.00	0	0.0	0	0.0	0	0	0	0
C	1.52	1	0.0	0	0.0	0	1	0	0
H	1.11	1	111.1	1	0.0	0	1	2	0
H	1.11	1	111.1	1	-120.0	1	1	2	3
H	1.11	1	111.1	1	120.0	1	1	2	3
H	1.11	1	111.1	1	60.0	1	2	1	3
H	1.11	1	111.1	1	-60.0	1	2	1	3
H	1.11	1	111.1	1	180.0	1	2	1	3

In Table 4.9, the bond length, bond angle and dihedral angle are tabulated in columns 1, 3 and 5 respectively. The optimizability of each of these geometric parameters is designated by a number in columns 2, 4 and 6 respectively. A zero is specified if the geometric parameter remains constant throughout the process of achieving self-consistency. A one is recorded if the geometric parameter is allowed to vary as the optimization is carried out. Columns 7, 8 and 9 denote the other atoms with respect to which these values are specified, referred to here as the parent atom, angle atom and dihedral atom.

For each row of Table 4.9, the indices of the current atom, parent atom, angle atom and dihedral atom can be determined by the connectivity of those atoms in the BE matrix. The values of the bond angle, bond length and dihedral angle are determined by the stable state structural properties of those atoms.

Broadbelt⁴ developed an approximation method to parse the 2D structure of the BE matrix to a 3D representation based on the experimental data of atom structural properties for pure simple hydrocarbon mixtures. However, this algorithm is limited in two key aspects: it lacks the data for five-member and heteroatom ring; and the BE matrix connectivity search algorithm ignores the case of cata-condensed ring structures (e.g., phenanthrene, pyrene etc.). Therefore, a modified algorithm based on Broadbelts⁴ was developed for this thesis.

Typical structural data for internal coordinate are listed below in Table 4.10.

Table 4.10 Typical atom data as internal coordinate

Bond Type	Bond length
C-H	1.11
C-C	1.52
C=C	1.32
C#C	1.13
Hybridization of parent atom	Bond angle
sp	180
sp ²	120
sp ³	109
sp ³ -five member ring	100
Non cyclic atom	
Hybridization of parent atom	Dihedral angle
Not dependent on parent atom	(180,60,-60)
sp ²	(-180)
sp ³	(120,-120)
cyclic atom	
aromatic	(0,0,0,0,0,0,)
cyclohexadienyl	(30,-20,0,10,0,-20)
cyclohexenyl	(40,-10,0,-20,50,-60)
cyclohexyl	(60,-60,60,-60,60,-60)
cyclopentadienyl	(15,-10,-,5,0)
cyclopentenyl	(20,-5,0,-10,25)
cyclopentyl	(30,-30,30,-30,30)

In Table 4.10, the first part indicates bond length. The bond type is listed in left column and the right column gives the corresponding value for the bond length. The second part deals with the bond angle. The bond angle (right column) is determined by the hybridization of the parent atom (left column). The last part indicates the dihedral angle. The value of the dihedral angle is chosen from within the parenthesis based on a predefined search order and the properties of the current and parent atoms. The current atom is classified as non-cyclic atom and cyclic atom. For a non-cyclic atom, the current atom can be dependent or non-dependent upon the parent atom's hybridization. For a cyclic atom, the dihedral angle is determined by the ring type of the cyclic atom.

This built-in algorithm can provide fast conversion from a 2D BE matrix to a 3D coordinate geometry for regular hydrocarbon components without any external package. However, the structural data of this

algorithm is limited to the bonds between carbons and hydrogen. To extend this conversion to a generic context, a general algorithm that can cover various atom and bond types (e.g., C, S, N, O, H, etc.) is required. As it was not a goal of this thesis, an external package was employed to solve this problem.

Openbabel⁵ is a free open-sourced package that converts between over 110 chemical file formats. Openbabel has the internal data necessary to convert a 2D file format into a 3D coordinate file for most structures in the world. A Openbabel library written in the Python programming language is freely available for external development, and, was used for this thesis in order to obtain the 3D coordinates.

At first, a parser converts the BE matrix “.dat” file to a chemical markup language “.cml” file. The “.cml” file can then be converted to a MOPAC coordinate file using the Openbabel library. This external method not within the SPCM can handle the most molecules in complex feedstocks, but requires the installation of the Openbabel package.

4.2.3.2 Selected properties calculated by MOPAC

MOPAC uses the internal 3D coordinate files to calculate the critical thermodynamic properties of reaction. MOPAC is a general-purpose, semi empirical-molecular orbital package for the study of solid state and molecular structures and reactions. The semi-empirical Hamiltonians MNDO⁶, AM1⁷, PM3⁸, PM6, and MNDO-*d*^{9, 10}, etc. are used for the calculation of molecular orbitals in order to obtain the heat of formation and its derivative with respect to molecular geometry.

The heat of formation is the most useful pure component property with regard to reaction modeling. E.q.4.4 shows its calculation in MOPAC.

$$\Delta H_f = 23.061(E_{elect} + E_{nuc} - \sum_i E_{(isol-atom)}) \quad (4.4)$$

Where, E_{elect} is the electronic energy, E_{nuc} is the nuclear-nuclear repulsion energy, $-E_{isol}$ is the energy required to strip all the valence electrons off all the atoms in the system, and E_{atom} is the total heat of

atomization of all the atoms in the system. The index i is over all atoms in the system. The MNDO or PM3 method was selected for hydrocarbon components based on past experience.

Besides the heat of formation, other thermodynamic properties, such as heat capacity, entropy, and internal energy, can be calculated using the vibrational frequencies (energies), moments of inertia, symmetry number, and temperature. To calculate these properties, the FORCE method must be specified and the symmetry number of a component must be entered.

For the purpose of this thesis, the only calculation of the heat of formation can be automatically calculated by the SPCM. The other properties can be evaluated manually.

4.2.4 Miscellaneous properties: image and chemical name

An intuitive view of the molecule is very convenient in the development of a complex model. Instead of the tedious system of referential names (specie1, specie2, etc.), using a graphical 2D depiction of a molecule or even its IUPAC name can provide a highly efficient way to build and analyze composition and reaction models. Therefore, an automated algorithm to generate the 2D image of a molecule and its IUPAC name was developed for this thesis.

The mechanism of generating a molecule's image is analogous to the 3D structure parsing. In this case, Openbabel is used to generate a 2D coordinate file instead of a three dimensional one. That file can then be processed by other Python image libraries to create the graphical image of the molecule. This methodology was included in the SPCM.

The IUPAC name can be captured by another free, opened-source python library: webel. Webel uses the CML file to identify the molecule and then retrieves its IUPAC name from a free NIH web server¹¹.

In addition, CambridgeSoft's ChemDraw and ChemScript can be used to generate the graphical file and the IUPAC name respectively.

4.2.5 Supplemental property database

Although the combination of the above methods provides a strong basis for property estimation in SPCM, each approach has its own limitations. A supplemental scheme was built into the SPCM in order to provide an open, flexible database of property estimations. Database entries for specific properties can be entered or corrected based on other estimation methods, relevant literature or experimental data. If the built-in methods of the SPCM do not provide good estimations, the user can make note in the database that further evaluation is needed. Moreover, the property database of the SPCM allows for the avoidance of duplicated property estimation calculations. Simply looking up precalculated values will eliminate a large computational burden when dealing with large complex molecules (e.g., a resid's molecule). Although kinetic models vary based on chemistry and feedstock, the species in those systems are often duplicated. It is unnecessary to evaluate the properties of a molecule that has been calculated previously for an existing model. The database stores the calculated property information of a molecule the first time it occurs in any model. Any subsequent run of a model will obtain this information directly from the database rather than make the calculation again.

Any database should rely on a unique key value to represent each row of data. For this property database, the molecule's SIMILES code was chosen. SMILES code is based on the canonical graph representation of the molecule, and uses a system of implicit hydrogens. SIMLES is an open, public standard widely used in the field of chemical informatics.

All structural, thermodynamic, and miscellaneous properties are recorded in the database. An Excel-VBA based interface allows user to search, edit, update, and delete the information for molecules in the database.

4.2.6 Automation of SPCM for pure components

The SPCM flowsheet for pure components is shown in Figure 4.2. It contains two functional programs: PropGen and DictGen. PropGen is a C code program that is used to calculate the structural and thermodynamic properties of pure components by the group contribution method as discussed in the previous sections. DictGen is a Python language program which generates the 3D MOPAC input file, the 2D graphical depiction file, and the IUPAC name for pure components via external python libraries (e.g., Openbabel, Image Library etc.).

In Figure 4.2, each species in CME's MCA will first have its BE matrix and SIMLES code generated. Then the database is checked for the species existence.

If it already exists, the property information is obtained directly from the database directly and sent to CME for further composition modeling. If the molecule does not exist in the database, it is sent to PropGen for the calculation of the structural and selected thermodynamic properties. Next, the molecule is processed by DictGen in order to obtain the graphical representation, the IUPAC name and the three dimensional coordinate file used by MOPAC. MOPAC7 is then launched in order to calculate the final selected properties. Finally, The calculated properties are sent back to CME. The SPCM recursively follows this logic until all species in the MCA are traversed. Then, CME provides a user-friendly interface (Figure 4.3) for updating and correcting the property information before it is synchronize with the database. Finally, the database can be used by KME for LFER reaction indices, the heat of reactions and product properties.

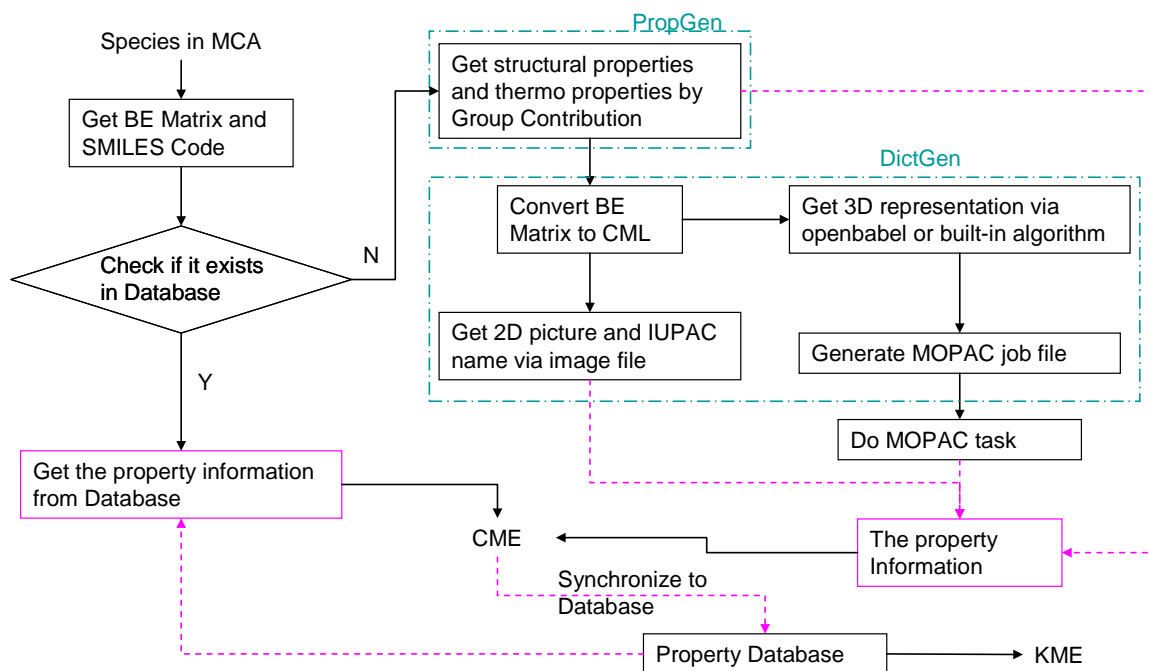


Figure 4.2 Structure-Property Correlations for pure Component

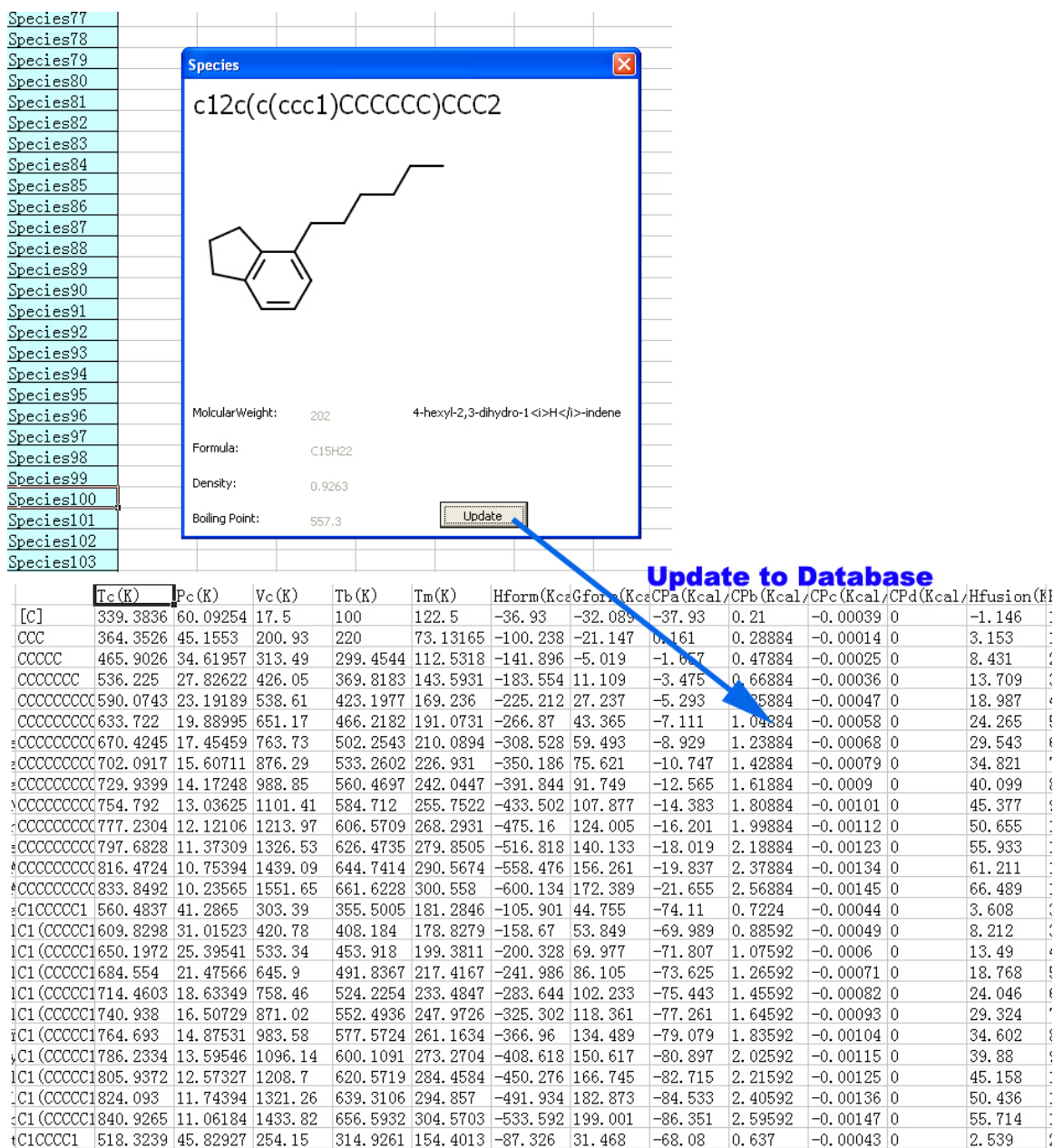


Figure 4.3 Synchronization of the Property Database via a user friendly interface

4.3 Bulk property estimation

Bulk properties serve as the objective function during MCA0 optimization in composition modeling, as well as the final product properties in reaction modeling. The estimation of bulk properties is based on

the properties for pure components in the system. The SPCM uses two classifications for those properties: built-in properties and user-supplied properties.

The built-in properties are a collection of widely used composition modeling properties such as molecular weight, H to C ratio, Aromatic Hydrogen fraction, Alpha Hydrogen fraction, the carbon fraction on aromatic rings (CA), the carbon fraction on naphthenic rings (CN), global SARA and PONA weight/mol fractions, global sulfur weight percent, global nitrogen weight content, density, and simulated distillation (volume or weight) and so on. Except for the simulated distillation data, all the properties can be calculated with a weight-based or mol-based mixing rule based on the pure components' properties. The data of simulated distillation is defined by the sorted set of boiling points for all the molecules in the system.

Not every feed stock process needs the same bulk properties. For example, gasoil uses a PONA table, while resid usually uses SARA fractions. Therefore, the SPCM provides different built-in property templates based on different feedstocks.

User-supplied properties attend the practical need for non-standardized measurements. Detailed modeling often contains thousands of species with in depth isomeric details. However, realistic data available to modeler is usually only of O (20-30). In these cases, lumps of species are simply created based on some given constraint. For example, IP8 is a lumped component for all isoparaffins with a carbon number of 8. That lump summarizes all the isomers variously branched. These lumped criteria can be specified via an intuitive Excel-based interface and will immediately serve as part of the bulk property calculation code for both the simulation and tuning modes of reaction modeling.

4.4 The reaction's thermodynamic properties of reaction

In reaction modeling, the enthalpy change of reaction (ΔH_{rxn}), the entropy change of reaction (ΔS_{rxn}), and the Gibbs energy change of reaction (ΔG_{rxn}) are very important, not only because serve as reaction indices for the LFER approximation, but also because they can be used to calculate the equilibrium constant and energy balance. The estimation of each of these three reaction properties is derived from the properties of the reactants and products shown by E.q.4.5.

$$\begin{aligned}\Delta H_{rxn} &= \sum \Delta H_{\text{product}} - \sum \Delta H_{\text{reactant}} \\ \Delta G_{rxn} &= \sum \Delta G_{\text{product}} - \sum \Delta G_{\text{reactant}} \\ \Delta S_{rxn} &= \frac{\Delta H_{rxn} - \Delta G_{rxn}}{T}\end{aligned}$$

$\Delta H_{\text{reactant}}$ is the heat of formation of each reactant
 $\Delta H_{\text{product}}$ is the heat of formation of each product
 $\Delta G_{\text{reactant}}$ is the Gibbs energy of formation of each reactant
 $\Delta G_{\text{product}}$ is the Gibbs energy of formation of each product

If the components in the reaction are not temperature-sensitive, the standard (298K) enthalpy of formation (ΔH_f^θ) and the standard Gibbs energy of formation (ΔG_f^θ) of the pure component can be used. Otherwise, a departure function can be applied to calculate the enthalpy of formation and the Gibbs energy of formation at the given temperature (E.q.4.6). The heat capacity is obtained from the Joback method.

$$\begin{aligned}
\Delta H_f^T &= \Delta H_f^\theta + \int_{298}^T C_p dT \\
\Delta G_f^T &= \Delta G_f^\theta + \int_{298}^T C_p dT \\
\Delta S_f^T &= \Delta S_f^\theta + \int_{298}^T \frac{C_p}{T} dT
\end{aligned}
\tag{4.6}$$

ΔH_f^θ is the standard heat of formation,

ΔG_f^θ is the standard Gibbs energy of formation

ΔS_f^θ is the standard entropy of formation at 298K

ΔH_f^T , ΔG_f^T , ΔS_f^T are the heat of formation, the Gibbs energy of formation and the entropy of formation at a given temperature

4.5 Summary and discussion

In this chapter, the structure-property correlation is discussed. The SPCM plays a fundamental role in supporting the composition modeling in CME as well as the kinetic modeling in KME.

This chapter addresses first the estimation of pure component properties. In this thesis, every molecule in the system is represented by a BE matrix, which is an explicit 2D representation of a molecule's structure. Structural related properties are calculated directly from the BE matrix in a straightforward process. Thermodynamic properties are evaluated by group contribution methods and quantum chemistry software. Two group contribution methods were discussed in this chapter. The Joback method was applicable to light feedstocks, and the Gani method with its 370 detailed structural groups, was applicable to heavier complex feedstocks. Then the chapter turned to the application of the quantum chemistry tool, MOPAC. In order to fulfill MOPAC's calculation requirements, a conversion from a 2D to a 3D structure was addressed. At the same time, the generation of the IUPAC name and a 2D image of each species were discussed. Finally, a database of pure component properties was discussed and its logic was outlined.

The prediction of pure component density was noted as sub-par compared with the other estimated physical properties. This deficiency should be addressed by the future incorporation of other advanced methods or by user correction at the database level. For the properties calculated by MOPAC, only the heat of formation is automatically to the KMT suite. Other properties such as Gibbs energy, heat capacity, and entropy could be integrated into the SPCM in a future revision.

The calculation of bulk properties is discussed based on the estimation of the pure components' properties. Besides the predetermined built-in properties, the SPCM allows the modeler to specify specific lumping strategies for reaction modeling.

Finally, this chapter is illustrated how to calculate the thermodynamic properties of reaction from the pure components' properties.

Reference

-
- ¹ K.G. Joback, R.C. Reid, *Chem. Eng. Comm.* 57 (1987) 233–243
- ² Marrero and R. Gani, *Fluid Phase Equil.* 183/184 (2001), pp. 183–208.
- ³ <http://www.openmopac.net/index.html>
- ⁴ Broadbelt, L. J., Stark, S. M., and Klein, M. T., *Chem. Eng. Sci.*, 49, pp. 4991-5101 (1994a)
- ⁵ http://openbabel.org/wiki/Main_Page
- ⁶ M. J. S. Dewar and W. Thiel. *J. Am. Chem. Soc.*, 99:4899-4907, 1977.
- ⁷ M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart. *J. Am. Chem. Soc.*, 107:3902-3909, 1985.
- ⁸ J. J. P. Stewart. *J. Comp. Chem.*, 10:209-220, 1989.
- ⁹ W. Thiel and A. Voityuk. *Theor. Chim. Acta*, 81:391-404, 1992.
- ¹⁰ W. Thiel and A. Voityuk. *J. Phys. Chem.*, 100:616-629, 1996.
- ¹¹ <http://cactus.nci.nih.gov/chemical/structure>

Chapter 5 Composition modeling of selected complex feedstocks

5.1 Introduction

Chapter 2 discussed how a set of analytical data could be transformed into a statistical representation of a complex feedstock thereby obtaining the MCA0, the set of initial concentration values necessarily for kinetics modeling. The goal of this chapter is to validate this approach by applying it to a variety of feedstocks. This methodology has been successfully applied to petroleum middle distillates, natural gas condensates, petroleum heavy gas oil (HGO), and petroleum vacuum gas oil (VGO), and petroleum resids.

5.2 Petroleum middle distillates

In order to validate the hybrid approach of representing complex feedstocks statistically, a case study of a light petroleum fraction-middle distillate was performed. Both bulk properties and detailed composition data were measured for this fraction. Bulk properties were used as the optimization criteria, the detailed composition data was used for model validation. The analytical data, the CME generated composition model, and the model results will be discussed in detail in the following sections.

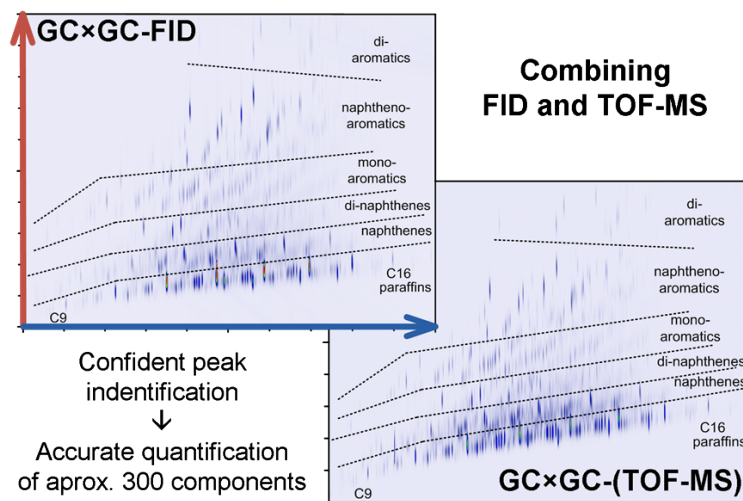
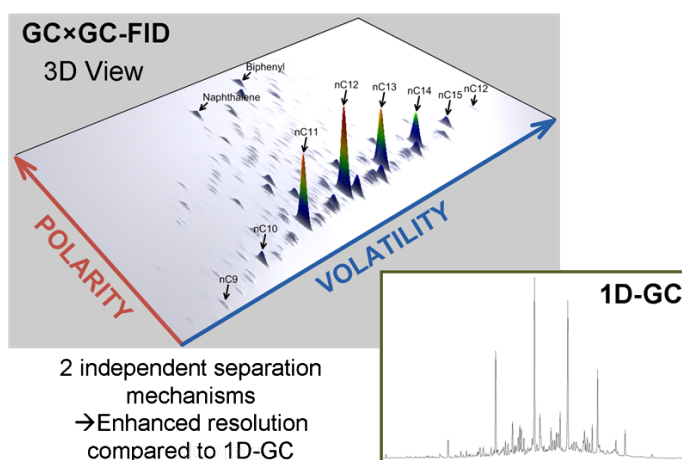
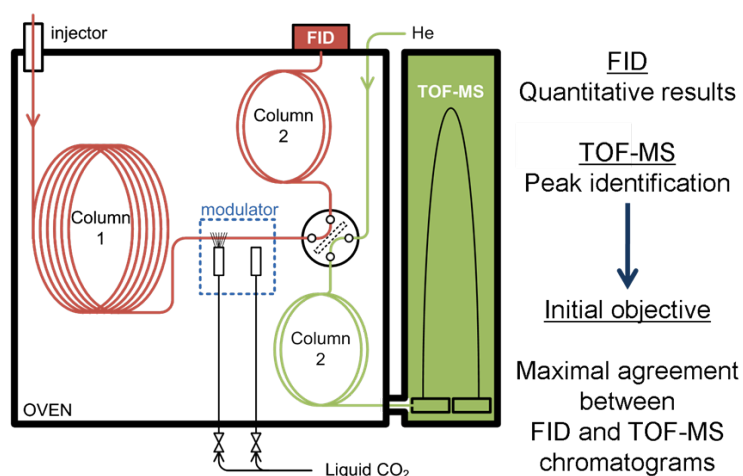
5.2.1 Analytical characterization

This modeling method was applied to two dissimilar middle distillates. The analytical properties for these two samples were provided by Ghent University¹. The overall weight PIONA (paraffin, isoparaffin, olefin, naphthenic, and aromatic) fraction, density (g/cm^3) and ASTM D2887 simulated distillation curve (K) were used as the objective function to optimize the attribute PDF parameters for the feedstocks are summarized in Table 5.1.

Table 5.1 Bulk properties for petroleum middle distillates

	Sample1	Sample2
WTP_frac:	0.143	0.235
WTI_frac:	0.211	0.254
WTO_frac:	0.000	0.000
WTN_Frac:	0.458	0.341
WTA_frac:	0.188	0.171
Density(g/cm3):	0.7987	0.7983
IBP(K):	386	422
5BP(K):	408	455
10BP(K):	416	466
20BP(K):	438	475
30BP(K):	450	483
40Bp(K):	462	489
50BP(K):	475	498
60BP(K):	487	503
70BP(K):	498	511
80BP(K):	515	521
90BP(K):	533	533
95Bp(K):	543	541
EBP(K):	577	561

The detailed composition data is provided by a comprehensive 2D Gas Chromatography (GC)² that can analyze over 300 components with high accuracy. The detailed description is shown as Figure 5.1. The detailed data is shown in a later section for model validation.

Figure 5.1 Comprehensive 2D GC²

5.2.2 Homologous series determination

For these two samples, the core series were explicitly determined. Their simulated distillation data indicated the max ring number of aromatic cores in these samples is two because the end boiling point cut was less than 600 K. Based on analytical literature and previous experience, five saturates and five aromatics structures were specified for this model as shown in Figure 5.2.







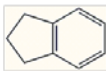
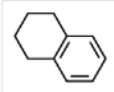
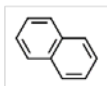
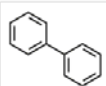
					Saturates
					Aro#1
					Aro#2

Figure 5.2 Selected homologous series for middle distillates

5.2.3 Structural attribute sampling protocol and PDF form

The quantitative optimization was tested as two cases by two different structural attribute sampling protocols and attribute PDF forms as discussed in Chapter 2. In case 1, the feedstock was modeled as a set of homologous series with side chain information. The global fractions of these series were set as discrete values and a gamma distribution was imposed as the attribute PDF for carbon number and the number of side chain of each series. In case 2, the set of homologous series was decomposed to several elemental structural attributes, and the feedstock was sampled from multiple elemental structural attributes covering the core series' structure and side chain information. Figure 5.3 shows the specific sampling protocol for this case based on the general sampling protocol mentioned in Chapter 2. Both feedstocks were modeled by each of the above two cases.

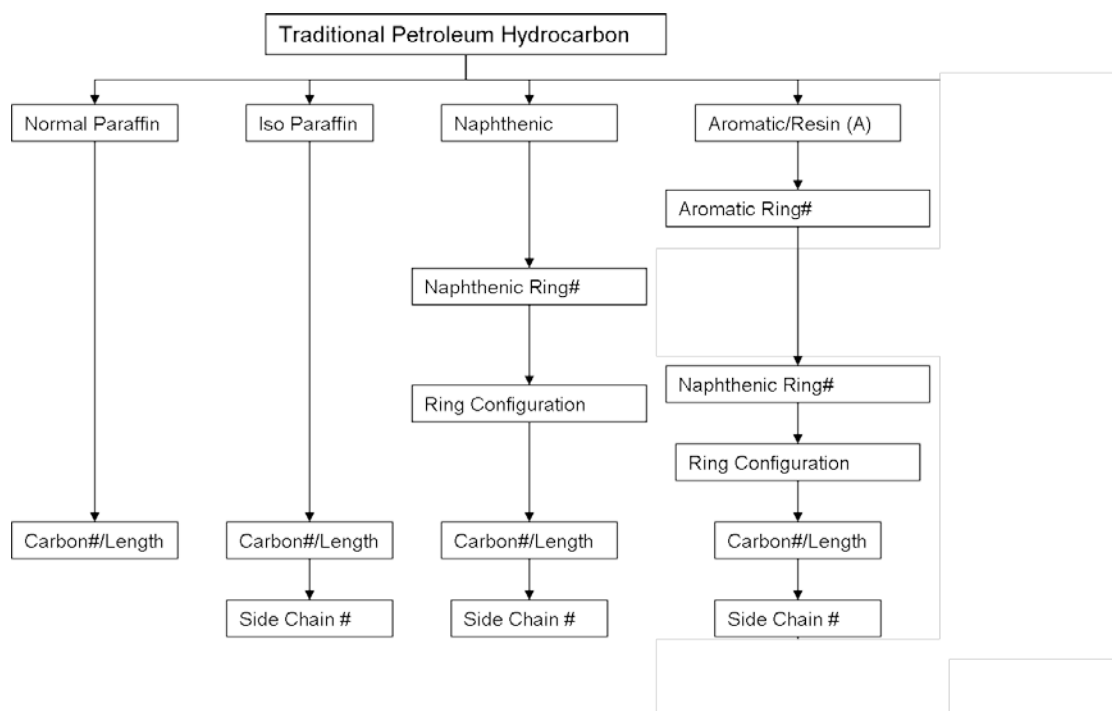


Figure 5.3 sampling protocol for middle distillates

5.2.4 Conditional probability

In order to ensure physically realistic molecules, conditional probability distributions were used. Since the petroleum middle distillates were bounded by both an initial boiling point and a final boiling point, it was necessary to ensure molecules that were too small or too large were not considered. By constraining the boiling point cut range in CME for each series, the range of the carbon number was determined via structure property correlation before the optimization. Based on findings in the literature³, iso-paraffin and single ring components were limited to one, two or three side chains. Two ring components were limited to only one or two side chains.

In addition to the overall carbon constraints, it was necessary to add conditional probability constraints on the number and length of sidechains for the isoparaffin, naphthenic, and aromatic fractions. For instance, if a molecule contained only one sidechain carbon, then only one sidechain could possibly be

attached to the molecule core. Furthermore, since each sidechain contains at least one carbon, any one sidechain can be no longer than the total number of sidechain carbons minus the number of sidechains other than itself.

5.2.5 Optimization of the PDF parameters

The PDF parameters for selected petroleum middle distillates were optimized using a global simulated annealing technique. The chi square objective function of Eq. 5.1 was a combination of various molecular properties that could be easily compared to experimental data:

$$\begin{aligned} \chi^2 = & \left(\frac{Density_{\text{exp}} - Density_{\text{pred}}}{Weight_{\text{Density}}} \right)^2 \\ & + \left(\frac{1}{\#Comps} \right) \sum_{i=1}^{\#Comps} \left(\frac{PIONAWt_{\text{exp}} - PIONAWt_{\text{pred}}}{Weight_{\text{PIONA}}} \right)^2 \\ & + \left(\frac{1}{\#Fracs} \right) \sum_{i=1}^{\#Fracs} \left(\frac{SimDis_{\text{exp}} - SimDis_{\text{pred}}}{Weight_{\text{SimDis}}} \right)^2 + \dots \end{aligned} \quad (5.1)$$

The numerator is the square of the difference between the model prediction and the measurement for calibration. The denominator is a weighting factor analogous to the standard deviation of an experimentally determined value.

5.2.6 Validation results

The model predictions for the petroleum middle distillates are herein discussed for each of the two sampling protocols respectively: case 1 and case 2.

5.2.6.1 Case 1 validation results

The optimized results of two middle distillate samples with case 1 are shown in Table 5.2. The distribution of the number of side chain was excluded from tuning, and was given a fixed distribution

based on Ranzi³ as shown in Table 5.3. The branch point locations were assumed to follow a uniform distribution. For the most part, the model predictions were accurate to within an experimental standard deviation.

Table 5.2 Model predictions for the analytical properties of two middle distillates with case 1.

	Sample1		Sample2	
	Calc	Exp	Calc	Exp
WTP_frac:	0.143	0.143	0.238	0.235
WTI_frac:	0.211	0.211	0.252	0.254
WTO_frac:	0.000	0.000	0.000	0.000
WTN_Frac:	0.458	0.458	0.338	0.341
WTA_frac:	0.188	0.188	0.172	0.171
Density(g/cm3):	0.8174	0.7987	0.8249	0.7983
IBP(K):	385	386	422	422
5BP(K):	406	408	455	455
10BP(K):	422	416	465	466
20BP(K):	436	438	475	475
30BP(K):	451	450	483	483
40Bp(K):	462	462	492	489
50BP(K):	475	475	495	498
60BP(K):	487	487	502	503
70BP(K):	499	498	512	511
80BP(K):	515	515	521	521
90BP(K):	533	533	533	533
95Bp(K):	544	543	542	541
EBP(K):	580	577	560	561

Table 5.3 Typical weight distribution of branched isomers of C₈³

Isomers	Ponca	Occidental	Texas	Internalweights
2-Methylheptane	46.3	36.9	42.1	45.8
3-Methylheptane	15.4	28.5	23.4	22.9
4-Methylheptane	10.3	10.2	9.3	11.5
2,3-Dimethylhexane	3.6	5.4	6.3	3.4
2,4-Dimethylhexane	3.1	5.5	4.2	3.4
2,5-Dimethylhexane	3.1	5.7	4	3.4
3,4-Dimethylhexane	6.7	2.6	3.7	3.4
2,2-Dimethylhexane	0.5	–	0.3	–
3,3-Dimethylhexane	1.5	1.7	0.4	–
2,3,4-Trimethylpentane	0.3	–	1.1	1.2
2,2,3-Trimethylpentane	0.2	–	–	–
2,3,3-Trimethylpentane	0.3	–	0.6	–
3-Ethylhexane	4.6	3.5	3.1	3.8
2-Methyl-3-ethylpentane	3.1	–	1.5	1.2
3-Methyl-3-ethylpentane	1	–	–	–

Table 5.2 shows good model correlation for both of the sample feedstocks, with the model for Sample 1 slightly outperforming that of Sample2. Each of the analytical properties for these two samples was predicted to within one experimental standard deviation except for density. The reason for the larger discrepancy in density may be caused its less adequate structure-property estimation. However, the accuracy of the PIONA weight percents and the simulated distillation curve shows more persuasive proof about the model's precision.

As an additional validation, the experimentally determined and predicted carbon number distributions corresponding to each PIONA fraction, were compared via detailed composition data. The overall statistical analysis between calculations and experimental data is listed in Table 5.4.

Table 5.4 the overall statistical analysis for middle distillate Sample1 and Sample2 with case 1

	Sample1	Sample2
	Deviation	Deviation
avg	0.3918	0.5085
std.	0.4992	0.6842

Figure 5.4 shows the experimentally determined and predicted carbon number distribution for the paraffin fraction of middle distillate Sample1 with case 1. The predicted curve matches the experimentally determined curve decently well. The predicted values between carbon numbers 10 to 15 are slightly higher than those of the analytical data.

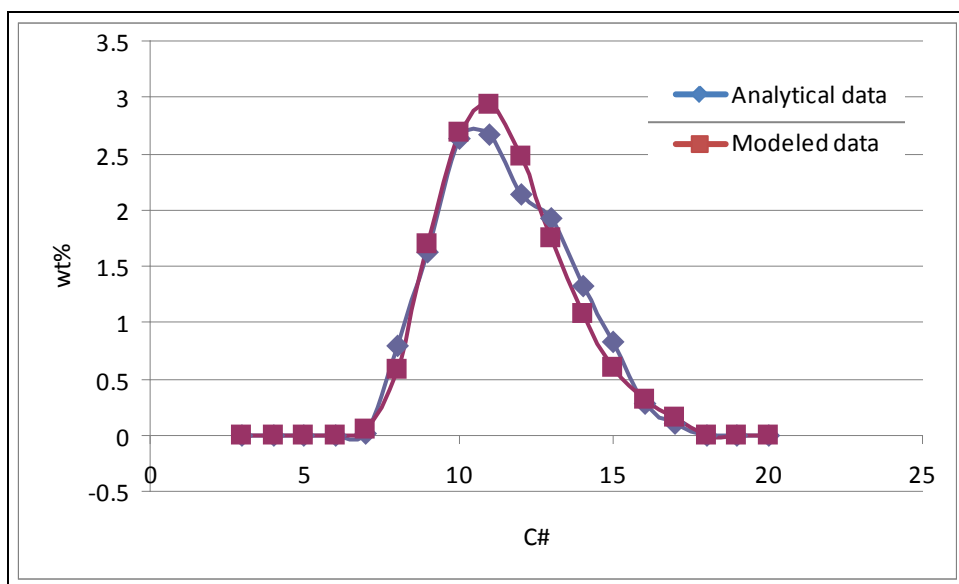


Figure 5.4 Predicted and experimental carbon number distribution for the paraffin fraction of middle distillate Sample1 with case 1.

Figure 5.5 shows the predicted and experimentally determined carbon number distributions for the isoparaffin fraction of middle distillate Sample1 with case 1. The predicted curve matches the experimentally determined curve moderately well. The curve of analytical data is not ideally smooth. The predicted values for carbon number larger than 10 show a deviation from the analytical data curve. The values for carbon numbers 10-14 are overestimated, while the values of carbon number from 15 onward end are underestimated. Since the deviation shows an opposite symmetry trend, the influences are cancelled out and only contribute a small effect to the overall weight fraction in objective function.

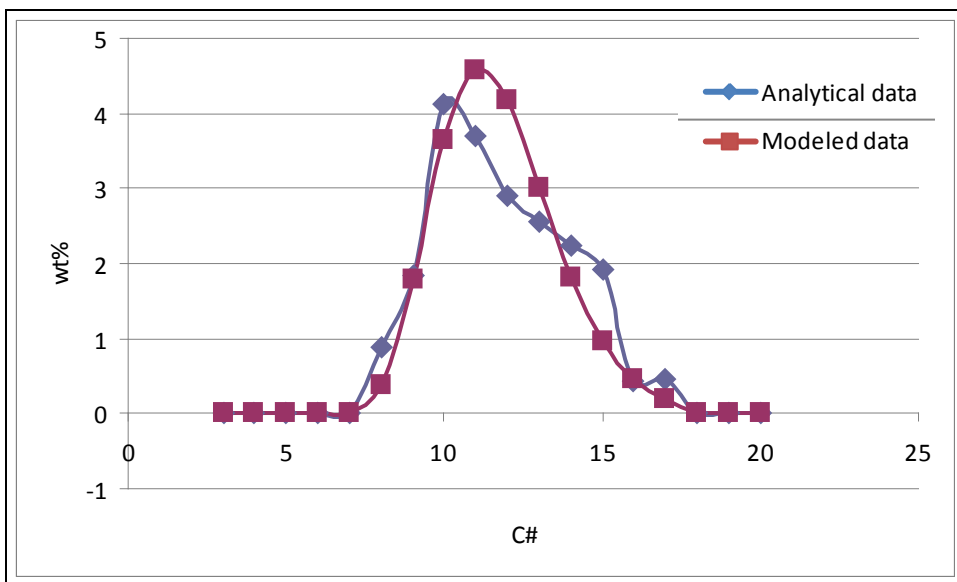


Figure 5.5 Predicted and experimental carbon number distribution for the isoparaffin fraction of middle distillate Sample1 with case 1.

Figure 5.6 shows the predicted and experimentally determined carbon number distribution for the naphthenic fraction of middle distillate Sample1 with case 1. The predicted curve matches the experimentally determined curve moderately well. The modeled data shows a sharper curve than analytical data. The predicted values for carbon number larger than 10 show a deviation from the analytical data curve. But opposite to isoparaffin case, the values for carbon numbers 10-14 are underestimated, while the values for carbon number from 15 onward end are overestimated. Since the deviation also shows as an opposite symmetry trend, the influences are cancelled out and only contribute a small effect to the overall weight fraction in the objective function.

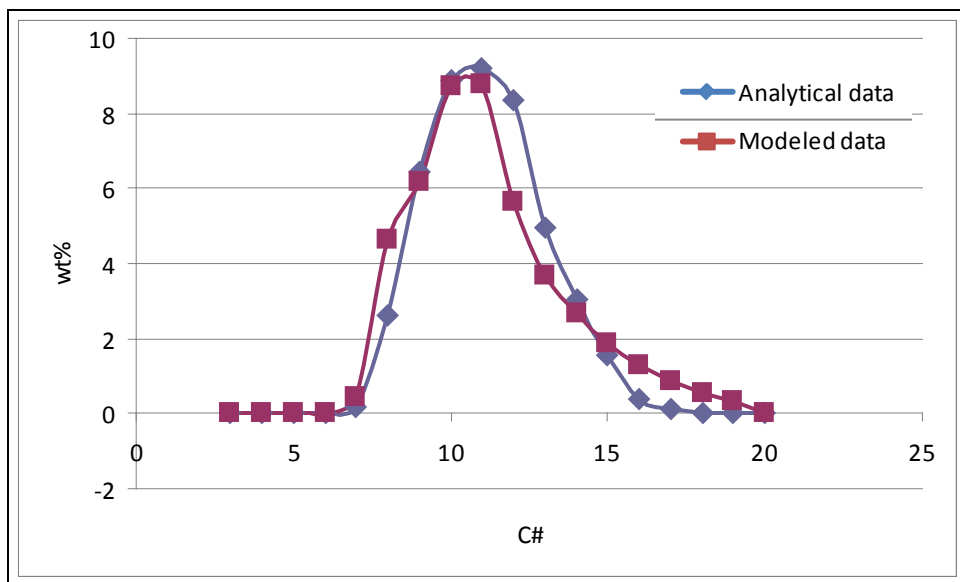


Figure 5.6 Predicted and experimental carbon number distribution for the naphthenic fraction of middle distillate Sample1 with case 1.

Figure 5.7 shows the predicted and experimentally determined carbon number distribution for the aromatic fraction of middle distillate Sample1 with case 1. The predicted curve does not show a good match with the experimentally determined curve. The modeled data nearly matches the position of the peak with the analytical data; however, the curve is quite sharper, and the values around the peak are overestimated. Since case 1 uses discrete values for the global fractions of each aromatic core, a constraint on these values may need to be imposed by adding additional analytical data for tuning the weight compositions of the aromatic rings. Another solution may be to use a structural attribute PDF on aromatic ring number, as in case 2.

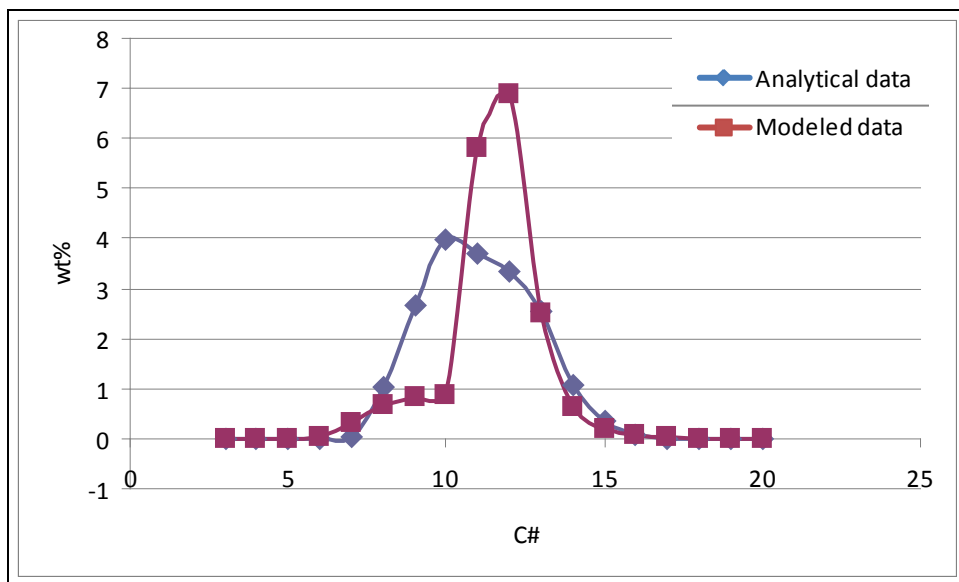


Figure 5.7 Predicted and experimentally determined carbon number distributions of the aromatic fraction of middle distillate Sample1 with case 1.

Figure 5.8 shows the experimentally determined and predicted carbon number distribution for the paraffin fraction of middle distillate Sample2 with case 1. The predicted curve matches the experimentally determined curve moderately well. The predicted values close to the peak are underestimated.

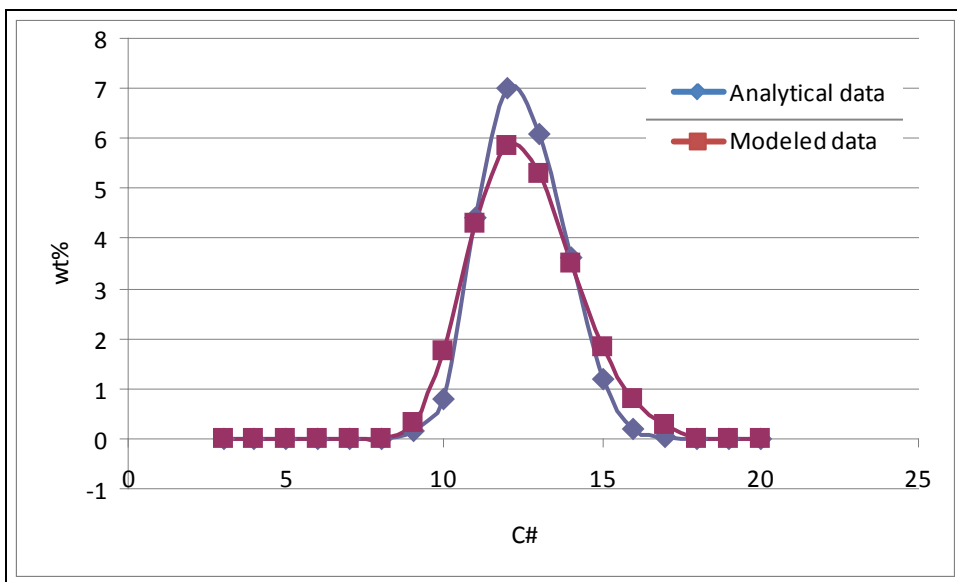


Figure 5.8 Predicted and experimental carbon number distribution for the paraffin fraction of middle distillate Sample2 with case 1.

Figure 5.9 shows the predicted and experimentally determined carbon number distributions for the isoparaffin fraction of middle distillate Sample2 with case 1. The predicted curve matches the experimentally determined curve decently well.

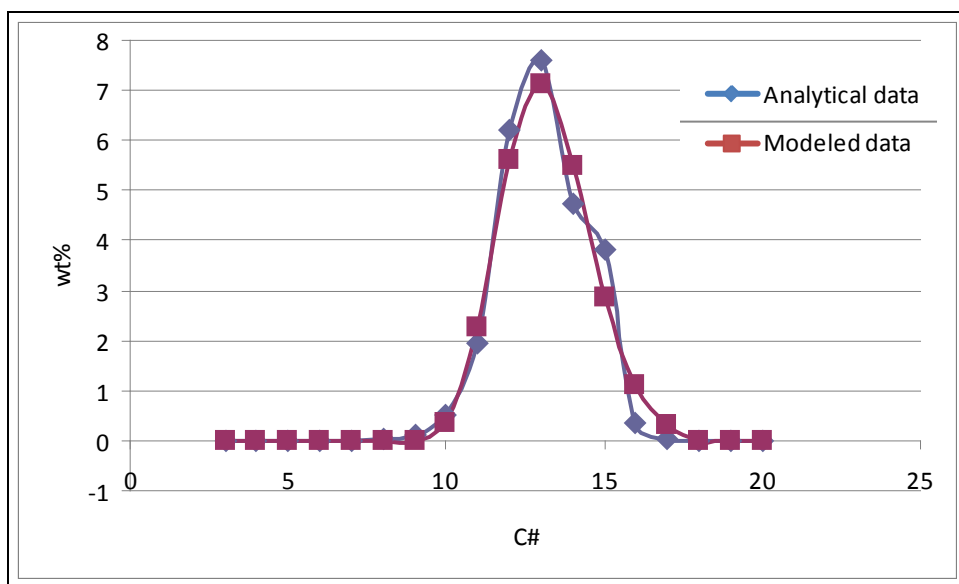


Figure 5.9 Predicted and experimental carbon number distribution for the isoparaffin fraction of middle distillate Sample2 with case 1.

Figure 5.10 shows the predicted and experimentally determined carbon number distribution for the naphthenic fraction of middle distillate Sample2 with case 1. The predicted curve matches the experimentally determined curve moderately well. The modeled data shows a bit of a right-shift to the analytical data, and its peak value is underestimated.

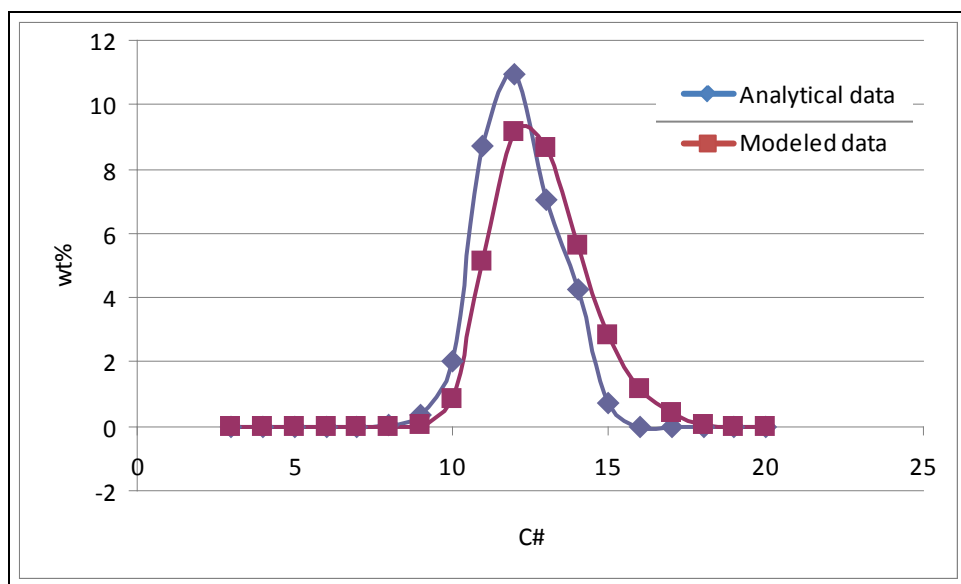


Figure 5.10 Predicted and experimental carbon number distribution for the naphthenic fraction of middle distillate Sample2 with case 1.

Figure 5.11 shows the predicted and experimentally determined carbon number distribution for the aromatic fraction of middle distillate Sample2 with case 1. The predicted curve does not show a good match with the experimentally determined curve. The modeled data shows a obvious right-shift to the analytical data close to the peak area, but the height of the peak area is consistent. Since case A uses discrete values for the global fractions of each aromatic core, a constraint on their values may need to be imposed by adding additional analytical data for tuning the weight compositions of the aromatic rings. Another solution may be to use a structural attribute PDF on the aromatic ring number, as in case 2.

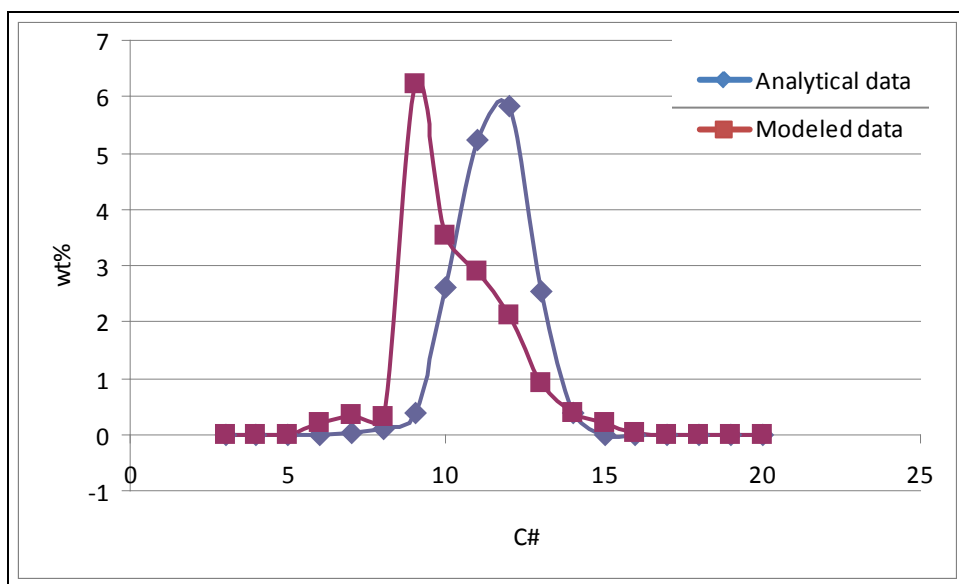


Figure 5.11 Predicted and experimentally determined carbon number distributions of the aromatic fraction of middle distillate Sample2 with case 1.

5.2.6.2 Case 2 validation results

The optimized results of two middle distillate samples with case 2 are shown in Table 5.5. The distribution of the number of side chain was excluded from tuning, and was given a fixed distribution based on Ranzi³ as shown in Table 5.3. The branch point locations were assumed to follow a uniform distribution. For the most part, the model predictions were accurate to within an experimental standard deviation.

Table 5.5 Model predictions for the analytical properties of two middle distillates with case 2.

	Sample1		Sample2	
	Calc	Exp	Calc	Exp
WTP_frac:	0.143	0.143	0.234	0.235
WTI_frac:	0.211	0.211	0.254	0.254
WTO_frac:	0.000	0.000	0.000	0.000
WTN_Frac:	0.458	0.458	0.341	0.341
WTA_frac:	0.188	0.188	0.171	0.171
Density(g/cm3):	0.8081	0.7987	0.8086	0.7983
IBP(K):	385	386	422	422
5BP(K):	411	408	455	455
10BP(K):	422	416	465	466
20BP(K):	440	438	476	475
30BP(K):	454	450	484	483
40Bp(K):	465	462	489	489
50BP(K):	475	475	498	498
60BP(K):	487	487	505	503
70BP(K):	498	498	512	511
80BP(K):	512	515	521	521
90BP(K):	530	533	533	533
95Bp(K):	544	543	542	541
EBP(K):	577	577	563	561

Table 5.5 shows good model correlation for both of the sample feedstocks. Each of the analytical properties for these two samples was predicted to within one experimental standard deviation except for density. The reason for this discrepancy in density may be its less adequate structure-property estimation method. However, the accuracy of PIONA weight percents and the simulated distillation curve shows more persuasive proof about the model's precision.

As an additional validation, the experimentally determined and predicted carbon number distributions corresponding to each PIONA fraction were compared via detailed composition data. The overall statistical analysis between calculations and experimental data is listed in Table 5.6.

Table 5.6 the overall statistical analysis for middle distillate Sample1 and Sample2 with case 2.

	Sample1	Sample2
	Deviation	Deviation
avg	0.2015	0.5118
std.	0.2946	0.4790

Figure 5.12 shows the experimentally determined and predicted carbon number distribution for the paraffin fraction of middle distillate Sample1 with case 2. The predicted curve matches the experimentally determined curve decently well. The predicted values close to the peak area are higher than the analytical data.

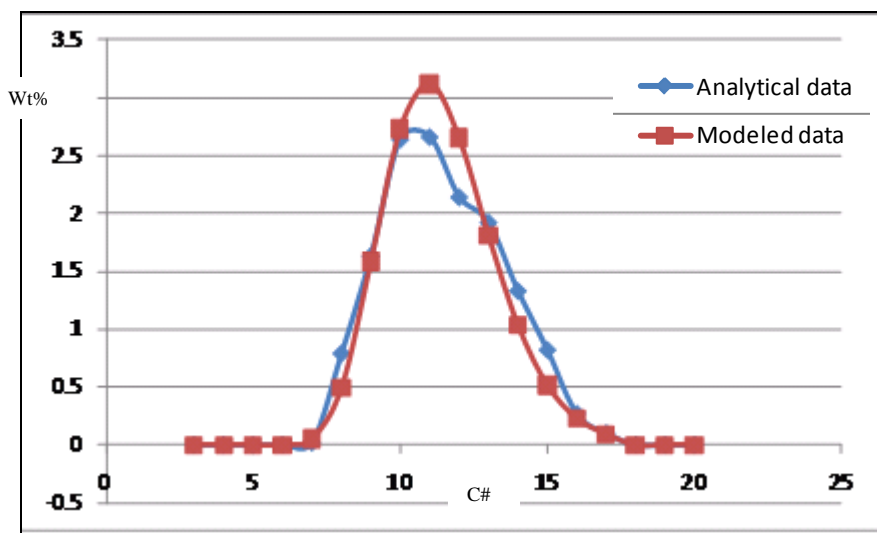


Figure 5.12 Predicted and experimental carbon number distribution for the paraffin fraction of middle distillate Sample1 with case 2.

Figure 5.13 shows the predicted and experimentally determined carbon number distributions for the isoparaffin fraction of middle distillate Sample1 with case 2. The predicted curve matches the experimentally determined curve moderately well. The curve of analytical data is not ideally smooth. The predicted values for carbon number larger than 10 shows a deviation from the analytical data curve. The values for carbon numbers 10-14 are overestimated, while the values of carbon number from 15 onward end are underestimated. Since the deviation also shows an opposite symmetry trend, the

influences are cancelled out and only contribute a small effect to the overall weight fraction in the objective function.

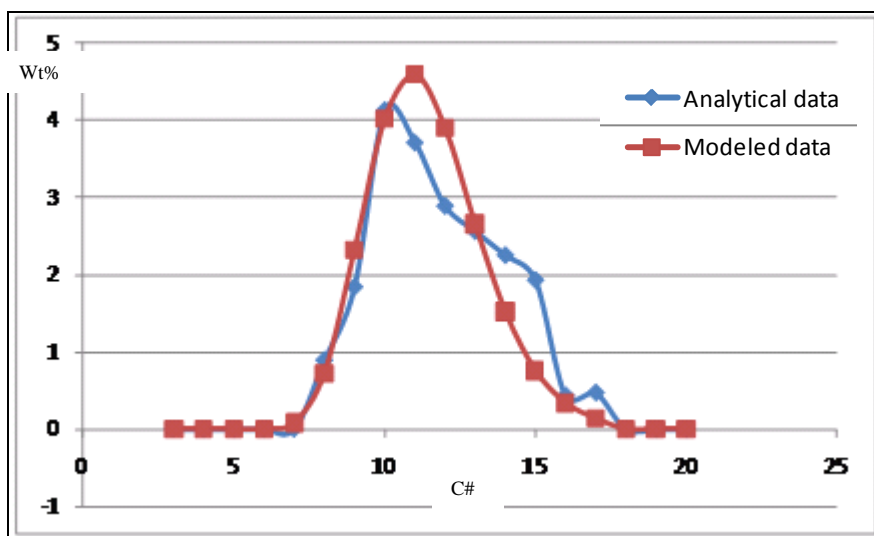


Figure 5.13 Predicted and experimental carbon number distribution for the isoparaffin fraction of middle distillate Sample1 with case 2.

Figure 5.14 shows the predicted and experimentally determined carbon number distribution for the naphthenic fraction of middle distillate Sample1 with case 2. The predicted curve matches the experimentally determined curve decently well. The predicted value at the peak is a bit higher than the analytical data.

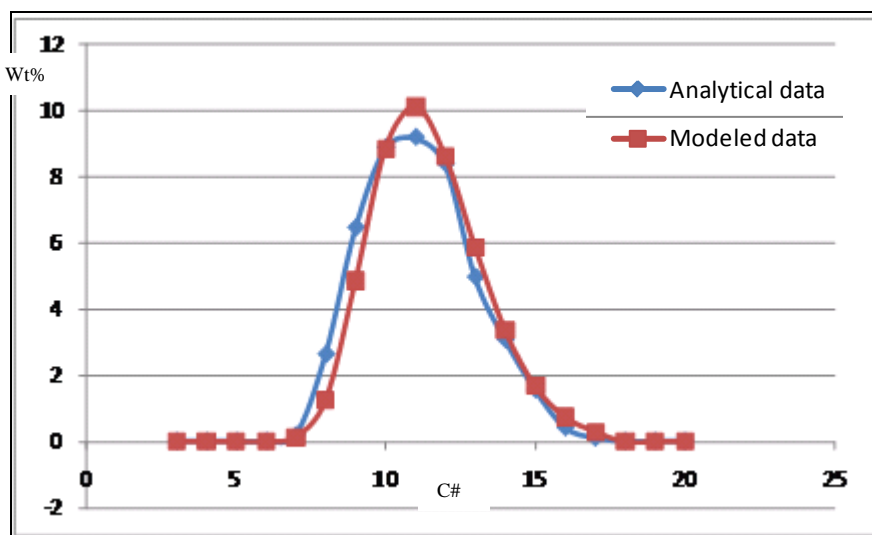


Figure 5.14 Predicted and experimental carbon number distribution for the naphthenic fraction of

middle distillate Sample1 with case 2.

Figure 5.15 shows the predicted and experimentally determined carbon number distribution for the aromatic fraction of middle distillate Sample1 with case 2. The predicted curve matches the experimentally determined curve moderately well. The curve of the modeled data is a bit sharper and higher than the analytical data close to the peak. Compared to the model of middle distillate Sample1 with case 1, the results are attained quite an improvement.

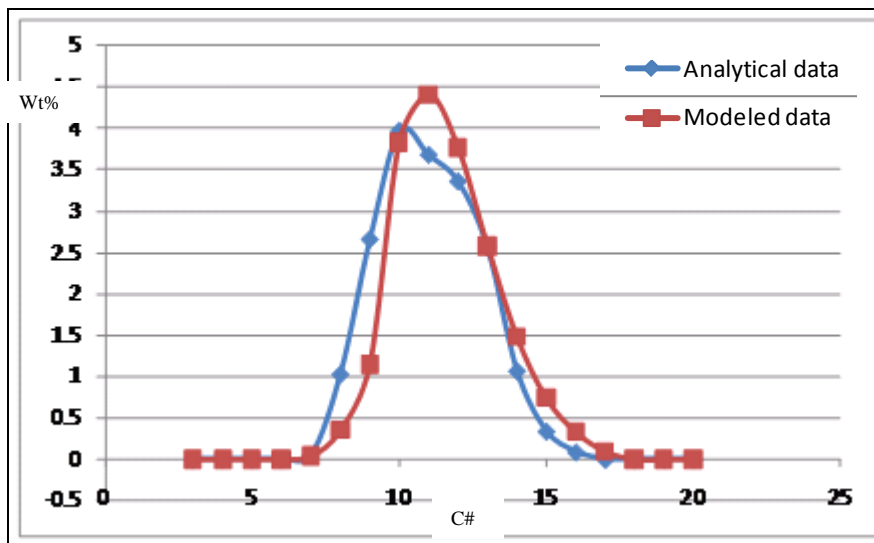


Figure 5.15 Predicted and experimentally determined carbon number distributions of the aromatic fraction of middle distillate Sample1 with case 2.

Figure 5.16 shows the experimentally determined and predicted carbon number distribution for the paraffin fraction of middle distillate Sample2 with case 2. The predicted curve matches the experimentally determined curve decently well. The peak area of the predicted values was a bit of a right-shift compared to the analytical data.

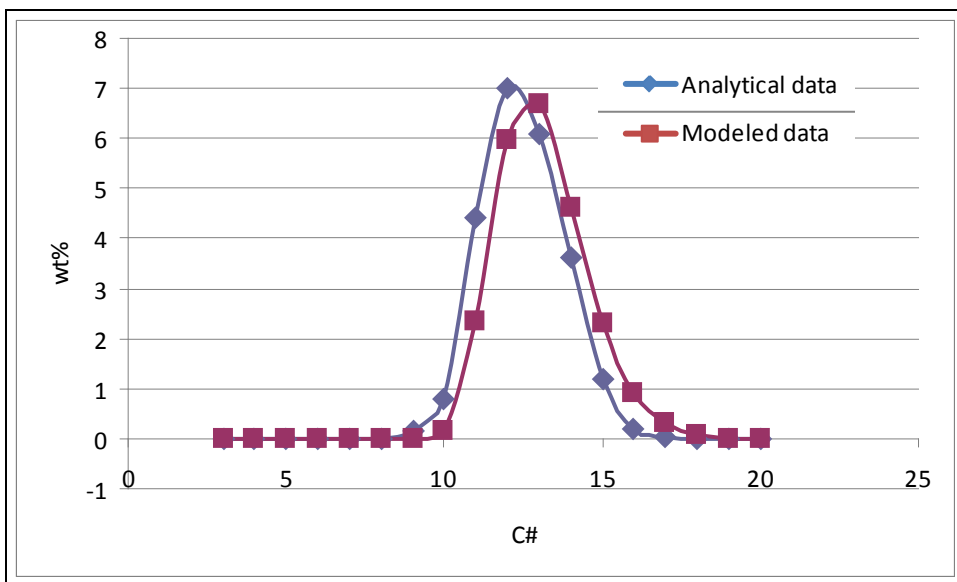


Figure 5.16 Predicted and experimental carbon number distribution for the paraffin fraction of middle distillate Sample2 with case 2.

Figure 5.17 shows the predicted and experimentally determined carbon number distributions for the isoparaffin fraction of middle distillate Sample2 with case 2. The predicted curve matches the experimentally determined curve decently well although the curve of analytical data is not ideally smooth.

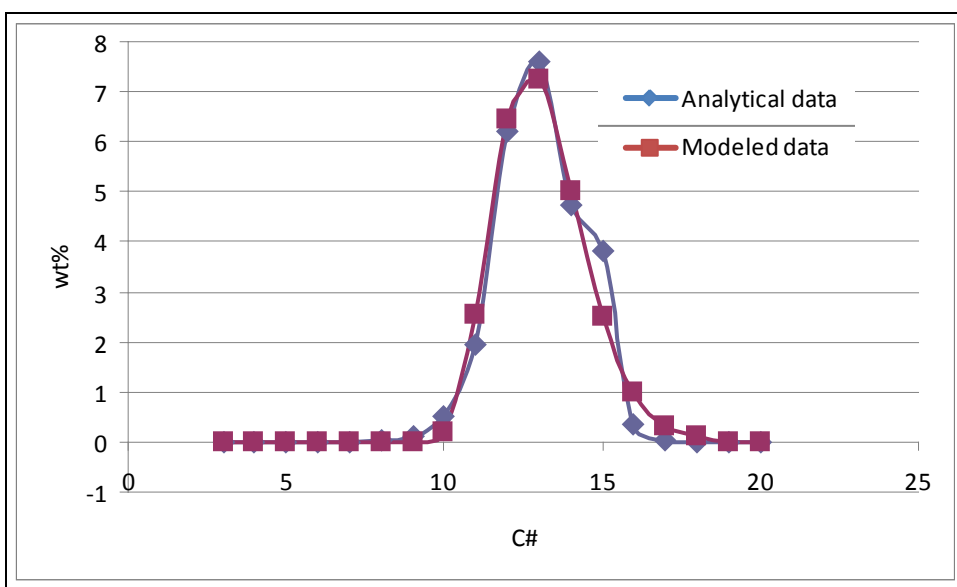


Figure 5.17 Predicted and experimental carbon number distribution for the isoparaffin fraction of

middle distillate Sample2 with case 2.

Figure 5.18 shows the predicted and experimentally determined carbon number distribution for the naphthenic fraction of middle distillate Sample2 with case 2. The predicted curve matches the experimentally determined curve moderately well. The predicted values close to the peak are lower than the analytical data and show a bit of a right-shift.

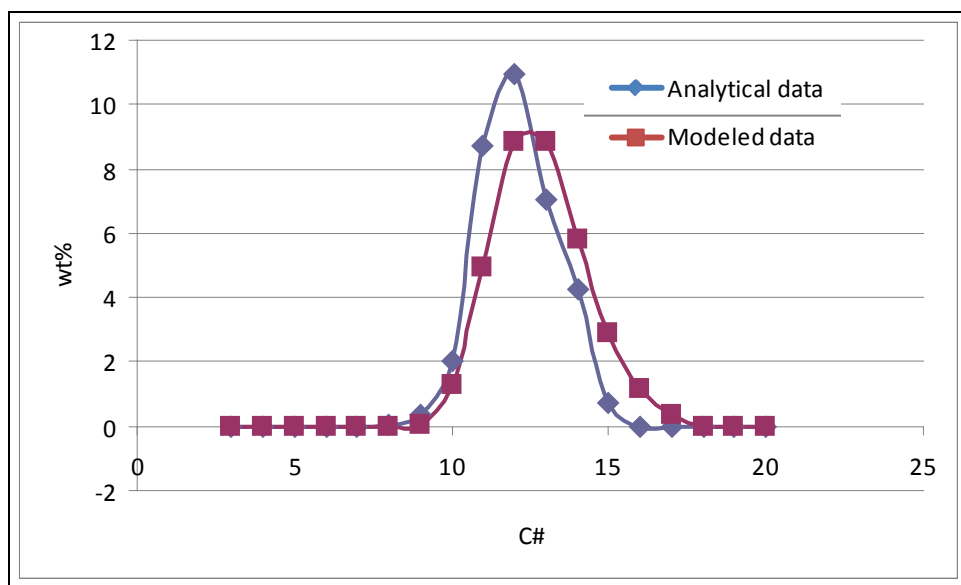


Figure 5.18 Predicted and experimental carbon number distribution for the naphthenic fraction of middle distillate Sample2 with case 2.

Figure 5.19 shows the predicted and experimentally determined carbon number distribution for the aromatic fraction of middle distillate Sample2 with case 2. The predicted curve matches the experimentally determined curve acceptably. The curve of the modeled data is a bit lower than the analytical data and shows a left-shift close to the peak. However, compared to the model of middle distillate Sample2 with case 1, the results are quite an improvement.

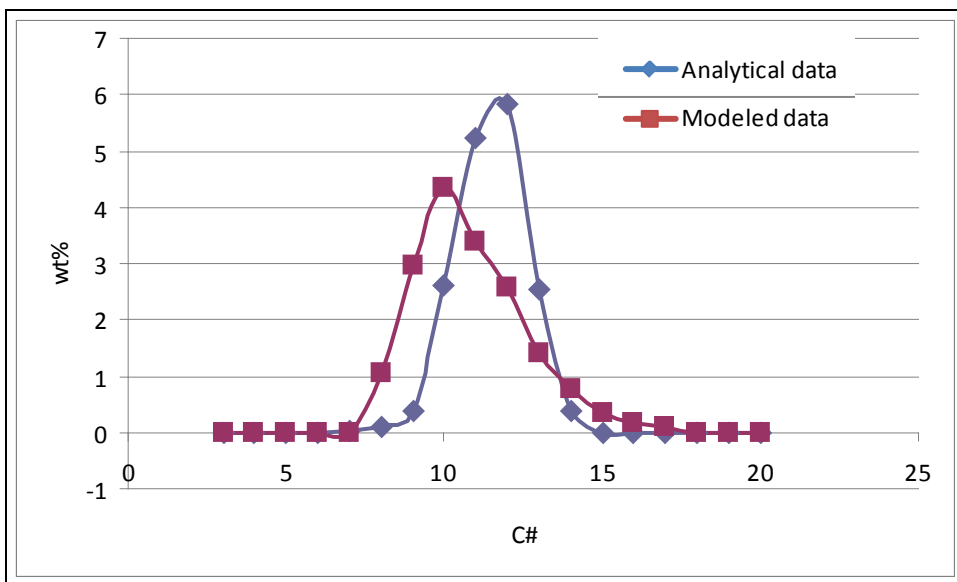


Figure 5.19 Predicted and experimentally determined carbon number distributions of the aromatic fraction of middle distillate Sample1 with case 2.

5.2.6.3 Conclusion

The validation results for two samples using two differently sampling cases show an overall good fit for the petroleum middle distillates. For both cases, the shape of the carbon number distribution for both samples can be captured so that the continuous attribute PDFs (e.g, gamma) can describe the side chain information quite well. For case1, using discrete values for the global fractions of each core series showed some limitations, especially with regard to the carbon number distributions for aromatic compositions. Additional measurements may need to be worked into the objective function in order to obtain a better fit. The method of case 2 naturally constrains these core series' fractions, so it obtained an overall good fit for each of the calibrated items. In conclusion, the method of case 2 requires less analytical information than case 1 for calibration, and thus has more compatibility for heavier complex feedstocks (which generally have limited analytical information).

5.3 Natural gas condensates

In addition to validating the modeled results of a petroleum middle distillate, this method was also applied to the heavier conversion produced by natural gas condensates. Bulk properties were used as the optimization targets. Unlike petroleum middle distillates, detailed composition data can not be obtained for this fraction, but GC-MS data can be used for the model validation. The analytical information, the composition model, and the model results will be discussed in detail in the following sections.

5.3.1 Analytical characterization

This modeling method was applied to two dissimilar natural gas condensates. The analytical properties for these two samples were provided by Ghent University¹. The properties used in the objective function were similar to those of the petroleum middle distillates, and are summarized in Table 5.7. The GC-MS data is shown in a later section for model validation.

Table 5.7 Bulk properties for natural gas condensates

	GC663	GC659
WTP_frac:	0.175	0.143
WTI_frac:	0.325	0.149
WTO_frac:	0.000	0.000
WTN_Frac:	0.347	0.356
WTA_frac:	0.152	0.351
WTSulfur(%):	0.028	0.120
Density(g/cm3):	0.903	0.738
IBP(K):	0	0
5BP(K):	0	0
10BP(K):	0	0
20BP(K):	311	329
30BP(K):	336	342
40BP(K):	343	354
50BP(K):	359	369
60BP(K):	377	385
70BP(K):	405	410
80BP(K):	436	437
90BP(K):	485	491
95Bp(K):	0	0
EBP(K):	629	607

5.3.2 Homologous series determination

For these two samples, the core series were explicitly determined. Based on analytical literature and previous experience, five saturate and ten aromatic structures (including one thiophenic component) were specified for this model, and are shown in Figure 5.20.

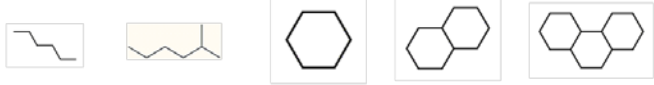
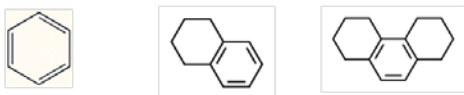

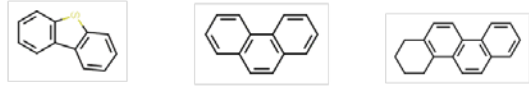
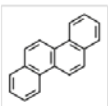
	Saturates
	Aro#1
	Aro#2
	Aro#3
	Aro#4

Figure 5.20 Selected homologous series for natural gas condensates

5.3.3 Structural attribute sampling protocol and PDF form

Similar to the petroleum middle distillates, quantitative optimization was tested by two different structural attribute sampling protocols and attribute PDF forms, as addressed in Chapter 2. In case 1, the feedstock was modeled as a set of homologous series with side chain information. The global fractions of these series were set as discrete values and a gamma distribution was imposed as the attribute PDF for carbon number and the number of side chain of each series. In case 2, the set of homologous series was decomposed to several elemental structural attributes, and the feedstock was sampled from multiple elemental structural attributes covering the core series' structure and side chain information. Figure 5.21 shows the specific sampling protocol for this case based on the general sampling protocol mentioned in Chapter 2. Both feedstocks were modeled by each of the two above cases.

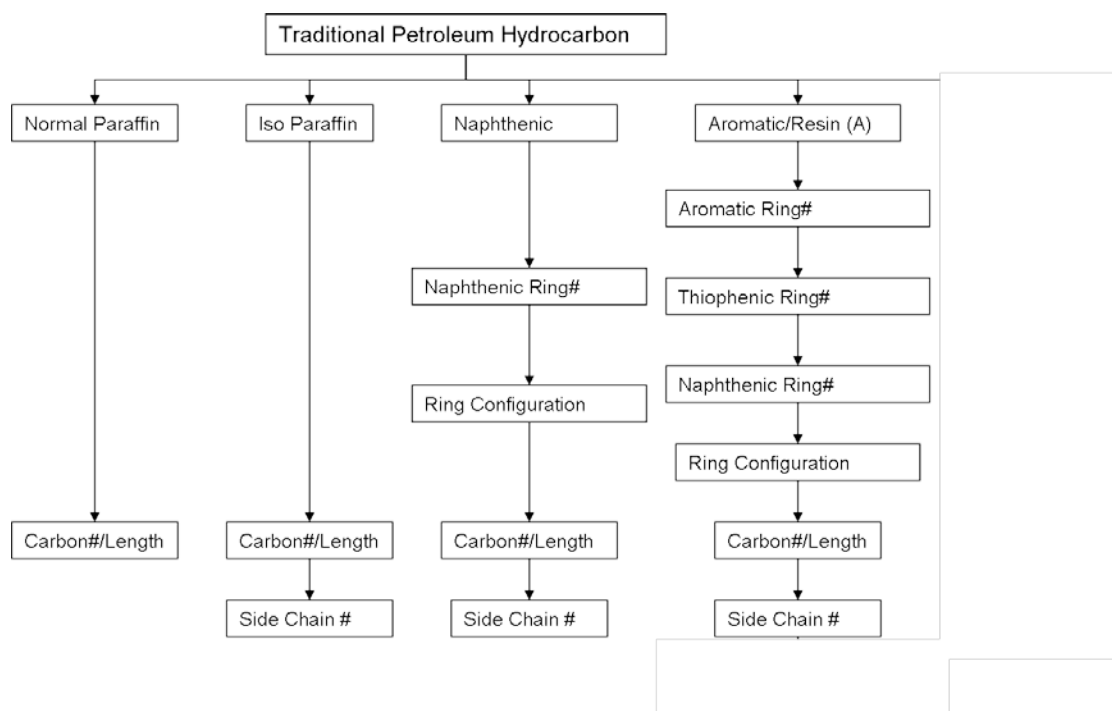


Figure 5.21 sampling protocol for natural gas condensates

5.3.4 Conditional Probability

In order to ensure physically realistic molecules, conditional probability distributions were used. Since the natural gas condensates were bounded by both an initial boiling point and a final boiling point, it was necessary to ensure that molecules that were too small or too large were not considered. By constraining the boiling point cut range in CME for each series, the range of carbon numbers was determined via structure property correlation before the optimization. Based on findings in the literature³, iso-paraffin components were limited one or two sidechains. The ring components were limited to only one side chain.

In addition to the overall carbon constraints, it was necessary to add conditional probability constraints to number and length of sidechains for the isoparaffin, naphthenic, and aromatic fractions. For instance, if a molecule contained only one sidechain carbon, then only one sidechain could possibly be attached to

the molecule core. Furthermore, since each sidechain contains at least one carbon, any one sidechain can be no longer than the total number of sidechain carbons minus the number of sidechains other than itself.

5.3.5 Optimization of the PDF parameters

The PDF parameters for selected natural gas condensates were optimized using a global simulated annealing technique. The chi square objective function of Eq. 5.2 was a combination of various molecular properties that could be easily compared to experimental data:

$$\begin{aligned} \chi^2 = & \left(\frac{Density_{\text{exp}} - Density_{\text{pred}}}{Weight_{\text{Density}}} \right)^2 + \left(\frac{SulWt_{\text{exp}} - SulWt_{\text{pred}}}{Weight_{\text{SulWt}}} \right)^2 \\ & + \left(\frac{1}{\#Comps} \right) \sum_{i=1}^{\#Comps} \left(\frac{PIONAWt_{\text{exp}} - PIONAWt_{\text{pred}}}{Weight_{\text{PIONA}}} \right)^2 \\ & + \left(\frac{1}{\#Fracs} \right) \sum_{i=1}^{\#Fracs} \left(\frac{SimDis_{\text{exp}} - SimDis_{\text{pred}}}{Weight_{\text{SimDis}}} \right)^2 + \dots \end{aligned} \quad (5.2)$$

Eq. 5.2 is similar to Eq. 5.1 for petroleum middle distillates, the only difference being the addition of sulfur weight content.

5.3.6 Validation results

The model predictions for the natural gas condensates are herein discussed for each of the two sampling protocols: case 1 and case 2 respectively.

5.3.6.1 Case 1 validation results

The optimized results of two natural gas condensates samples with case 1 are shown in Table 5.8. The distribution of the number of side chain was excluded from tuning and was given a fixed distribution based on Ranzi³ as shown in Table 5.3. The branch point locations were assumed to follow a uniform

distribution. For the most part, the model predictions were accurate to within an experimental standard deviation.

Table 5.8 Model predictions for the analytical properties of two natural gas condensates with case 1.

	GC663		GC659	
	Calc	Exp	Calc	Exp
WTP_frac:	0.150	0.143	0.177	0.175
WTI_frac:	0.150	0.149	0.326	0.325
WTO_frac:	0.000	0.000	0.000	0.000
WTN_Frac:	0.352	0.356	0.346	0.347
WTA_frac:	0.347	0.351	0.151	0.152
WTSulfur(%):	0.12	0.12	0.022	0.028
Density(g/cm3):	0.7672	0.7380	0.7244	0.7110
IBP(K):	284.2	-	243.1	-
5BP(K):	293.0	300.9	272.7	284.6
10BP(K):	308.0	310.5	288.0	295.0
20BP(K):	324.8	326.7	312.8	312.5
30BP(K):	341.9	341.6	331.3	328.8
40Bp(K):	357.2	356.6	351.5	345.2
50BP(K):	376.5	372.3	361.6	362.6
60BP(K):	388.2	389.6	381.5	381.8
70BP(K):	408.2	409.9	407.2	404.4
80BP(K):	436.9	435.8	431.7	433.4
90BP(K):	485.9	475.2	477.8	477.7
95Bp(K):	511.9	510.6	517.7	517.7
EBP(K):	623.3	-	633.6	-

Table 5.8 shows good model correlation for both of the sample feedstocks. Each of the analytical properties for these two samples were predicted to within one experimental standard deviation except for density. The reason for this discrepancy in density may be its less adequate structure-property estimation method. However, the accuracy of the PIONA weight percents and the simulated distillation curve shows more persuasive proof about the model's precision.

As an additional validation, the modeled data were compared to the GC-MS data. Figure 5.22 and Figure 5.23 show the group wise comparison between the modeled data and the analytical GC-MS data. From an overall point of view, the calculations match the analytical data moderately well. Groups that

show a greater composition fraction were matched better than those with a slighter composition fraction.

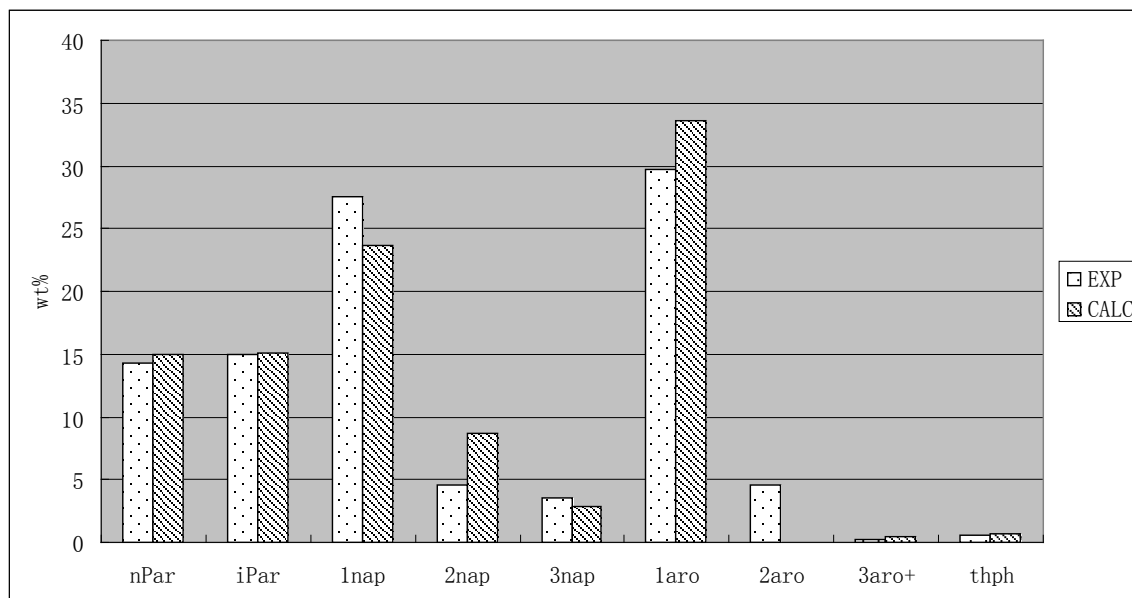


Figure 5.22 Calculated and experimental group composition data for GC 663 for case 1

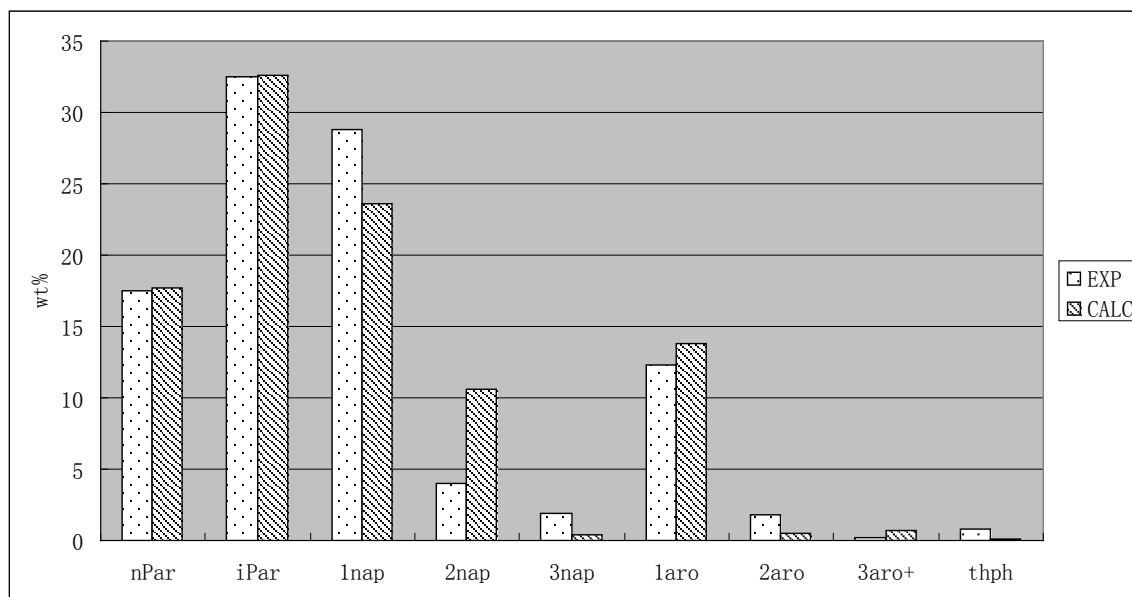


Figure 5.23 Calculated and experimental group composition data for GC659 with case 1

5.3.6.2 Case 2 validation results

The optimized results of two natural gas condensates samples with case 2 are shown in Table 5.9. The distribution of the number of side chain was excluded from tuning, and was given a fixed distribution from Ranzi³ as shown in Table 5.3. The branch point locations were assumed to follow a uniform distribution. For the most part, the model predictions were accurate to within an experimental standard deviation.

Table 5.9 Model predictions for the analytical properties of two natural gas condensates with case 2.

	GC663		GC659	
	Calc	Exp	Calc	Exp
WTP_frac:	0.144	0.143	0.174	0.175
WTI_frac:	0.149	0.149	0.325	0.325
WTO_frac:	0.000	0.000	0.000	0.000
WTN_Frac:	0.356	0.356	0.348	0.347
WTA_frac:	0.351	0.351	0.153	0.152
WTSulfur(%):	0.12	0.12	0.028	0.028
Density(g/cm3):	0.7805	0.7380	0.7258	0.7110
IBP(K):	279.4	-	259.1	271.3
5BP(K):	309.7	300.9	282.8	284.6
10BP(K):	325.9	310.5	306.0	295.0
20BP(K):	346.3	326.7	315.8	312.5
30BP(K):	356.7	341.6	325.9	328.8
40Bp(K):	361.7	356.6	345.2	345.2
50BP(K):	373.0	372.3	360.5	362.6
60BP(K):	387.3	389.6	381.7	381.8
70BP(K):	400.0	409.9	407.0	404.4
80BP(K):	435.6	435.8	430.9	433.4
90BP(K):	486.1	475.2	477.7	477.7
95Bp(K):	532.0	510.6	514.9	517.7
EBP(K):	609.5	-	628.0	632.3

Table 5.9 shows good model correlation for both of the sample feedstocks. Each of the analytical properties for these two samples was predicted to within one experimental standard deviation except for density. The reason for this discrepancy in density may be its less adequate structure-property estimation method. However, the accuracy of the PIONA weight percents and the simulated distillation curve shows more persuasive proof about the model's precision.

As additional validation, the modeled data were compared to the GC-MS data. Figure 5.24 and Figure 5.25 show the group wise comparison between the modeled data and the analytical GC-MS data. From an overall point of view, the calculations match the analytical data moderately well. Groups that show a greater composition fraction were matched better than those with a slighter composition fraction. Compared to the modeled data with case 1, the results of naphthenic and aromatic components show much better consistent with the experimental data.

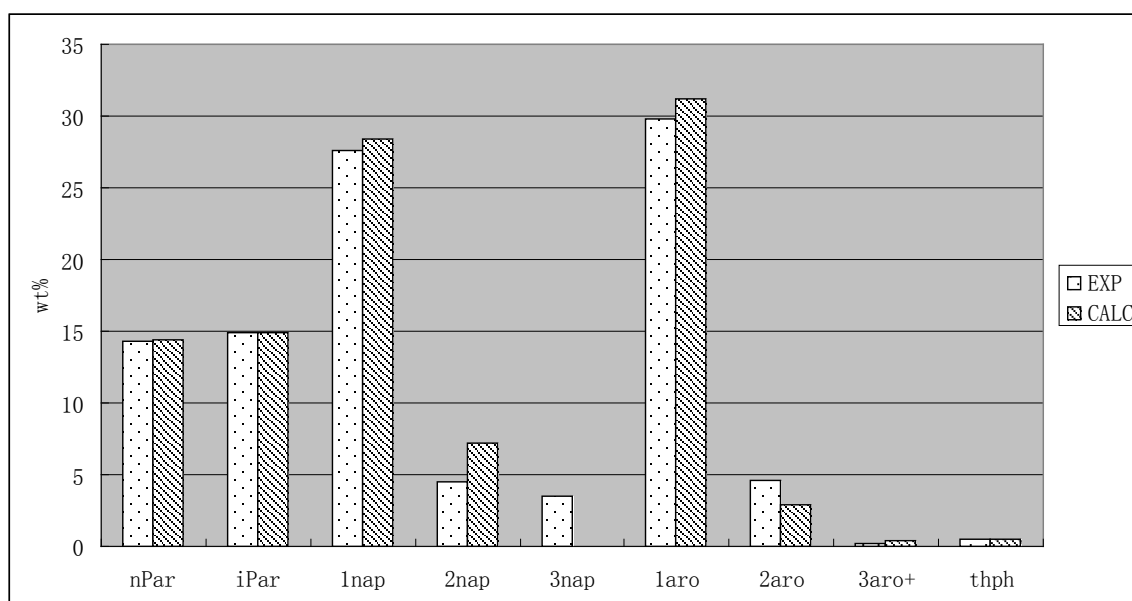


Figure 5.24 Calculated and experimental group GC-MS data for GC 663 with case 2

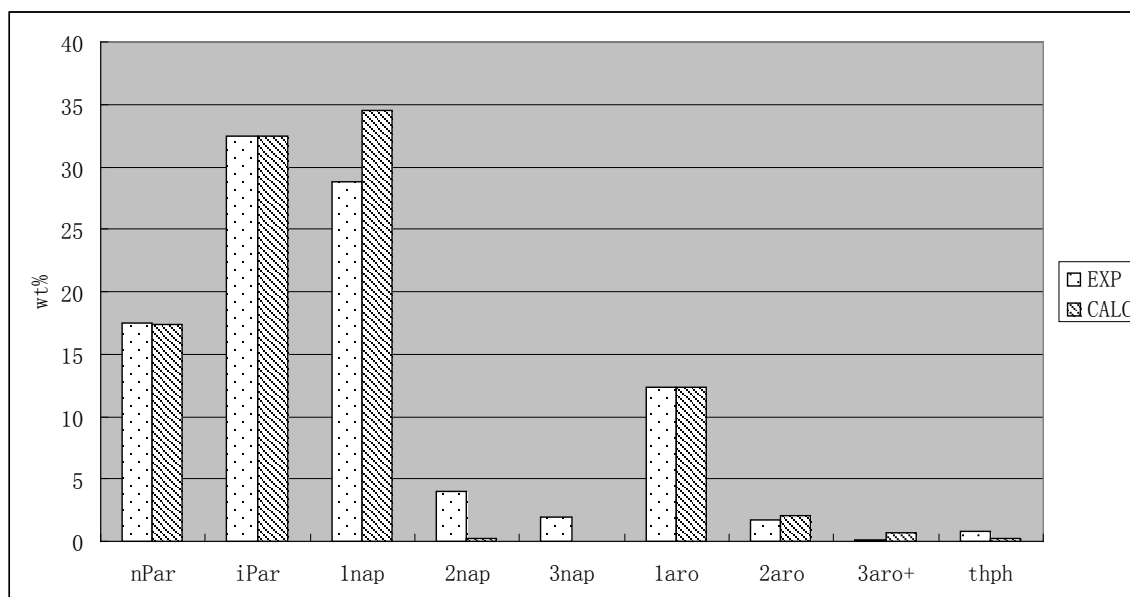


Figure 5.25 Calculated and experimental group GC-MS data for GC659 with case 2

5.3.6.3 Conclusion

The validation results for two samples using two different sampling protocols show an overall good fit for the natural gas condensates. This additional proof of concept gave a persuasive basis to apply CME on much heavier feeds. Over case 1 with fewer constraints and less analytical data, case 2 was determined to be more appropriate for heavy feedstocks. The next section will show the validation of CME to much heavier feedstocks solely using the sampling protocol of case 2.

5.4 Petroleum heavy gas oil

In addition to validating the modeled results of two above examples, this method was also applied to the much heavier conversion produced by petroleum heavy gas oil (HGO). The compositions of a HVGO conversion were much heavier than those of natural gas condensates but had a narrower boiling point cut range. For the HGO sample studied in the thesis, bulk properties were measured. In addition, Ghent University provided with a detailed composition data obtaining by GC-MS. Bulk properties were used

as the optimization criteria, the detailed composition data was used for model validation. The analytical information, the composition model, and the model results will be discussed in detail in the following sections.

5.4.1 Analytical characterization

This modeling method was applied to one sample of HGO. The analytical properties for this sample was provided by Ghent University¹. The properties used in the objective function is similar to these of the above two examples, and are summarized in Table 5.10.

Table 5.10 Bulk properties for the HGO sample

Name	
MW (g/mol):	300.4
WTP_frac:	0.156
WTI_frac:	0.245
WTO_frac:	0.000
WTN_Frac:	0.378
WTA_frac:	0.221
WTSulfur_%:	0.148
WTNitrogen_ppm:	0
Density (g/cm3):	0.8621
IBP (K):	510
5BP (K):	560
10BP (K):	579
20BP (K):	593
30BP (K):	606
40Bp (K):	616
50BP (K):	626
60BP (K):	637
70BP (K):	648
80BP (K):	661
90BP (K):	678
95Bp (K):	689
EBP (K):	729

5.4.2 Homologous series determination

For this sample, the core series were explicitly determined. Based on analytical literature and previous experience, six saturate and ten aromatic structures (including two thiophenic components) were specified for this model, and are shown in Figure 5.26.

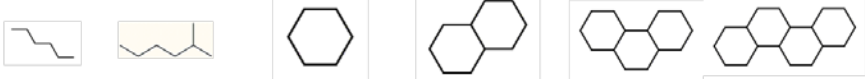
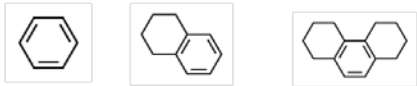
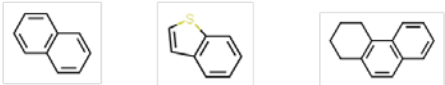
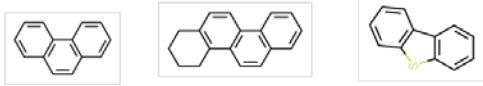
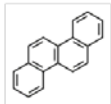
	Saturates
	Aro#1
	Aro#2
	Aro#3
	Aro#4

Figure 5.26 Selected homologous series for HGO

5.4.3 Structural attribute sampling protocol and PDF form

The quantitative optimization was tested followed case 2 as discussed in Chapter 2 in order to further validate the method of case 2 is more appropriate to heavy feedstocks. In case 2, the set of homologous series was decomposed to several elemental structural attributes, and the feedstock was sampled from multiple elemental structural attributes covering the core series' structure and side chain information. The specific sampling protocol for this HGO sample was the same as that for natural gas condensates as shown in Figure 5.21.

5.4.4 Conditional probability

In order to ensure physically realistic molecules, conditional probability distributions were used. Since the HGO sample was bounded by both an initial boiling point and a final boiling point, it was necessary to ensure that molecules that were too small or too large were not considered. By constraining the boiling point cut range in CME for each series, the range of carbon numbers was determined via structure property correlation before the optimization. Based on findings in the literature³, iso-paraffin components were limited one or two sidechains. The ring components were limited to only one side chain.

In addition to the overall carbon constraints, it was necessary to add conditional probability constraints on the number and length of sidechains for the isoparaffin, naphthenic, and aromatic fractions. For instance, if a molecule contained only one sidechain carbon, then only one sidechain could possibly be attached to the molecule core. Furthermore, since each sidechain contains at least one carbon, any one sidechain can be no longer than the total number of sidechain carbons minus the number of sidechains other than itself.

5.4.5 Optimization of the PDF parameters

The PDF parameters for selected HGO sample were optimized using a global simulated annealing technique. The chi square objective function of Eq. 5.3 was a combination of various molecular properties that could be easily compared to experimental data:

$$\begin{aligned}
\chi^2 = & \left(\frac{Density_{\text{exp}} - Density_{\text{pred}}}{Weight_{Density}} \right)^2 + \left(\frac{SulWt_{\text{exp}} - SulWt_{\text{pred}}}{Weight_{SulWt}} \right)^2 \\
& + \left(\frac{MW_{\text{exp}} - MW_{\text{pred}}}{Weight_{MW}} \right)^2 \\
& + \left(\frac{1}{\#Comps} \right) \sum_{i=1}^{\#Comps} \left(\frac{PIONAWt_{\text{exp}} - PIONAWt_{\text{pred}}}{Weight_{PIONA}} \right)^2 \\
& + \left(\frac{1}{\#Fracs} \right) \sum_{i=1}^{\#Fracs} \left(\frac{SimDis_{\text{exp}} - SimDis_{\text{pred}}}{Weight_{SimDis}} \right)^2 + \dots
\end{aligned} \tag{5.3}$$

Eq. 5.3 is similar to Eq. 5.2 for petroleum middle distillates, the only difference being the addition of average molecular weight.

5.4.6 Validation results

5.4.6.1 Case 2 validation results

The optimized results of the HGO sample with case 2 are shown in Table 5.11. The distribution of the number of side chain was excluded from tuning, and was given a fixed distribution based on Ranzi³ as shown in Table 5.3. The branch point locations were assumed to follow a uniform distribution. For the most part, the model predictions were accurate to within an experimental standard deviation.

Table 5.11 Model predictions for the analytical properties of the HGO sample with case 2.

Name	Calc	Exp
MW (g/mol):	300.4	300.4
WTP_frac:	0.156	0.156
WTI_frac:	0.246	0.245
WTO_frac:	0.000	0.000
WTN_Frac:	0.378	0.378
WTA_frac:	0.220	0.221
WTSulfur_%:	0.150	0.148
WTNitrogen_ppm:	0	0
Density (g/cm ³):	0.8465	0.8621
IBP (K):	540	510
5BP (K):	564	560
10BP (K):	577	579
20BP (K):	590	593
30BP (K):	601	606
40Bp (K):	612	616
50BP (K):	620	626
60BP (K):	630	637
70BP (K):	639	648
80BP (K):	650	661
90BP (K):	666	678
95Bp (K):	681	689
EBP (K):	726	729

Table 5.11 shows good model correlation for the HGO sample. Each of the analytical properties for this sample was predicted to within one experimental standard deviation except for density. The reason for this discrepancy in density may be its less adequate structure-property estimation method. However, the accuracy of PIONA weight percents and the simulated distillation curve shows more persuasive proof about the model's precision.

As an additional validation, the experimentally determined and predicted carbon number distributions corresponding to each PIONA fraction were compared via detailed composition data.

Figure 5.27 shows the experimentally determined and predicted carbon number distribution for the paraffin fraction of the HGO sample with case2. The predicted curve matches the experimentally determined curve decently well although the curve of analytical data is not ideally smooth. The peak area of the predicted values was a bit of a right-shift compared to the analytical data.

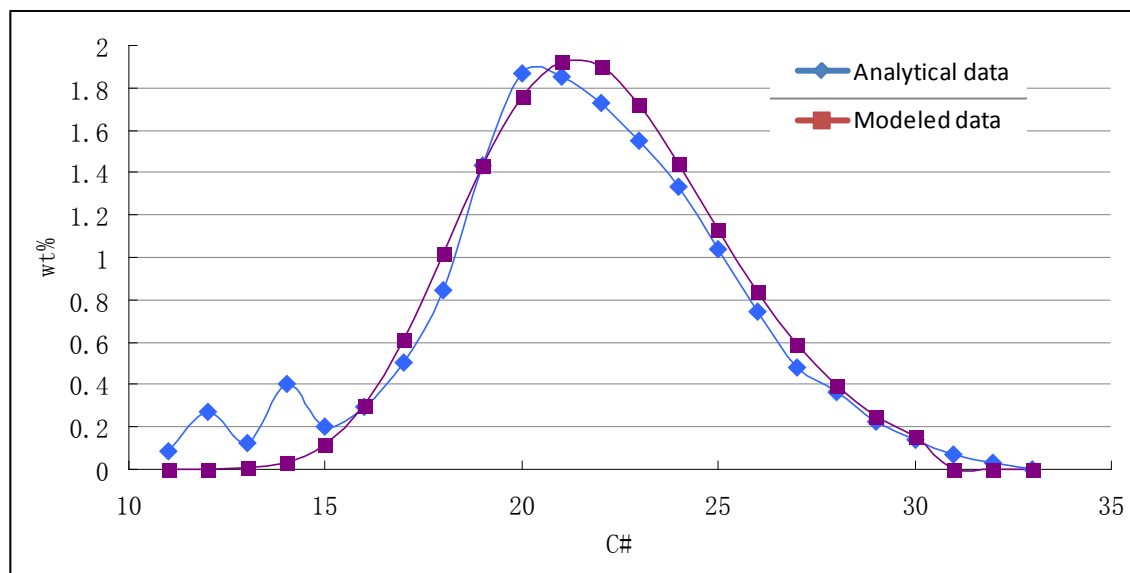


Figure 5.27 Predicted and experimental carbon number distribution for the paraffin fraction of HGO sample with case 2.

Figure 5.28 shows the predicted and experimentally determined carbon number distributions for the isoparaffin fraction of HGO sample with case 2. The predicted curve matches the experimentally determined curve moderately well. The predicted values for carbon numbers 18-25 are overestimated, while the values for carbon number from 25 onward end are underestimated. Since the deviation also shows as an opposite symmetry trend, the influences are cancelled out and only contribute a small effect to the overall weight fraction in the objective function.

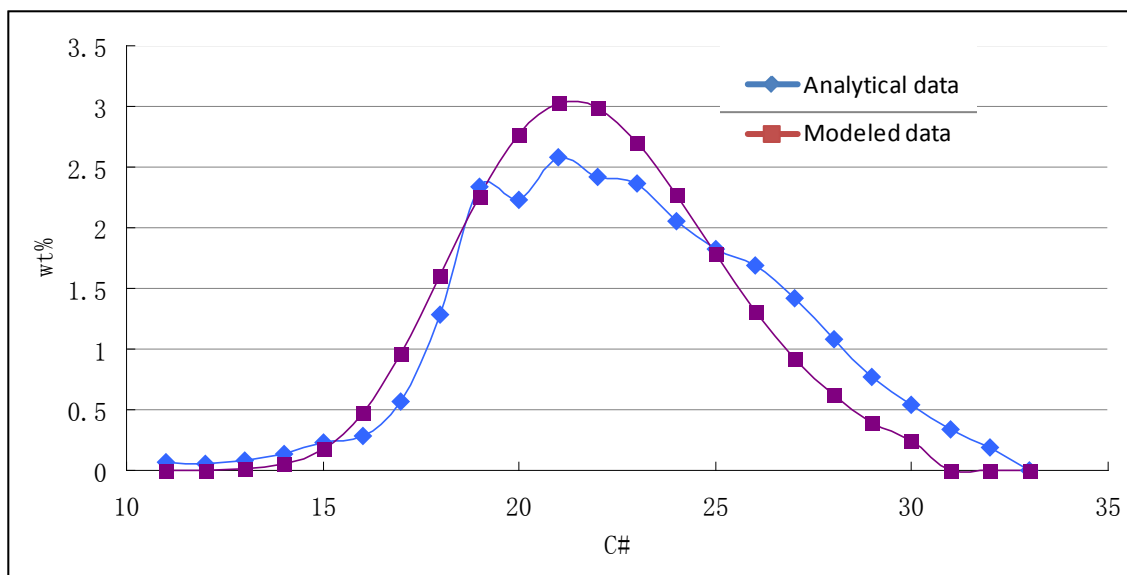


Figure 5.28 Predicted and experimental carbon number distribution for the isoparaffin fraction of HGO sample with case 2.

Figure 5.29 shows the predicted and experimentally determined carbon number distribution for the naphthenic fraction of HGO sample with case 2. The predicted curve matches the experimentally determined curve moderately well. The predicted values for carbon numbers 20-25 are overestimated, while the values for carbon number from 25 onward end are underestimated. Since the deviation also shows as an opposite symmetry trend, the influences are cancelled out and only contribute a small effect to the overall weight fraction in the objective function.

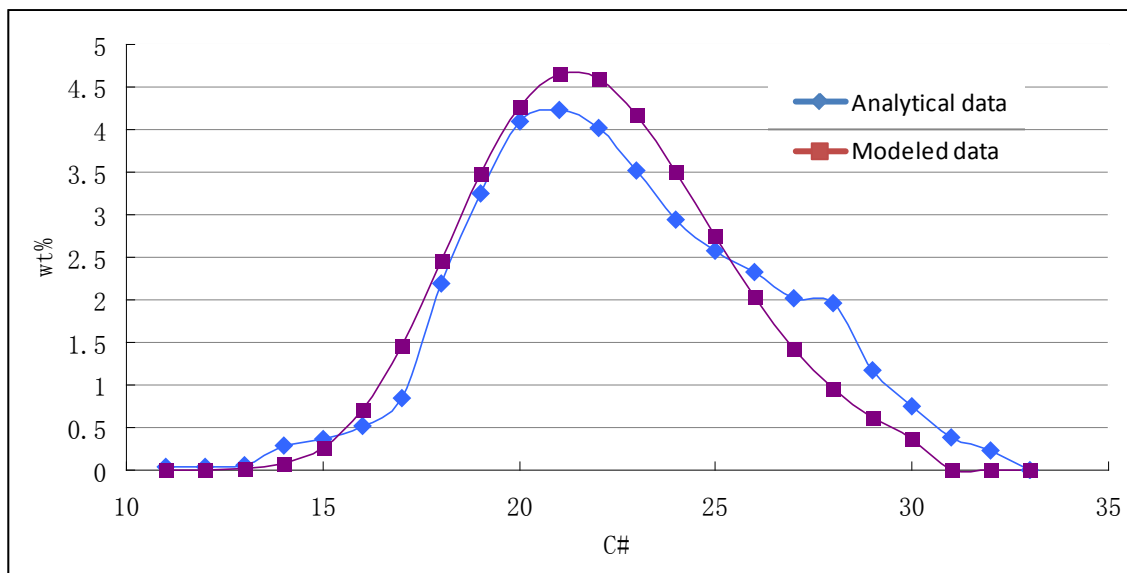


Figure 5.29 Predicted and experimental carbon number distribution for the naphthenic fraction of HGO sample with case 2.

Figure 5.30 shows the predicted and experimentally determined carbon number distribution for the mono, di-aromatic fraction of the HGO sample with case 2. The predicted curve matches the experimentally determined curve moderately well. The curve of the modeled data is higher than the analytical data and shows a bit right-shift close to the peak.

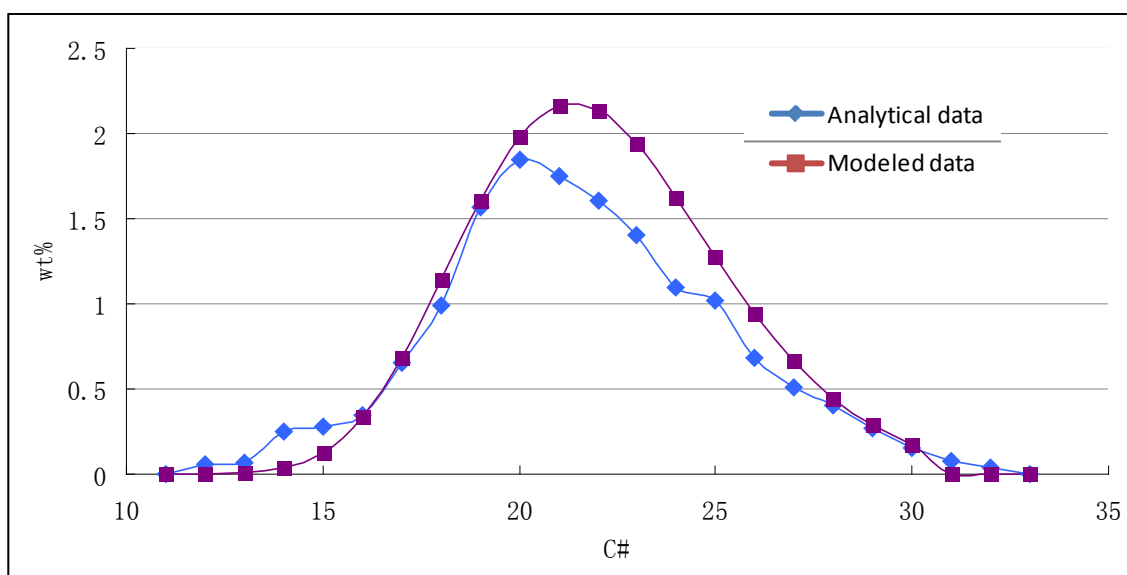


Figure 5.30 Predicted and experimentally determined carbon number distributions of the mono,di-aromatic fraction of the HGO sample with case 2.

Figure 5.31 shows the predicted and experimentally determined carbon number distribution for the tri+-aromatic fraction of the HGO sample with case 2. The predicted curve matches the experimentally determined curve decently well although the curve of analytical data is not ideally smooth.

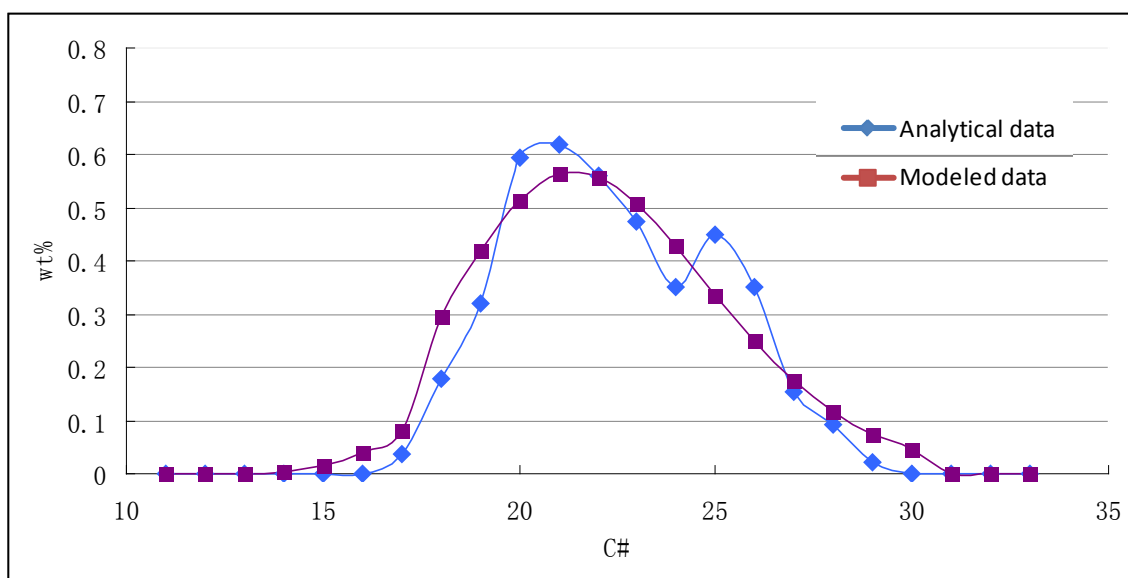


Figure 5.31 Predicted and experimentally determined carbon number distributions of the tri+-aromatic fraction of the HGO sample with case 2.

Figure 5.32 shows the predicted and experimentally determined carbon number distribution for the benzothiophene fraction of the HGO sample with case 2. The predicted curve matches the experimentally determined curve decently well although the curve of analytical data is not ideally smooth.

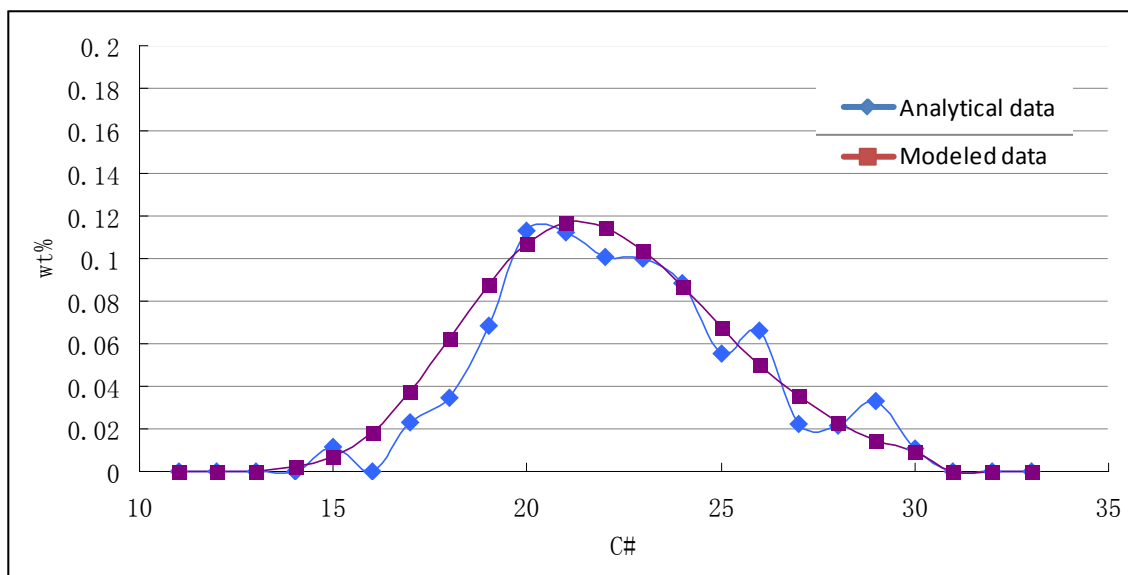


Figure 5.32 Predicted and experimentally determined carbon number distributions of the benzothiophene fraction of the HGO sample with case 2.

Figure 5.33 shows the predicted and experimentally determined carbon number distribution for the di-benzothiophene fraction of the HGO sample with case 2. The predicted curve does not match the experimentally determined curve well. It is hard to catch the analytical data by the continuous PDF because the experimental curve shows a scatter manner. In addition, the quantitative amount of di-benzothiophene is quite small, therefore, it only contribute a trivial effect to overall model calibration effect.

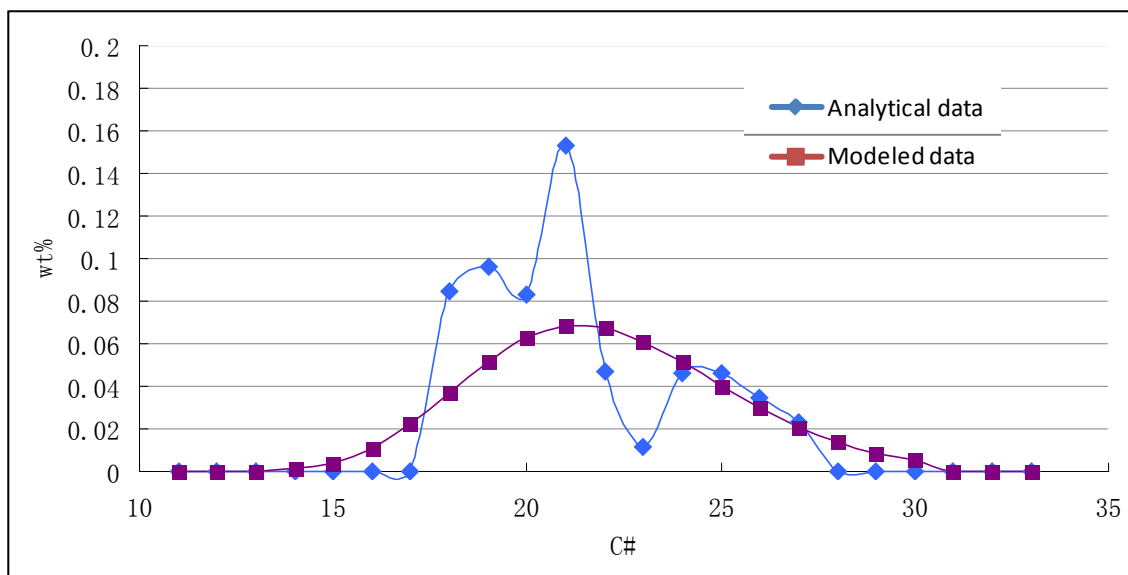


Figure 5.33 Predicted and experimentally determined carbon number distributions of the di-benzothiophene fraction of the HGO sample with case 2.

5.4.6.2 Conclusion

The validation results for this sample using the sampling protocol of case2 show an overall good fit for the HGO conversion. This additional proof of concept gave a persuasive basis to apply CME on much heavier feeds with case 2. The next section will show the application of CME to heavier feedstocks using extrapolated predictions based on the approach in Chapter 2.

5.5 Vacuum gas oil

This section illustrates a CME generated composition model of vacuum gas oil (VGO). VGO is much heavier than three above examples, and as such detailed compositional data was unavailable for the purpose of validation. This section details a predictive study of VGO's composition based solely on the measured bulk properties used for model optimization.

5.5.1 Analytical characterization

This method was applied to one sample of VGO. The analytical properties for this sample were provided by Ghent University¹. The properties used in the objective function is similar to these of the above two examples, and are summarized in Table 5.12.

Table 5.12 Bulk properties for VGO

H/C	1.995
WTP_frac:	0.061
WTI_frac:	0.241
WTO_frac:	0.000
WTN_Frac:	0.478
WTA_frac:	0.220
Density(g/cm3):	0.903
IBP(K):	543
5BP(K):	573
10BP(K):	0
20BP(K):	0
30BP(K):	0
40Bp(K):	0
50BP(K):	686
60BP(K):	0
70BP(K):	0
80BP(K):	0
90BP(K):	0
95Bp(K):	774
EBP(K):	826

5.5.2 Homologous series determination

For this VGO sample, the core series were explicitly determined. Based on analytical literature and previous experience, five saturate and ten aromatic structures up (to four rings) were specified for this model, and are shown in Figure 5.34.


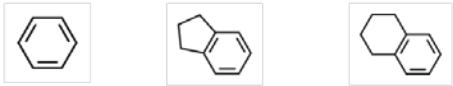
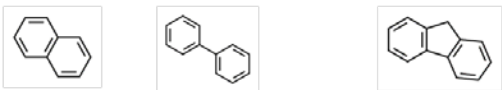

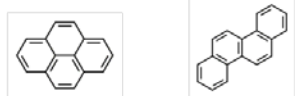
	Saturates
	Aro#1
	Aro#2
	Aro#3
	Aro#4

Figure 5.34 Selected homologous series for VGO

5.5.3 Structural attribute sampling protocol and PDF form

Similar to the previous two examples, the quantitative optimization was tested by two different structural attribute sampling protocols and attribute PDF forms, as addressed in Chapter 2. In case 1, the feedstock was modeled as a set of homologous series with side chain information. The global fractions of these series were set as discrete values and a gamma distribution was imposed as the attribute PDF for carbon number and the number of side chain of each series. In case 2, the set of homologous series was decomposed to several elemental structural attributes, and the feedstock was sampled from multiple elemental structural attributes covering core the series' structure and side chain information. The specific sampling protocol for this VGO sample was the same as that for petroleum middle distillates as shown in Figure 5.3.

5.5.4 Conditional Probability

In order to ensure physically realistic molecules, conditional probability distributions were used. Since the VGO sample was bounded by both an initial boiling point and a final boiling point, it was necessary to ensure that molecules that were too small or too large were not considered. By constraining the

boiling point cut range in CME for each series, the range of carbon numbers was determined via structure property correlation before the optimization. Based on findings in the literature³, iso-paraffin components were limited to one or two sidechains. Single ring components were limited to one, two or three sidechains. Multiple ring components were limited to only one or two side chain.

In addition to the overall carbon constraints, it was necessary to add conditional probability constraints on the number and length of sidechains for the isoparaffin, naphthenic, and aromatic fractions. For instance, if a molecule contained only one sidechain carbon, then only one sidechain could possibly be attached to the molecule core. Furthermore, since each sidechain contains at least one carbon, any one sidechain can be no longer than the total number of sidechain carbons minus the number of sidechains other than itself.

5.5.5 Optimization of the PDF Parameters

The PDF parameters for selected VGO sample were optimized using a global simulated annealing technique. The chi square objective function of Eq. 5.4 was a combination of various molecular properties that could be easily compared to experimental data:

$$\begin{aligned}
 \chi^2 = & \left(\frac{Density_{exp} - Density_{pred}}{Weight_{Density}} \right)^2 + \left(\frac{SulWt_{exp} - SulWt_{pred}}{Weight_{SulWt}} \right)^2 \\
 & + \left(\frac{H/C_{exp} - H/C_{pred}}{Weight_{H/C}} \right)^2 \\
 & + \left(\frac{1}{\#Comps} \right) \sum_{i=1}^{\#Comps} \left(\frac{PIONAWt_{exp} - PIONAWt_{pred}}{Weight_{PIONA}} \right)^2 \\
 & + \left(\frac{1}{\#Fracs} \right) \sum_{i=1}^{\#Fracs} \left(\frac{SimDis_{exp} - SimDis_{pred}}{Weight_{SimDis}} \right)^2 + \dots
 \end{aligned} \tag{5.4}$$

Eq. 5.4 is similar to Eq. 5.1 for petroleum middle distillates, the only difference being the addition of H/C ratio.

5.5.6 Optimization results

The VGO sample was optimized by CME using both case1 and case2 method. The optimized results of both methods are shown in Table 5.13. The distribution of the number of side chain distribution was excluded from tuning, and was given a fixed distribution based on Ranzi³ as shown in Table 5.3. The branch point locations were assumed to follow a uniform distribution. For the most part, the model predictions of both cases were good to within an experimental standard deviation. The method of case2 had a better fit than the method of case1 for some items such as H to C ratio. As was the case for the other two examples, the density did not fit well. This discrepancy might be improved by implementing new methods for density calculation. However, the accuracy of the fit of simulated distillation and PONA are quite good for both cases, thereby showing show persuasive proof about the model's precision.

This example shows that CME can handle a heavier complex feedstock, such a VGO. Therefore CME provides high value in overcoming the formidable difficulty involved in obtaining a full detailed compositional analysis that can be used by a detailed kinetic model. The next section covers the modeling of extremely heavy feedstocks, specifically petroleum resids.

Table 5.13 Model optimization results for the VGO sample

	Exp	CalcbyCase1	CalcbyCase2
H/C	1.995	1.900	1.963
WTP_frac:	0.061	0.063	0.054
WTI_frac:	0.241	0.237	0.250
WTO_frac:	0.000	0.000	0.000
WTN_Frac:	0.478	0.478	0.478
WTA_frac:	0.220	0.222	0.219
Density(g/cm3):	0.903	0.747	0.842
IBP(K):	543	546.000	541
5BP(K):	573	576.000	567
10BP(K):	-	598.000	579
20BP(K):	-	626.000	602
30BP(K):	-	646.000	644
40Bp(K):	-	664.000	669
50BP(K):	686	682.000	685
60BP(K):	-	697.000	700
70BP(K):	-	714.000	717
80BP(K):	-	732.000	735
90BP(K):	-	758.000	758
95Bp(K):	774	777.000	774
EBP(K):	826	808.000	817

5.6 Petroleum resids

This section illustrates how to model the compositions of an extreme feedstock: resid. Resid is far too complex to obtain detailed compositional data. Therefore, this section details a predictive study of resid's composition based solely on the measured bulk properties used for model optimization. The purpose of this section is to illustrate the logic necessary to model resids, therefore a selection of light resids was used to simplify the model process.

5.6.1 Analytical characterization

The available analytical information of a resid is different that that of the other feeds discussed above. Firstly, the compositional analysis for resid is SARA instead of PIONA, which indicates that the

compositions found in resid are mainly multiple aromatic rings and their aggregated ring structures. Moreover, simulated distillation can not provide a complete sample for resid's fractions. The average molecular weight is used as a criterion to evaluate how heavy the feedstock is. Finally, heteroatoms take a significant role in resids. Proton NMR and other advanced techniques can help understand resids' structures insightfully.

For the model in this chapter, a variety of light resids was selected to simplify the illustration. Petroleum resids located in China are much different than those found in other places such as the middle-east. These Chinese resids often have lower sulfur (and other heteroatom) contents than others, which will help simplify the structures' determination. Therefore, two Chinese resid samples with lower sulfur content were selected: DaQing and ShengLi. Additionally, for the purpose of determining CME's current capability, two higher sulfur-containing resids were selected as: GuDao China and Arabic Light. The properties used as the objective function for resids are summarized in Table 5.14. Since NMR data was unavailable for the Chinese samples, it was excluded from the measurements.

Table 5.14 Bulk properties for resids

Name	ShengLi	DaQing	GuDao	Arab Light
MW:	784	1030	913	842
H/C:	1.63	1.74	1.56	1.42
WTP_frac:	0.012	0.024	0.01	0.016
WTN_Frac:	0.185	0.385	0.147	0.147
WT_Arom.+Resin_Frac:	0.801	0.591	0.815	0.779
WT_Asphaltene_Frac:	0.002	-	0.028	0.06
WTSulfur_%:	1.35	0.27	2.43	4.00
WT_HGOCut:	0.10	0.10	0.10	0.10
WT_VGOCut:	0.13	0.11	0.90	0.09
WT_ResidCut:	0.86	0.88	0.90	0.91

5.6.2 Homologous series determination

Campbell et al^{4,5} used MolGen and Monte Carlo simulation to study several resid samples. Based on this and other experience of the Klein group, the core series was explicitly determined for the resid




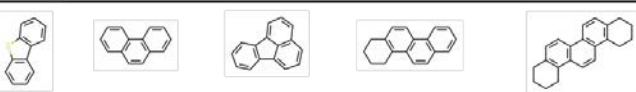
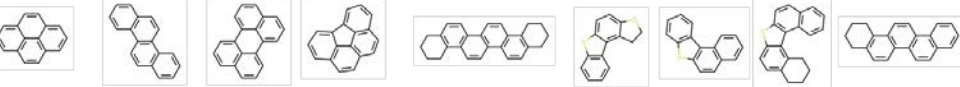
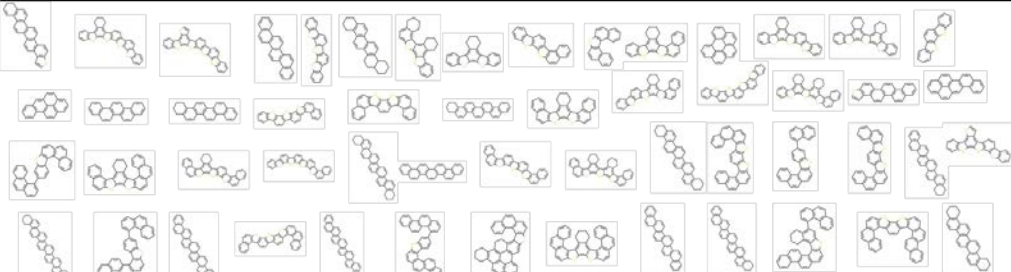
	Saturates
	Aro#1
	Aro#2
	Aro#3
	Aro#4
	Aro#5+

Figure 5.35 Selected homologous series for resins

5.6.3 Structural attribute sampling protocol and PDF form

For the resid model, the method of case1 discussed in Chapter 2 is not applicable, due to the high computational burden necessary for this method. Therefore, only the case 2 methodology in Chapter 2 was applied to these resids. Each feedstock was sampled from multiple elemental structural attributes covering the core series' structure and side chain information. The specific sampling protocol for resids is shown in Figure 5.36. In addition, in order to control the complexity of the computation, the side chains were lumped into three "families": paraffin, naphthene, and aromatics.

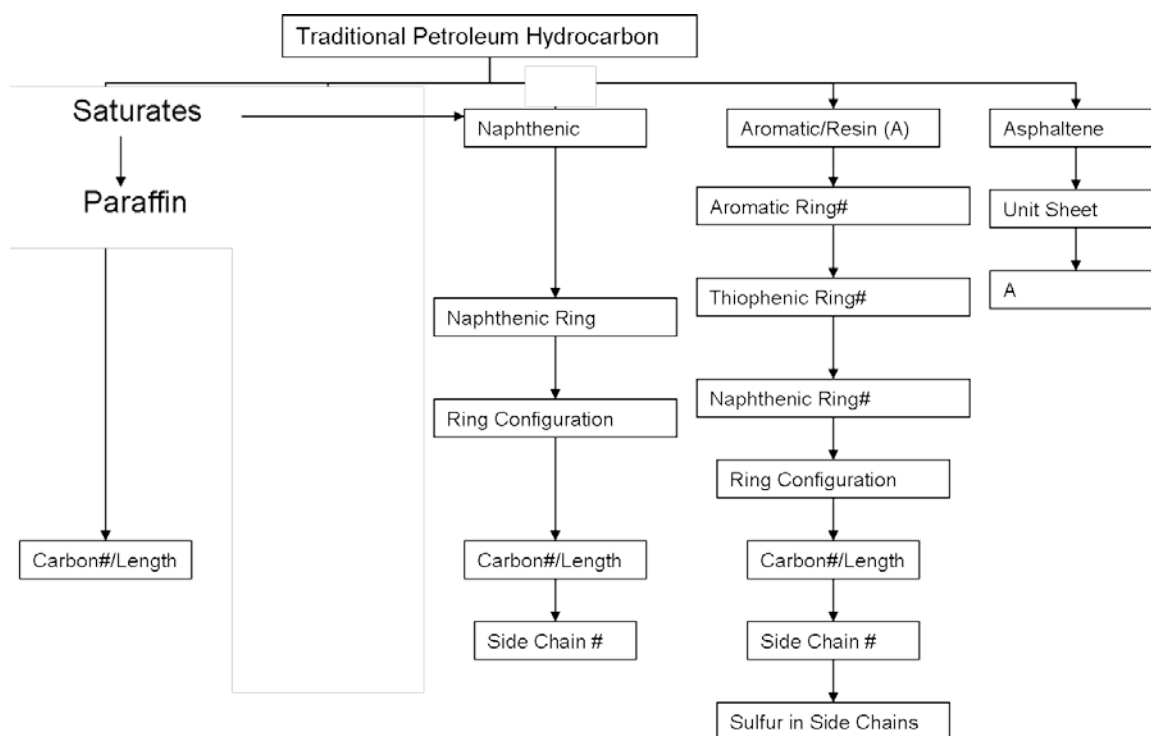


Figure 5.36 sampling protocol for resids

5.6.4 Conditional probability

In order to ensure physically realistic molecules, conditional probability distributions were used. Since the resid samples were only bound by an initial boiling point, the minimum carbon number for each series was thusly limited. The maximum carbon number for the side chains was determined by the past experience. All series in the resid were assumed to have one side chain.

5.6.5 Optimization of the PDF parameters

The PDF parameters for selected Resids' samples were optimized using a global simulated annealing technique. The chi square objective function of Eq. 5.5 was a combination of various molecular properties that could be easily compared to experimental data:

$$\begin{aligned}
\chi^2 = & \left(\frac{MW_{\text{exp}} - MW_{\text{pred}}}{W_{MW_{\text{exp}}}} \right)^2 + \left(\frac{HtoC_{\text{exp}} - HtoC_{\text{pred}}}{W_{HtoC_{\text{exp}}}} \right)^2 + \\
& + \left(\frac{Sulf_{\text{exp}} - Sulf_{\text{pred}}}{W_{Sulf_{\text{exp}}}} \right)^2 + \frac{1}{\#Frac s} \sum_{i=1}^{\#Frac s} \left(\frac{SimDis_{\text{exp}} - SimDis_{\text{pred}}}{W_{SimDis_{\text{exp}}}} \right)^2 \\
& + \frac{1}{\#Comps} \sum_{i=1}^{\#Comps} \left(\frac{SARA_{\text{exp}} - SARA_{\text{pred}}}{W_{SARA}} \right)^2
\end{aligned} \tag{5.5}$$

5.6.6 Optimization results

The four resids' samples were optimized by CME using the method of case2. The optimized results are shown in Table 5.15. The results show that the ShengLi and DaQing resids were well modeled by CME. All items including SARA, H/C ratio, molecular weight, sulfur content, and incomplete distillation were decently optimized. Therefore the resultant compositions could be used for further kinetic model study.

The GuDao and Arab Light resids did not get good results. Although SARA of GuDao looks good, the H/C ratio and sulfur content have a large bias. The incomplete distillation agreement was even worse. The Arab light resid did not fit the data at all. The reason CME could not match these two feeds is rooted in the identity determination discussed in the previous section. The current version of CME does not create sufficient sulfur containing side chain structures. The identities selected in the previous section work only for low sulfur resids such as ShengLi and DaQing. Therefore, CME could not provide good optimization for high sulfur resids that contained more key structures than this model provides. This issue could be solved by adding such necessary structure information into future version of CME.

Although the two high sulfur content resids were not well modeled, the good correlations for the two low sulfur resids show that, conceptually, CME can model extremely heavy feedstocks by the approach addressed in Chapter 2. If more heteroatom structures were incorporated, CME could probably handle

heavy resids with high sulfur contents. Therefore, modeling asphaltenes with CME also looks promising.

Table 5.15 Model optimization results for the resids' samples

Name	ShengLi		DaQing		GuDao		Arab Light	
	Exp	Calc	Exp	Calc	Exp	Calc	Exp	Calc
MW:	784	784	1030	1031	913	910	842	810
H/C:	1.63	1.63	1.74	1.75	1.56	1.44	1.42	1.35
WTP_frac:	0.012	0.013	0.024	0.025	0.01	0.010	0.016	0.008
WTN_Frac:	0.185	0.185	0.385	0.384	0.147	0.146	0.147	0.107
WT_Arom.+Resin_Frac:	0.801	0.801	0.591	0.591	0.815	0.808	0.779	0.870
WT_Asphaltene_Frac:	0.002	0.001	-	0	0.028	0.036	0.06	0.015
WTSulfur_%:	1.35	1.35	0.27	0.27	2.43	2.311	4.00	2.674
WT_HGOCut:	0.10	0.02	0.10	0.00	0.10	0.000	0.10	0.006
WT_VGOCut:	0.13	0.13	0.11	0.12	0.90	0.009	0.09	0.002
WT_ResidCut:	0.86	0.85	0.88	0.88	0.90	0.990	0.91	0.992

5.7 Summary and discussions

This chapter mainly shows specific and varied examples to support the composition modeling approach developed in Chapter 2. Three feedstocks were used to validate the approach: petroleum middle distillates, natural gas condensates and petroleum HGO. Both of these feeds were modeled with bulk properties and validated using their analytical and modeled compositional data. Furthermore, two other feedstocks (VGO and resids) were studied to ensure the methodology could be used for heavier feedstocks.

The petroleum middle distillate samples included the detailed composition data, so the two optimization methods (case1 and case 2) discussed in Chapter 2 could be fully evaluated. The overall good results of modeling petroleum middle distillates proved that modeling a complex feedstock statistically with fixed identity determination and quantitative structural attribute PDFs sampling is a valid methodology. In addition, the results showed that the method of case 2 (where the feedstock was optimized by a set of continuous elemental structural attribute PDFs) required less analytical information than the method of

case 1 (where the feedstock is optimized by a set of discrete values for the global core series fractions and attribute PDFs for the side chain information). Because of the maintenance of accuracy with less analytic information, case 2 was shown to be more useful for complex feedstocks.

The further validation of the method using natural condensates and HGO enhanced the above conclusion and proved that the method could be used for heavier feedstocks.

The application of the method to VGO showed that CME could handle even heavier complex feedstocks.

Finally, the method was applied to petroleum resids. Although a simplified identity model was employed on resids, CME showed a good capability to model lighter resids with lower sulfur content . More complex resids, such as those with high-sulfur content, could be modeled in the future by adding more heteroatoms structures to CME.

In summary, the methods behind CME were verified by analytical data, and it was shown that CME could effectively model complex feedstocks with limited analytical information.

Reference

- ¹ Laboratory for Chemical Technology, Department of Chemical Engineering, Ghent University, Belgium
- ² Van Geem, Pyl, et al. J. Chrom. A. 2010
- ³ E. Ranzi , et al. Progress in Energy and Combustion Science 27 (2001) 99–139
- ⁴ Campbell, D.M., Stochastic Modeling of Structure and Reaction in Hydrocarbon Conversion, Doctoral Dissertation, University of Delaware (1998)
- ⁵ Trauth, D.M., Structure of Complex Mixtures through Characterization, Reaction, and Modeling, Ph. D. Dissertation, University of Delaware, 1993.

Chapter 6 Modeling lignin pyrolysis with ARM

6.1 Introduction

The worldwide energy crisis has accelerated renewed interest in the search for alternate feedstocks for the supply of energy such as biomass. The utilization of such feeds is complicated by not only the economics of upgrading and conversion but also the environmental footprint, including the impact on CO₂ issues. Therefore, the technique to address these feeds needs to be considered.

While each potential resource has its own set of feed-specific technical issues, there are some common difficulties. A significant character of these alternate feeds is a complex mixture of complex molecules that are often found as covalent or physically aggregated macromolecules. According to the report¹ from US DOE Office, it is critical to identify the structures and reaction pathways of these heavy feeds and to develop robust computational tools to model their reaction trajectories. This would contribute not only to the processes for utilization of these feeds but also to the design of catalysts for providing efficient and optimal conversion.

Therefore, the development of a modeling methodology for application to Biomass (such as Lignin, Coal, Shale, and Resid) is an urgent and significant issue for contemporary energy research. Biomass and even Gas Oils differ, only in the quantitative details of the feeds. That is, each of these feeds can be characterized, relative to petroleum naphtha, as carbon-rich, hydrogen-deficient resources laden with heteroatoms and often found in macromolecular form. As an example of biomass, lignin pyrolysis is studied in this chapter.

In this chapter, the significance of lignin and lignin pyrolysis is first considered. Then, a lignin pyrolysis model is developed via the set of modeling tools addressed from Chapter 2 to Chapter 4. The model

development will consider the structure of lignin, the analysis of reaction paths and kinetics, the development of pyrolysis network, and the prediction of pyrolysis products, in turn.

6.2 The significance of lignin and lignin pyrolysis

Interests in lignin are both as a possible feed in the biomass contribution to new energy supplies and also as a relatively well understood model of low-rank coals. There was an intense application in the use of coal in liquefaction and gasification processes around 1970's. Lignin, the microbially resistant component of the biomass coal precursor, was examined to explore on the reaction paths and kinetics involved in coal utilization. Because the investment of the utilization of lignin as an energy resource is less expensive, lignin is widely applied in marginal energy recovery rather than burning as the pulping waste product. These interests are also clear in the current energy situation because it can provide useful information for further biomass study. Pyrolysis is a fundamental component of the potential conversion processes, such as gasification, pyrolysis and liquefaction to bio-oils. Thus the study of lignin pyrolysis modeling seemed promising.

More generally, lignin can be described as a macromolecular feed with aromatic rings, ring substituents, and heteroatoms. This qualitative information can also describe coal, resins and shales, and thus developments in lignin pyrolysis modeling should support the development of models for these other feeds as well.

6.3 Modeling approach

In order to construct a detailed kinetic model of lignin pyrolysis, the four main steps of the molecule-based kinetic modeling for complex feeds (listed in Figure 6.1) must be addressed. The approaches of these four steps are extensively discussed in previous chapters. The first is the structure-composition model for the feed. In this thesis, CME was used as a generic tool to transform a

set of measurements into a structural representation and the quantitative mol fractions for lignin. As developed below, lignins are optimally complex in that their structures are actually very well known.

1. Structural and Composition Models

- **Measurements (GS-MS, NMR, etc.)**
- **Measurements to Molecules**

2. Reaction Modeling

- **Computer generated models in seconds**

3. Reactivity Correlations

- **Fundamentals based Chem Eng LFER kinetics correlations**
- **Order 10 [O(10)] LFER's for every process chemistry**

4. Property Estimation

- **Provides commercially relevant product yields**
- **Molecules to commercial lumps**
- **End-use vs. internal-use properties**

Figure 6.1 Molecular-level Modeling steps

Once the feed stock has been modeled, the second component creates the reaction model by transforming reactants into products using experimentally discerned reaction pathways. To address this lignin pyrolysis chemistry, the ARM approach developed in Chapter3 will be used. Unlike a traditional deterministic model that has one material balance for each of the N components along with energy and momentum balances, the ARM can model a lignin mixture containing N-components with fewer than N material balance equations.

The kinetics and reaction pathways of the lignin model were derived from experimental literatures². For this simpler system, independent individual parameters can be used for reactivity estimation. For a general lignin pyrolysis system that was a network of thousands of reactants and rate laws, the reaction family and LFER concepts of Chapter 3 would have to be applied, thereby reducing the parameters burden from thousands to of order 10-30.

The final component is the property estimation function, which was committed by the SPCM detailed in Chapter 4.

6.4 Modeling Lignin structure and composition

Native lignin is a phenolic co-polymer formed by a set of enzyme-initiated radicals on the phenolic² and corresponding reactive positions of the three monomers as shown in Figure 6.2. As a result, such a co-polymer thus contains single-ring aromatic cores bonded covalently by one of a handful of inter-unit linkages. The classic Freudenberg model³ (Figure 6.3) illustrates the types of these links and further provides a convenient quantitative distribution of the types. Isolated lignins will deviate from this prototype model by various configurations on differing initial distributions of the three monomers (e.g. different plant types) and also by structural changes induced by the method of isolation. These changes are well known in the biochemistry and pulping literature.

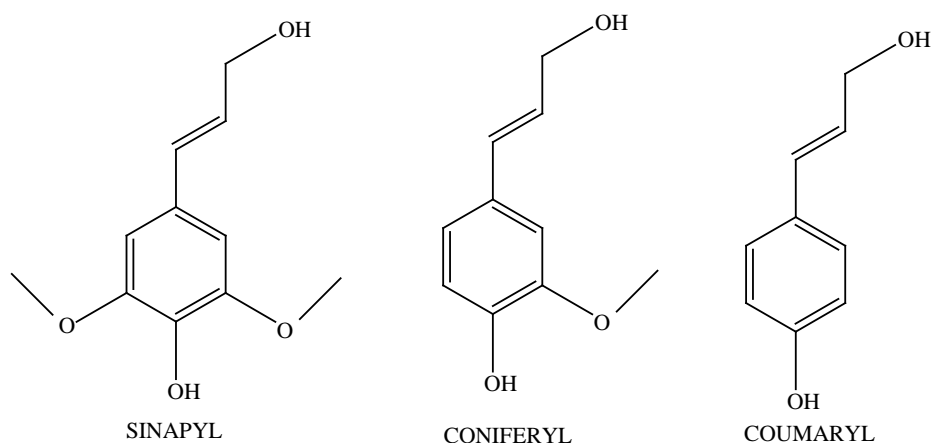


Figure 6.2 Lignin monomer alcohols

The Freudenberg model shown in Figure 6.3 was used as the example to study the composition model for the lignin pyrolysis. The Freudenberg model allowed the CME tool to model the structure and composition of biomass. The monomer alcohols, the crosslinking mechanism and the lignin pyrolysis mechanism gave a statistical view of the lignin structure as a set of single-ring aromatics each with two

attributes. The first attribute is the type of propanoid side chain (PC) attached to each aromatic ring and the second attribute is the type of the phenolic or methoxyphenol (MP) substituent on each aromatic ring. Since pyrolysis keeps the rings unchanged, the ARM model was based on the conservation of rings and the reaction of two independent attributes.

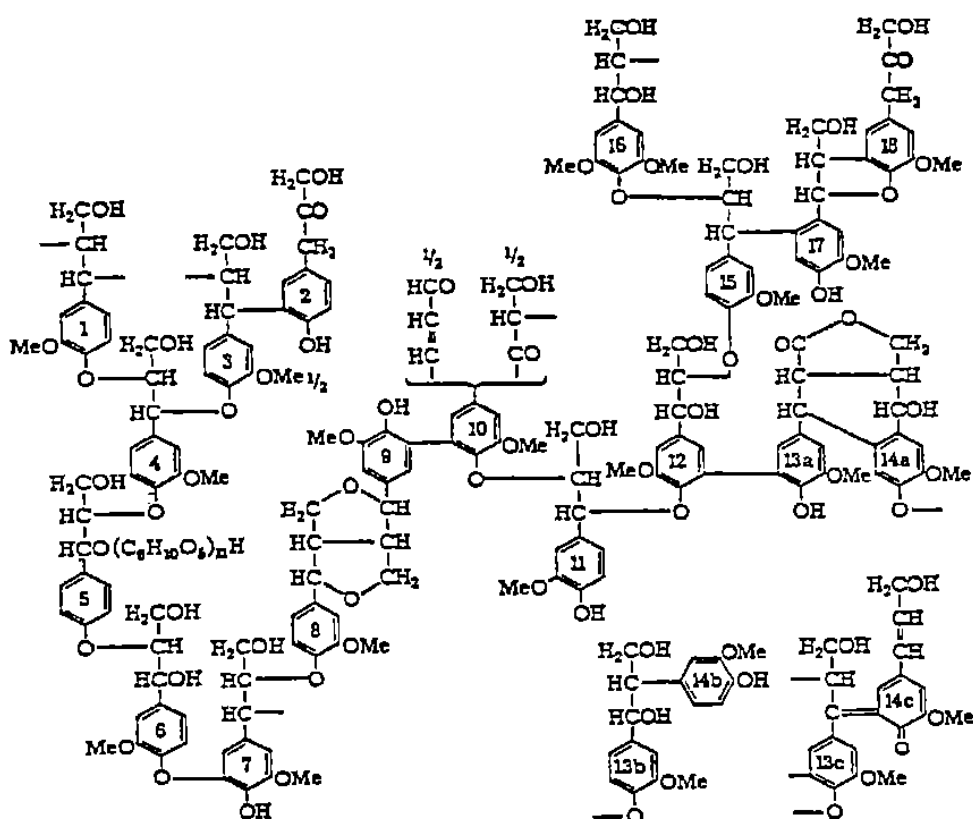


Figure 6.3 the Freudenberg model of lignin structure³

Parsing the Freudenberg structure (or analogous structural representations) provides quantitative values for the statistical distribution of the attributes. The two attributes (PC and MP) can be calculated directly by viewing each unit of the Freudenberg model as the statistical probability of the occurrence. As a result, the original Freudenberg model can be represented by the joint probability of two attributes as shown in Figure 6.4, for both the PC and MP attribute types. Note that both the PC and MP attribute sets have “free” and “etherified” subcategories. The free substituents are bonded to only one aromatic ring whereas the etherified substituents are part of an inter-aromatic ring linkage.

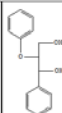
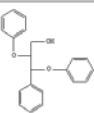
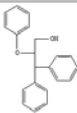
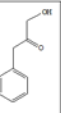
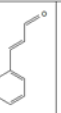






MP	PC										Sum
Normalized Distribution			0.33	0.11	0.11	0.11	0.06	0.11	0.11	0.06	1.00
 CONIFERYL	Free	0.28	1.67	0.56	0.56	0.56	0.28	0.56	0.56	0.28	5.00
	Etherified	0.56	3.33	1.11	1.11	1.11	0.56	1.11	1.11	0.56	10.00
 COUMARYL	Free	0.06	0.33	0.11	0.11	0.11	0.06	0.11	0.11	0.06	1.00
	Etherified	0.06	0.33	0.11	0.11	0.11	0.06	0.11	0.11	0.06	1.00
 SINAPYL	Free	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Etherified	0.06	0.33	0.11	0.11	0.11	0.06	0.11	0.11	0.06	1.00
Sum		1.00	6.00	2.00	2.00	2.00	1.00	2.00	2.00	1.00	18.00

Figure 6.4 Initial Lignin (Freudenberg) Structure in Terms of MP and PC Attributes²

For other lignin models, CME can get the best-fit optimization of any lignin type. In this manner, an optimal set of attribute probabilities would be generated so that the subsequently generated lignin structure provided a best fit match with the available analytical chemistry.

As discussed in Chapter 2, CME is an Excel-VBA based program which provides a friendly interface for sampling structural attributes and creating a molecular representation for complex feedstocks. Engaging CME will, at first, lead to a user-friendly interactive form, such as that shown in Figure 6.5, which will allow the user to specify and constrain the attribute types. In addition, a sample rule for lignin is implicitly generated.

AttributeNumber: 5 Change

CompID	CompName	CompAttName
1	PC1(Mol/s)	Att1
2	P17(Mol/s)	Att1
3	P24(Mol/s)	Att1
4	P16(Mol/s)	Att1
5	P22(Mol/s)	Att1
6	PC8(Mol/s)	Att1
7	PC15(Mol/s)	Att1
8	CO2(Mol/s)	Att3
9	MP7(Mol/s)	Att2
10	MP6(Mol/s)	Att2
11	MP1(Mol/s)	Att2
12	CH4(Mol/s)	Att4
13	MP15(Mol/s)	Att2
14	CO(Mol/s)	Att5
15	MP9(Mol/s)	Att2
16	MP11(Mol/s)	Att2
17	MP16(Mol/s)	Att2

AttributeList: Att5

Edit

Ok Cancel

Attribute name as Att1-5 represented five attribute types in this example: Att1 refers to PC attributes, Att2 refers to MP attributes and Att3-5 refer to free gas molecules: CO₂, CH₄ and CO. PC_i and MP_i represent different lignin structures in term of PC and MP attributes as see

Figure 6.5 Attributes Sampling in CME

As discussed in Chapter 2, CME can represent an attribute PDF as either a set of discrete values or a continuous functional form with limited parameters. Either way, it will optimize a chi-square objective function with chemical analysis terms P_{ij}^E listed below, to determine the optimal set of attribute PDF parameters and probabilities for both the MP and PC sets.

$$\chi^2 = \sum_i \sum_j \left(\frac{P_{ij}^E - P_{ij}^M}{\sigma_{ij}} \right)^2$$

Consequently, CME will provide the optimal set of attributes as an input data list for further reaction modeling.

In Chapter 2, the mechanisms behind CME were illustrated using petroleum feedstocks. Here, as a biomass example, the computational flow of CME as related to lignin prolysis is illustrated in Figure 6.6. At first, CME will provide a form and allow the user to select the MP and PC attribute types of the lignin to be modeled. If the quantitative mol fractions can be determined directly, CME will launch a straightforward mathematical parser to convert the molecular compositions to the values of each attribute types (e.g., the Freudenberg model in this chapter). For cases where the molecular compositions cannot be obtained directly, CME will apply the optimization loop of Figure 6.6. In this case, the objective function will be optimized by adjusting the parameters of the attribute PDFs. The optimized attribute values will be the final representation of the lignin and will be used as the initial conditions in the kinetic modeling.

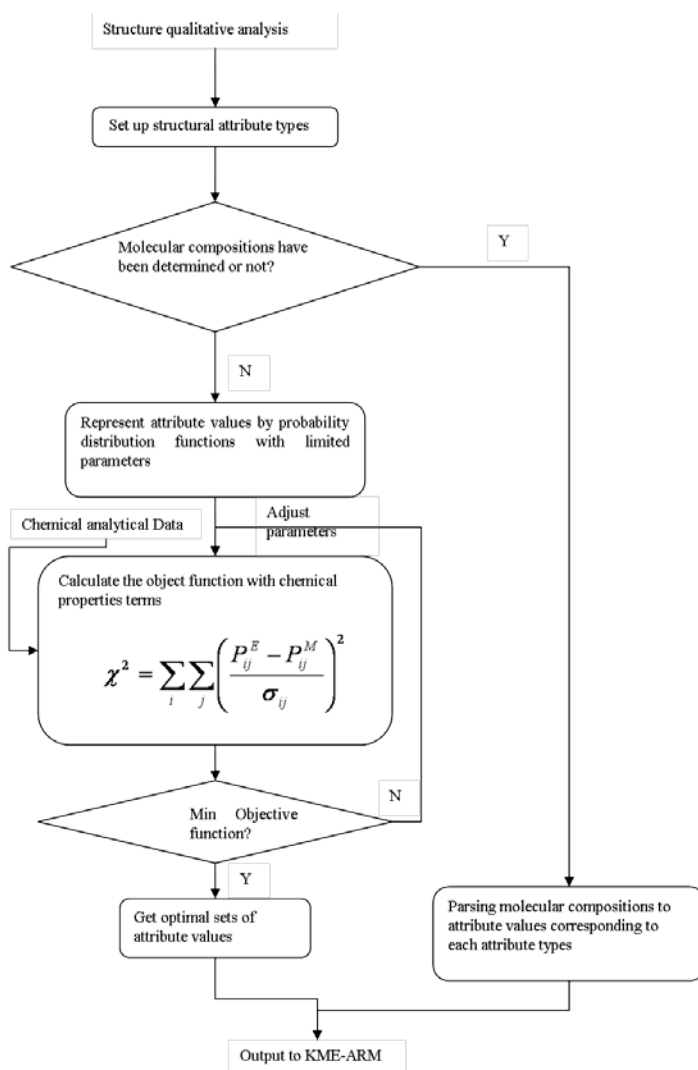
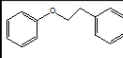
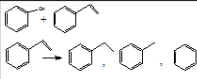

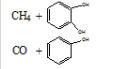

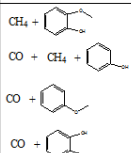
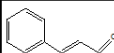
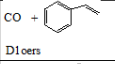

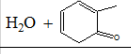
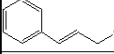
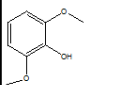
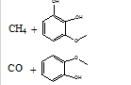
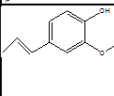
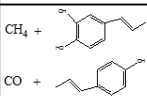
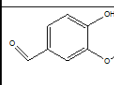
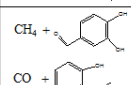

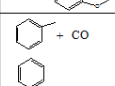
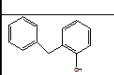
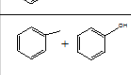
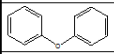
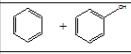
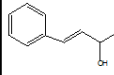
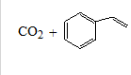
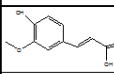
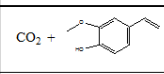

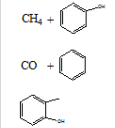

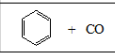


Figure 6.6 Flow sheet of CME strategy for modeling lignin structure

6.5 Lignin reaction pathways and kinetics

The structural analysis provides identities of the initially reactive moieties during lignin pyrolysis. Klein and Virk¹ studied the set of reactants and associated reaction pathways kinetics and mechanisms as listed in Table 6.1. In addition to the MCA0 shown in Figure 6.4, the full MCA can be attained from the lignin reaction pathways. The converged data base of Table 6.1 thus provides the basis for the subsequent kinetic model: the pathways reveal the reaction network and the rate parameters of the kinetics.

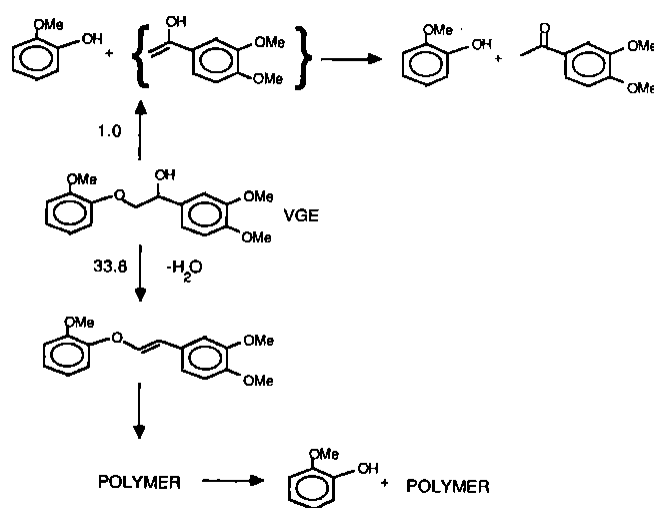
Table 6.1 Model Compounds for Lignin Pyrolysis²

Compound	Abbr	Structure	Model of	Products	log10A (1/s)	E(kcal/mol)
Phenethylphenyl ether	PPE		IL		11.1 5	45.0 22
Guaiacol	GUA		MP		10.9 11.5	43.7 47.4
Veratrole	YER		MP		13.9 14.1 14.8 11.2	55.9 58.4 60.1 49.2
Cinnamaldehyde	CAD		PC		12.1 8.6	48.2 33.7
Saligenol	SAL		PC		13.4	33.4
Cinnamyl alcohol	CAL		PC	CAD	4.2	21.8
2,6 Dimethoxy-phenol	DMP		MP		10.4 11.1	42.2 45.5
Isoeugenol	IEG		PC,MP		10.8 11.3	42.9 46.2
Vanillin	VAN		MP		12.2 10.2	47.3 38.5
Acetophenone	APH		PC		10.9 9.6	56.4 50.5
o-Hydroxy-diphenylmethane	OHD		IL		14.8	72.11
Phenylether	PET		IL		14.8	72.11
Cinnamic Acid	CIA		PC		8	21
Ferulic Acid	FEA		PC,MP		6.2	19.8
Anisole	ANI		CON		13.0 14.5 7.9	64.7 61.0 40.5
Benzaldehyde	BAD		CON		9.5	41.5
Model Codes: IL-Interunit Link, MP-Methoxyphenol PC-Propanoid Chain, CON-Control						

The reactants listed in the first column of Table 6.1 included the PC and MP attributes as well as the inter-aromatic unit links (IL) whose fragmentation leads to molecular weight reduction and the formation of single-ring phenolic products. Among the IL models is phenethylphenylether (PPE), a

beta-ether mimic to generate the hydrogen-balanced product pair phenol and styrene, whose main reaction path is illustrated in column 5 of Table 6.1. It is also shown in Table 6.1 that styrene undergoes secondary reaction to toluene, ethylbenzene, benzene and higher molecular weight adducts. A beta ether with the PPE backbone and additional lignin-like substituents as shown in Figure 6.7 is called VGE⁴. Its major reaction path is analogous to the dehydration and formation of the PPE with a double bond in the linkage. The fragmentation of this PPE analogue would not be in hydrogen balance. Likewise, the cracking reactions of the alpha ether, phenyl ether, diphenyl methane and diphenyl are not in hydrogen balance and so their fragmentation to light products consumes hydrogen elsewhere and is thus generally accompanied by the formation of an adduct.

“Beta Ether” Pyrolysis Details



18

Figure 6.7 VGE Pyrolysis Pathways⁴

The key MP mimic is guaiacol, which follows two reaction paths, one to methane and catechol and the other to phenol plus CO and H₂. The other MP models follow formally similar routes such as 2,6-Dimethoxyphenol, veratrole and anisole. The key PC mimics in Table 6.1 reveal routes to light gases (CO, CO₂, H₂), water, and various smaller hydrocarbon substituents.

In summary, Table 6.1 suggests that lignin pyrolysis products should include light gases (CH_4 , CO , CO_2 , H_2), light liquids (H_2O , MeOH), a collection of single-ring phenolics (variously substituted phenols, guaiacols and syringols), and a hydrogen-deficient char. The Arrhenius parameters and active energies in columns 6 and 7 of Table 6.1 provide an a priori basis for the prediction of the kinetics of the formation of these key products when it is employed in the material balance equations for each component. Thus Table 6.1 is directly linked to the creation of the lignin pyrolysis reaction network.

6.6 Attribute-based reaction modeling for lignin pyrolysis

The reaction products mentioned above arise as the result of the changes in state of the PC and MP substituents on the conserved aromatic ring. For this relatively simple lignin model, the number of implied molecules exceeds 624. For other feeds and more complex lignins this number will surely grow larger. Thus it seems worthwhile to develop methods to model the dynamics of the molecular composition with fewer than one equation per molecule. For that purpose we turn to the reactions of the attributes via the Attribute Reaction Model approach illustrated in Chapter 3.

As shown in Figure 6.8, a molecular composition containing N attribute types with M attribute values will have M^N molecules. The associated deterministic reaction model will thus have M^N balance equations. This can exceed 50,000 for petroleum resids and similar complex feed stocks, which is too large for many current applications. The essential idea of the ARM approach⁵ is to treat the reactions of the attributes as independent. Therefore, the ARM is a hybrid statistical model of N submodels. Each submodel has the number of equations equal to the number of values for a given attribute. Thus, the overall problem can scale as $N \times M$ instead of M^N , which is a considerable savings of computational burden. This is the approach followed in the lignin pyrolysis model, which allows convenient illustration since $N = 2$.

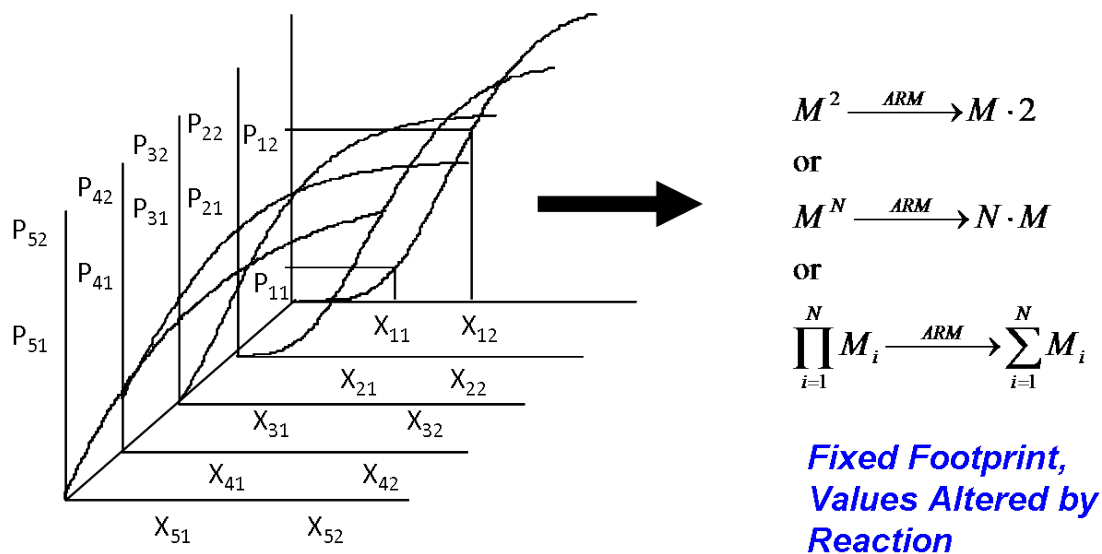


Figure 6.8 PDF sampling for molecular compositions representation

For lignin pyrolysis modeling, two independent sets of ODEs are used to describe the variations of MP and PC attributes. The temporal variations of all MP substituents can be represented by M equations as

E.q.6.1:

Attribute 1 Submodel :

$$\frac{dMP_1}{dt} = R_{11}$$

$$\frac{dMP_2}{dt} = R_{12}$$

...

$$\frac{dMP_M}{dt} = R_{1M}$$

(6.1)

Likewise, the variations of all PC substituents (PC) can be represented by N equations as E.q.6.2:

Attribute 2 Submodel :

$$\frac{dPC_1}{dt} = R_{21}$$

$$\frac{dPC_2}{dt} = R_{22}$$

...

$$\frac{dPC_N}{dt} = R_{2N}$$

(6.2)

Given initial conditions (e.g., from CME), these equations can be solved to obtain a reaction-altered distribution of attributes. Sampling these updated PDFs provides molecular information for the products with the juxtaposition of all combinations of attributes.

This logic has been incorporated into the Kinetic Modeling Editor (KME). KME with ARM support, as illustrated in Chapter 3 provided an automated Excel-based graphical interface for modeling lignin pyrolysis.

The computational flow of modeling lignin pyrolysis via KME with ARM is illustrated in Figure 6.9, whereas Figure 6.10 shows the key KME options selected for ARM support. Upon launch, KME opens a set of dialog forms that allows the user to setup modeling options such as model type, execution mode, reactor type and energy balance. For this lignin model, ARM was selected as the model type. The reaction pathways in Figure 6.7 were then entered in two separated ARM reaction sheets. Then, KME built the model and generated the C code for the ODE equations automatically. After the model was built, the model data were entered: feed data in the “INP” sheet, conditions in the “COND” sheet, and product measurements in the “OBS” sheet.

Essentially, for this lignin ARM strategy, two independent kinetic models (MP and PC) were created with shared information, such as reaction conditions, inlet flows, and the like. A final product parser created the molecular compositions as the juxtaposition of attributes as illustrated in Figure 6.11. The model was then launched to obtain the numerical solutions. After the model was solved, KME provided a set of tools to analyze the results. These included the species’ profiles through the reactor, parity plots, statistical analysis, and so on.

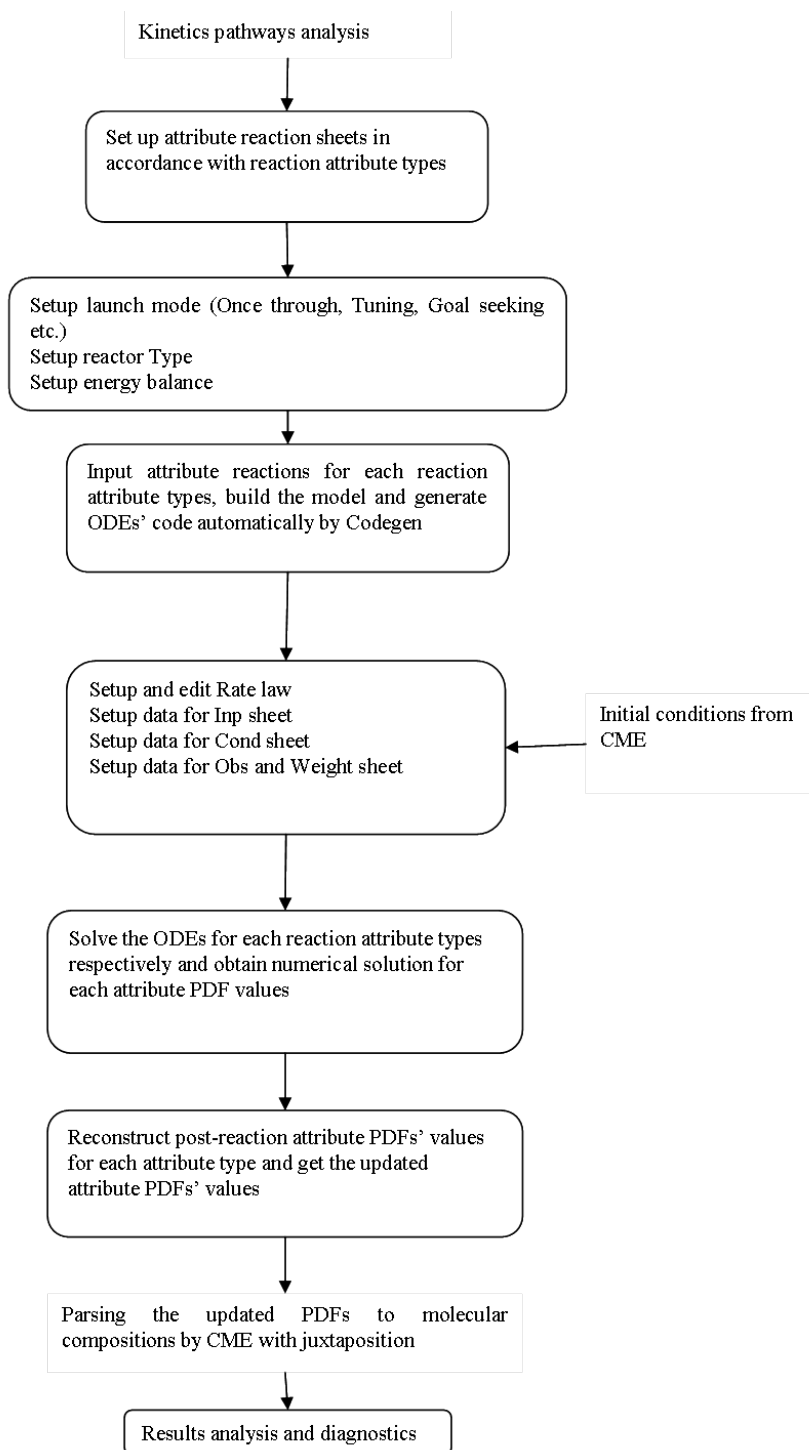


Figure 6.9 the conceptual flowsheet of KME with ARM for lignin pyrolysis

Problem Type | Model Type | Reactor Type | Temp Profile | Environment | Output | SolverSetting |

Option1: ☐ PowerLaw ☐ LHHW ☒ MicroKinetic

Option2: ☒ Individual Reactions ☐ Use Reaction Family

Option3: ☒ Use ARM Mode ☐ Non ARM Mode

☒ Isothermal ☐ User Selected Temp Profile ☐ Energy Balance

OK

Problem Type | Model Type | Reactor Type | Temp Profile | Environment | Output | SolverSetting |

☒ PFR ☐ BATCH ☐ CSTR ☐ RTR

OK

Problem Type | Model Type | Reactor Type | Temp Profile | Environment | Output | SolverSetting |

☒ Isothermal ☐ User Selected Temp Profile ☐ Energy Balance

OK

Figure 6.10 KME options

Figure 6.11 Two ARM Reaction Sets in KME

In summary, the current ARM model of lignin pyrolysis thus comprises the following steps. First, KME obtains the initial MP and PC attribute values from CME. It then generates the ODE's for the MP and PC submodels. The user then enters the input and output data, along with the reaction conditions and rate parameters, which allow the model to be run. The reaction-altered attributes are then juxtaposed to create the final product slate of light gases, light liquids, phenolics and char.

6.7 Results: application to lignin pyrolysis

The Lignin pyrolysis ARM model was constructed using the Freudenberg model for attribute input, the database of Table 6.1 for reaction pathways and kinetics, and the KME input sheets illustrated in Figure 6.11 and supplied separately as supporting material.

The reaction products, i.e., the juxtaposed attributes, are organized into four fractions: 1) light gases, 2) light liquids, 3) single-ring phenolics and 4) char, which are defined by difference. Representative results, intended only to show the nature of the CME/KME outputs, are shown in Figure 6.12 and Figure 6.13. Figure 6.12 depicts the reactor profile of the yields of the key products CH₄, CO, phenol, guaiacol and syringol. Figure 6.13, depicting the asymptotic yields of CH₄, CO, phenol, guaiacol and syringol as a function of the starting lignin's coniferyl/sinapyl alcohol ratio, shows the integrated nature of the tools and the link between CME and KME

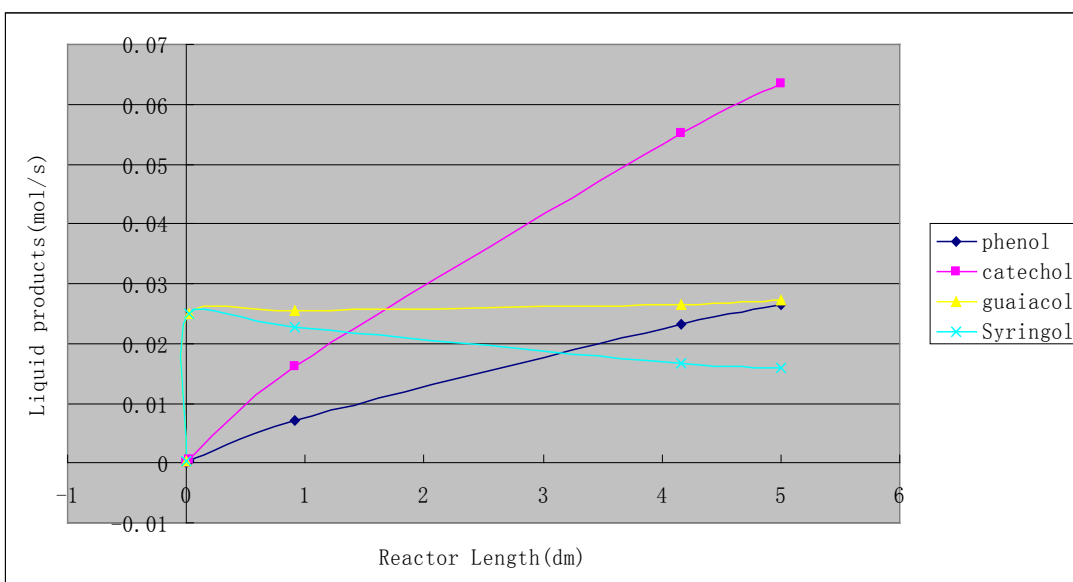
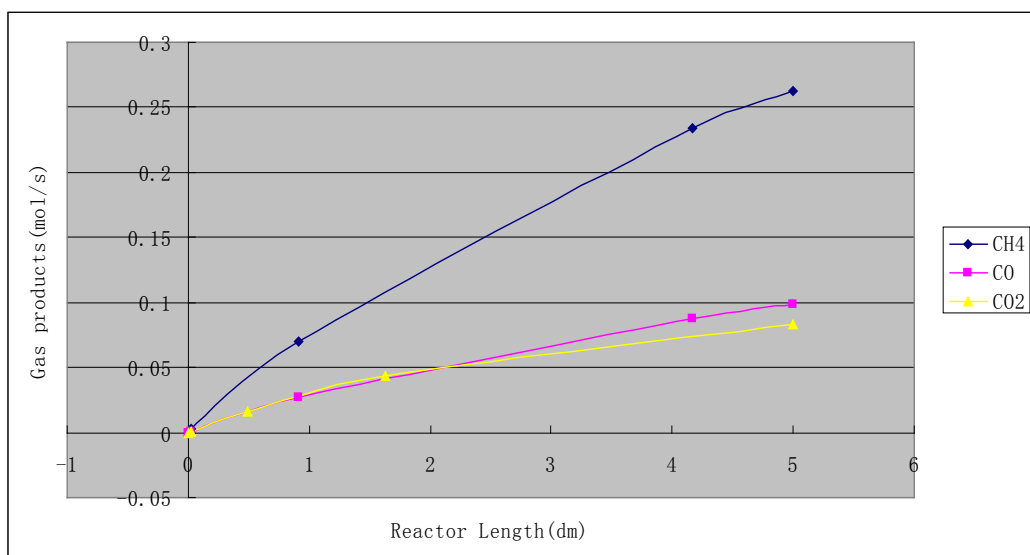


Figure 6.12 Pyrolysis products yield profile along PFR

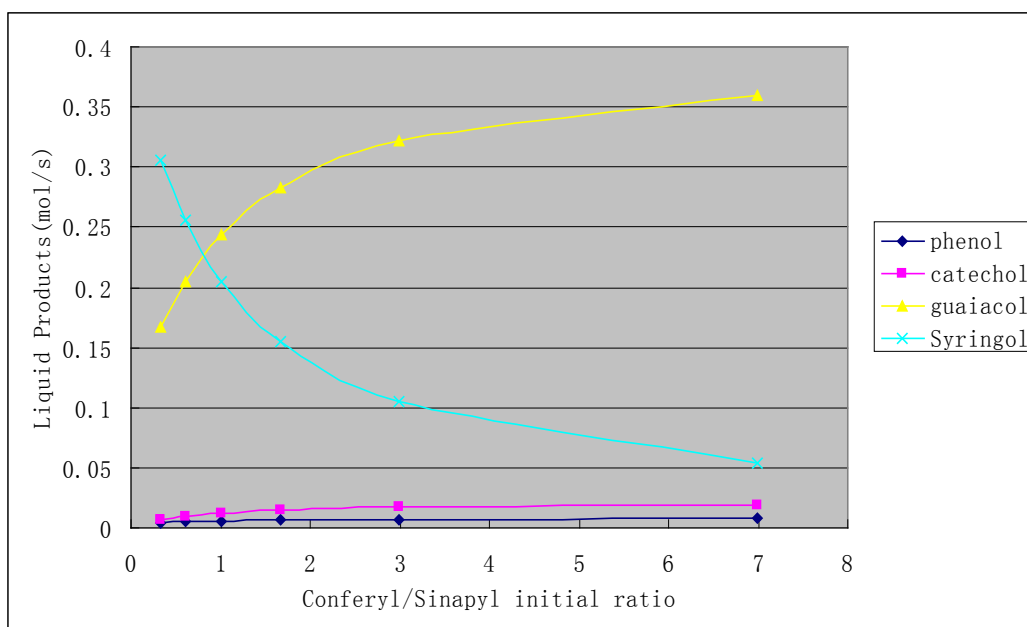
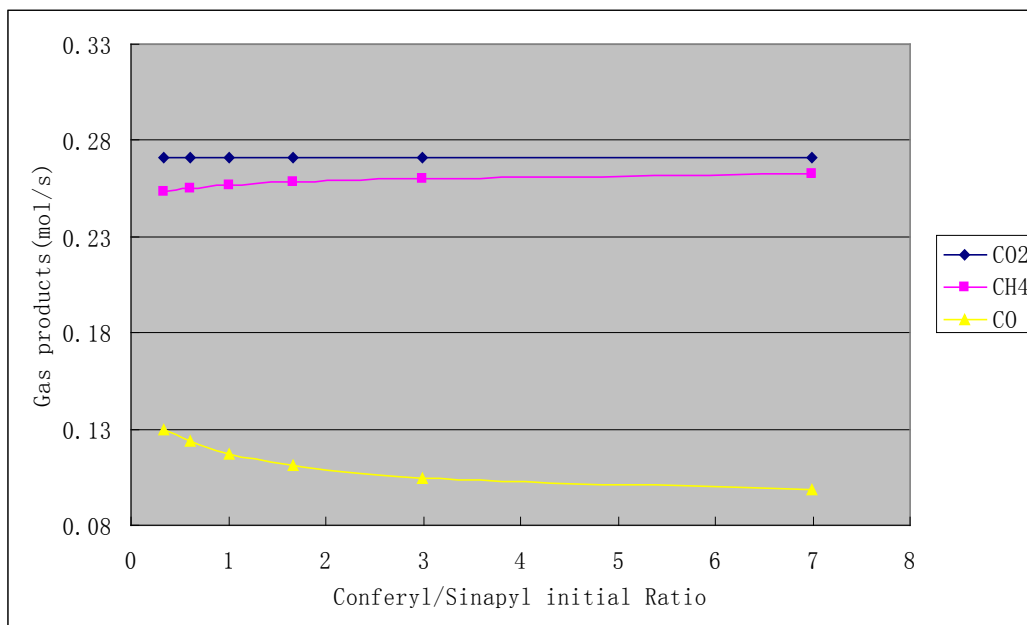


Figure 6.13 Asymptotic yields of key products parametric in the coniferyl/sinapyl alcohol ratio in starting lignin

6.8 Summary and discussions

This chapter used the application of lignin pyrolysis modeling as an example of the ARM approach mentioned in Chapter 3. The statistical approach embodied in the ARM paradigm provides a feasible solution for maintaining molecular level detail without the burden of one ODE for per a species in the model. By integrating the ARM approach with the user-friendly KME and CME tools, the complex lignin pyrolysis model can be created, solved and edited quite easily. This allows for engineering simulation of a wide range of lignins as well as for testing “what if” scenarios for lignin structure or manipulation strategies.

Reference

¹ Bell, A. T.; Gates, B. C.; Ray, D., Co-Chairs, Basic Research Needs: Catalysis for Energy, Report from the U. S. Department of Energy, Basic Energy Sciences Workshop, August 6-8, 2007.

² Klein, M.T; Virk, P.S. Modeling of Lignin Thermolysis, *Energy & Fuels* 2008, 22, 2175–2182

³ Freudenberg, K. and Neish, A.C "Constitution and Biosynthesis of Lignin", Springer-Verlag, New York (1968)

⁴ McDermott, John B., Klein, M.T., Obst, John R., Chemical modeling in the deduction of process concepts: a proposed novel process for lignin liquefaction, *Industrial & Engineering Chemistry Process Design and Development* 1986 25 (4), 885-889

⁵ Hou, Z.; Bennett, C.A.; Klein, M.T. "Attribute Reaction Model (ARM) Approach for Heavy Hydrocarbon Reaction Modeling", 235th ACS meeting, New Orleans 2007

Chapter 7 Automated modeling of resid pyrolysis with ARM

7.1 Introduction

As discussed in Chapter 1, a primary motivation for the development of this thesis is the impending worldwide energy issues. Two conceptual strategies have been studied to address this energy crisis. One is the development of new technology aimed at utilizing alternative feedstocks such as the pyrolysis of lignin as discussed in Chapter 6. The second strategy emphasizes the deep utilization of heavy hydrocarbon mixtures.

Resid pyrolysis is a heavy oil refining process that converts the heavy-end fractions to a set of light-middle fractions such as gasoline, kerosene, LCO and so on. In this chapter, resid pyrolysis is used as an example to show a conceptual strategy for developing a detailed kinetic model for such a complex, heavy chemistry.

7.2 Modeling approach

The challenge of molecule-based modeling of resid pyrolysis is derived not only from the complexity of the feedstock, but also from the complexity of the reaction modeling. There will often be thousands of “multi-functional” component species, thereby providing a challenge for detailed composition characterization in the presence of limited analytical information. In addition, the sheer number of ODEs for such a deterministic modeling problem (one ODE per specie) could exceed 50,000. Such computational problems carry a burden that can not be practically handled by current computer hardware.

To address this conflict, two key techniques have been employed in this chapter. In the first technique, CME, as discussed in Chapter 2 and Chapter 5, will provide an automated technique to characterize the

detailed compositions found in resids. The second approach will utilize the ARM methodology of KME, thereby providing a far fewer ODEs than the number of distinct species, while maintaining the full molecular detail of the model.

INGen will be employed to generate the reaction network for this pyrolysis, and finally, all these steps will be integrated in an excel-VBA based interface in order to provide an automated modeling environment for resid pyrolysis.

7.3 Model development

7.3.1 Resid composition modeling

Unlike the naphtha model discussed in Chapter 7, the complexity of a resid feedstock derives not only from isomeric details, but more relevantly from the large number of multiple ring structures and “multi-functional” components. In this case, CME’s statistical approach would be the best way to address this complexity. The automated composition modeling of two resid samples (ShengLi and DaQing) was fully discussed in Chapter 5. These two well modeled resid samples are used in this chapter as the feedstocks for modeling resid pyrolysis. In addition to the optimal MCA0, the optimal structural attribute PDFs were also obtained. However, those structural attributes are not the same as the reactive attributes used in ARM. Therefore, a further study into the reactive attributes for this resid pyrolysis modeling was necessary.

7.3.2 Reactive attributes determination

The reactive attributes are an extension of the structural attribute concept. In order to understand the concept of reactive attributes more clearly, recall the definition of the structural attribute as given in Chapter 2. A structural attribute is an element of structure defined by a PDF. The qualitative and quantitative information for a molecule can be defined by a collection of these PDFs. If any of these

structural attributes can undergo chemical transformation at the reaction conditions, it will be considered a reactive attribute.

For example, an alkyl benzene can be considered as having two structural attributes: an aromatic ring and an alkyl sidechain. Under general pyrolysis conditions, the aromatic ring will be stable but the alkyl sidechain may undergo a cracking reaction. Therefore, for such pyrolysis reactions, the alkyl sidechain would be considered as a reactive attribute but the aromatic ring would not. For other types of reactions, such as ring opening or ring saturation, the aromatic ring may also be reactive. In such cases, the aromatic ring would also be considered to be a reactive attribute.

Campbell¹ summarized a set of typical reactive attributes for various chemistries as shown in Table 7.1. In this table, only the main reactions of each process chemistry are considered. For example, while it is possible for a cracking reaction to occur on a sidechain during hydrotreating, only the reactions involving the removal of heteroatoms and the hydrogenation of aromatic rings will be considered.

Table 7.1 Reactive attributes for various chemistries¹.

Attribute	Pyrolysis	Catalytic Cracking	Hydrotreating	Reforming
Aromatic Rings	No	Yes	Yes	Yes
Naphthenic Rings	Yes	Yes	No	Yes
Alkyl Sidechains	Yes	Yes	No	Yes
Alkyl Sulfur	Yes	No	Yes	Yes
Ring Sulfur	No	No	Yes	Yes
Paraffins	Yes	Yes	No	Yes
Olefins	Yes	Yes	No	Yes

In addition to determining whether an attribute is reactive, the differences in their reactivity are need to be considered. Differences in reactivity take into account the relationship between the reaction rate and the reactive attribute. For example, although a C₅ and a C₂₀ alkyl sidechain are both the same structural attribute type, the C₂₀ will crack much quicker. In addition, while the thermal cracking of alkyl benzene and the thermal cracking of alkyl chrysene are both examples of alkyl side chain cracking on an

aromatic core, the reaction rates of one-ring and four-ring aromatics are quite different. Therefore, the reactive attributes corresponding to such reactivity differences must be considered. As a prerequisite, the structural attribute identities and the numerical values of the attributes must be defined. Therefore, although the resid samples modeled by CME consisted of only nine structural attributes, many more reactive attributes needed to be defined.

Resid pyrolysis mainly consist of the thermal cracking of a heavy feed (1000°+ F boiling fraction) to lighter products (1000°- F boiling fraction). The main pathways for this thermal conversion are the reactions of alkyl ring sidechains, paraffins, olefins, and their corresponding sulfur- containing species. As discussed in Chapter 5, the composition models for these two low sulfur feedstocks omitted structures where the sulfur was found on a side chain or a paraffin. In order to simplify this prototype model, reactions of paraffinic or olefinic species containing sulfur and ring species where sulfur is on a alkyl side chain were skipped. Typically, aromatic rings and thiophenic rings are thermally stable. Although naphthenic rings may dehydrogenate, this will typically have little effect on the product boiling point distribution. Therefore, the reactions of naphthenic rings were also ignored in order to simplify this pyrolysis model. In summary, the reactive attributes in this model are the alkyl side chains of both aromatic rings and naphthenic rings and the paraffins as shown in Table 7.2 for the two resids: ShengLi and DaQing. The carbon number ranges of the initially existing attributes (e.g., paraffins) derive from the optimized MCA0 found by CME. In addition, reactive attributes, such as olefins, that only exist in the products (MCA) are determined by the reaction network generation discussed in a later section.

Table 7.2 The carbon number ranges of the reactive attributes in a resid pyrolysis model

		DaQing	ShengLi
Paraffin	MaxCarbonNumber	87	43
	MinCarbonNumber	21	21
Naphthenics	MaxCarbonNumber	99	75
	MinCarbonNumber	20	20
Aromatics	MaxCarbonNumber	99	75
	MinCarbonNumber	24	24

Table 7.2 lists the range of the initial length of paraffins (MCA0) for the two sample feeds. The paraffin length attributes are defined from (initial) lengths of approximately 21 to 87 carbons for the DaQing model and 21 to 43 for the ShengLi model. However, each of these paraffins may crack to form smaller paraffins; lighter paraffins can also come from the cracking of alkyl sidechains. Therefore, as a result of reaction, the minimum paraffin length of a product (MCA) is simply one carbon.

Initially, there are no olefins in the system. However, they may be formed from the thermal cracking of various alkyl sidechain moieties and paraffins.

The differences in reactivity between naphthenic sidechains and aromatic sidechains were also considered. For naphthenic compounds, there were no resonance considerations. Therefore, all alkyl side chain reactions for naphthenic rings will be considered as a single reactive attribute type. The initial carbon number ranges for naphthenic rings are also shown in Table 7.2.

Aromatic compounds react much differently due to the added stability of the benzylic radicals and the instability of phenylic radicals. Therefore, it was necessary to divide the aromatic ring attribute into several reactive attributes based on the number of aromatic rings in the system. The hydroaromatic components are considered as aromatic components with a corresponding aromatic ring number. From the optimal results of CME, the aromatic ring number for the components in these two feeds is mainly above four. Therefore, in order to simplify this model, the aromatic components with one, two and three

rings are lumped into a single aromatic reactive attribute. In addition, three other reactive attributes for aromatics were specified: four rings, five rings, and six plus rings. The ranges of the initial alkyl length of these aromatic attributes are also shown in Table 7.2 for two feeds.

A sample calculation for paraffin attribute values is shown in equation 7.1:

$$\begin{aligned} &(\text{Attribute values of paraffins with carbon number } i) = \\ &(\text{paraffin mole fraction}) \times (\text{probability of paraffins with carbon number } i) \end{aligned} \quad (7.1)$$

The paraffin mole fraction and the probability of a paraffin with a given carbon number or length are obtained from the optimal attribute PDFs from CME.

The evaluation of the alkyl sidechain attribute values is a bit more complicated than the species defined by only a single attribute. Before the values of the various length sidechains for a given ring number can be determined, all components that have this given ring number but different ring configurations must be summarized. For instance, the attribute values of sidechains for two-ring naphthenic compounds can be computed by the following equation:

$$\begin{aligned} &(\text{Attribute values of two-ring naphthenic sidechains with carbon number } i) = \\ &(\text{naphthenics mole fraction}) \times \sum (\text{All fractions with two rings}) \times \\ &(\text{probability of naphthenic sidechains carbon number } i) \end{aligned} \quad (7.2)$$

The calculation of sidechain concentrations for aromatics is analogous to those of naphthenics. The only difference in the calculations for aromatics is that an aromatic core consists of aromatic rings, thiophenic rings, and naphthenic rings. Adjustments for these additional attributes are easily calculated from the results of CME.

7.3.3 Resid pyrolysis reaction network

After the reactive attributes have been determined, the next step is to obtain the reactions in this pyrolysis model. The mechanism of thermal cracking has been well studied in past years. As an example, a free radical mechanism for alkyl-aromatics pyrolysis is shown in Figure 7.1. Based on a pyrolysis mechanism, INGen² has the capability to assemble a pathway's model for resid pyrolysis. Unlike the naphtha model built in Chapter 7, the species used in INGen for this ARM model are determined from the above reactive attributes' determination in terms of CME results. Instead of using exact molecules, this ARM model uses a set of reactive attributes represented by particular species in INGen.

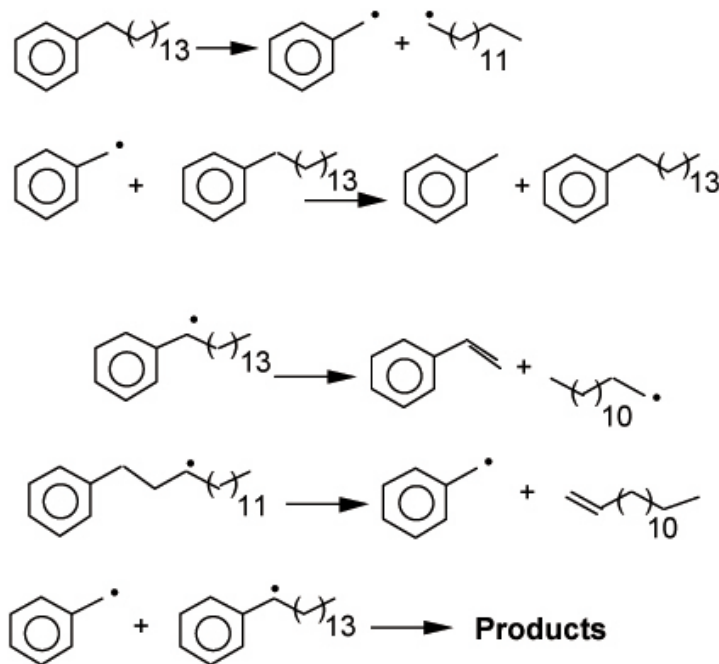


Figure 7.1 Free radical mechanism of alkyl aromatics pyrolysis³

For example, an octane molecule can be used to represent a reactive attribute of paraffin with carbon number 8. Alkyl side chain reactive attributes can also be represented by certain molecules. For example, an alkyl cyclohexane can be used to represent the reactive attributes for any naphthenic alkyl side chains. An alkyl benzene can be similarly used to describe an aromatic alkyl side chains. The reactive attributes

for various ring numbers can be recursively represented by this alkyl benzene series (through different alias names) because the reactions of those attributes will be taken independently according to the definition of ARM.

Although the reactive attributes have reduced the number of reactions significantly, the reaction model can still be quite large and unwieldy. To simplify this prototype model, several constraints were imposed upon INGen:

For paraffinic species, each bond is approximately equally reactive. However, it was assumed that all paraffins crack in the middle of the chain.

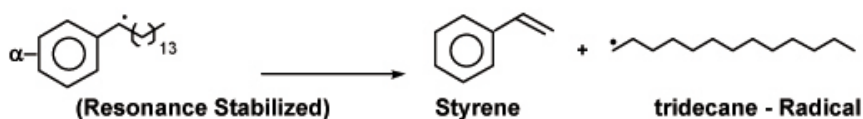
In addition to breaking at the middle of the chain, aromatic sidechains have thermodynamically favored reaction positions. As shown in

Figure 7.2, the “Highly Facile H-abstraction” and “Highly Facile beta-Scission” sites typically represent the two main reaction paths, while other reaction paths, such as reactions at the ring, are typically unfavorable. As a result, each alkyl sidechain was allowed to crack at only three positions: alpha to the ring, beta to the ring, and at the middle of the chain.

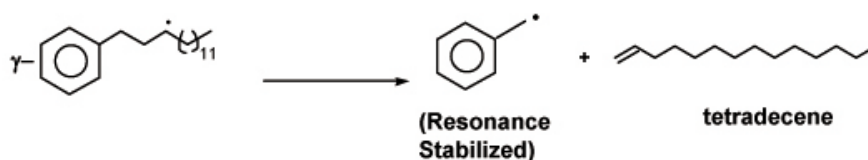
A PDB Pyrolysis Example

• Three Parallel Chains: Facility of H-abstraction vs. β -Scission

1: Highly Facile H-abstraction



2: Highly Facile β -Scission



3: Minor Pathways

Figure 7.2 An alkyl-aromatics pyrolysis pathway model (PentaDecyl Benzene-PDB)⁸

Naphthenic sidechains were considered to crack the same way as aromatic compounds. The selectivity for the various positions (alpha, beta) was considered to be equal. Overall, these rules further reduced the number of reactions.

A resid pyrolysis network was thus generated via INGen. The statistics of this reaction network are shown in

Table 7.3.

Table 7.3 The statistics of the reaction network for resid pyrolysis with ARM.

			DaQing	ShengLi
Index	RxnFamilyName	Reactants	RxnNumber	RxnNumber
1	ButAMidChainCracking	Four rings alkyl-aromatics	272	200
2	ButASideChainBetaCracking	Four rings alkyl-aromatics	91	67
3	ButASideChainGammaCracking	Four rings alkyl-aromatics	91	67
4	HexAMidChainCracking	Six rings+ alkyl-aromatics	272	200
5	HexASideChainBetaCracking	Six rings+ alkyl-aromatics	91	67
6	HexASideChainGammaCracking	Six rings+ alkyl-aromatics	91	67
7	NapMidChainCracking	Alkyl-naphthenics	274	202
8	NapSideChainBetaCracking	Alkyl-naphthenics	91	67
9	NapSideChainGammaCracking	Alkyl-naphthenics	91	67
10	ParMidChainCracking	Paraffins	145	97
11	PentAMidChainCracking	Five rings alkyl-aromatics	272	200
12	PentASideChainBetaCracking	Five rings alkyl-aromatics	91	67
13	PentASideChainGammaCracking	Five rings alkyl-aromatics	91	67
14	SngAMidChainCracking	One to three rings alkyl-aromatics	272	200
15	SngASideChainBetaCracking	One to three rings alkyl-aromatics	91	67
16	SngASideChainGammaCracking	One to three rings alkyl-aromatics	91	67
Total			2417	1769

There are 145 reactions of paraffins for the DaQing model and 97 reactions of paraffins for the ShengLi model; 456 reactions of alkyl side chains of naphthenics for the DaQing model and 336 reactions of alkyl side chains of naphthenics for the ShengLi model; 454 reactions of alkyl side chains of aromatics (with any ring number) for the DaQing model and 334 alkyl side chains of aromatics (with any ring number) for the ShengLi model. Totally, there were 2417 reactions for the DaQing model and 1769 reactions for the ShengLi model. Thus, instead of the tens of thousands of ODEs produced via deterministic modeling, the number of ODEs in this ARM model is much more practical.

7.3.4 Equations and rate constants of resid pyrolysis

After obtaining the reactions for each reactive attribute, the next step is to convert them to ODEs and implement the ARM methodology for model solving as illustrated in Chapter 3. As shown in Figure 7.3,

a molecular composition containing N attribute types with M attribute values will have M^N molecules. The associated deterministic reaction model would thus have M^N balance equations. This can exceed 50,000 for petroleum resids and similar complex feed stocks, which is too large a number for many current applications where solution time is important. The essential idea of the ARM approach⁴ is to treat the reactions of the attributes as independent. Therefore, the ARM is a hybrid statistical model of N submodels. Each submodel has a number of equations equal to the number of values for a given attribute. Thus, the overall problem can scale as $N \times M$ instead of M^N , which is a considerable savings of computational burden. This approach was followed for the resid pyrolysis model where $N = 6$ (paraffins, naphthenics, one-three rings aromatics, four rings aromatics, five rings aromatics and six ring+ aromatics).

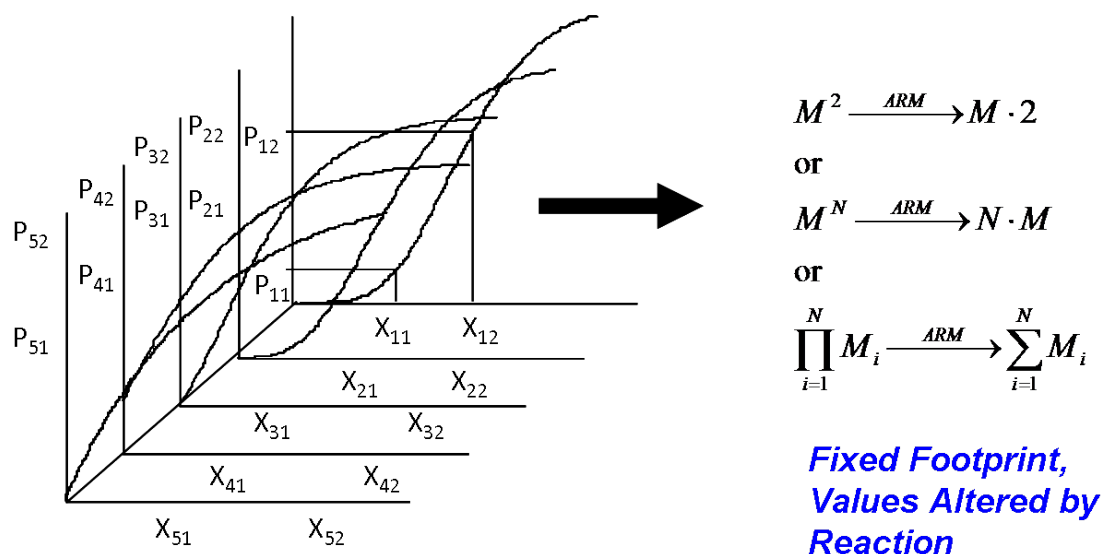


Figure 7.3 PDF sampling for molecular compositions representation

For the resid pyrolysis modeling, several independent sets of ODEs were used to describe the changes to the reactive attributes. For a given reactive attribute type i (e.g., paraffin length), the changes with time of all the values of this attribute during thermal cracking can be represented by M equations as shown in E.q.7.3:

Attribute i Submodel :

$$\begin{aligned}\frac{d\text{Att}_{i1}}{dt} &= R_{i1} \\ \frac{d\text{Att}_{i2}}{dt} &= R_{i2} \\ &\dots \\ \frac{d\text{Att}_{iM}}{dt} &= R_{iM}\end{aligned}\tag{7.3}$$

Sets of ODE submodels (like that above) were set up by KME for each attribute type in this model. The next step was to determine the appropriate kinetic parameters.

The above reaction model contains 2417 reactions for the DaQing model and 1769 reactions for the ShengLi model. The number of kinetic parameters for such a model is too huge to handle individually. Therefore, the reaction family and LFER concepts discussed in Chapter 3 is applied to this model. Nigam developed quantitative structure reactivity relationships (QSRR's) to mathematically define rate constants referenced to some model compound⁵. For this model, the complete reaction network can be defined by only a small number of reaction families regarding to reactive attributes: paraffin cracking, olefin cracking, naphthenic sidechain cracking, and aromatic sidechain cracking. The aromatic sidechain cracking is further differentiated according to ring number since the number of aromatic rings affects the selectivity of sidechain cracking.

For this reaction model, model compound data were available for each of the reaction families^{6,7,8,9}. These QSRR's were used without modification, so that the reaction model for these two resids was purely predictive.

Given the initial values from CME, and the appropriate kinetic parameters discussed above, the ODE submodels were solved in KME independently. As a result, reaction-altered distributions of the reactive attributes were obtained from the KME solution.

7.3.5 Regeneration of the attribute PDF's

After the reaction equations had been solved, the following step was to regenerate the values of the reactive attribute PDFs in order to obtain the molecular information of the products

For paraffin, olefin, and ring based alkyl side chains, the post reaction PDFs transformation process was accomplished through a straightforward renormalization of Eq. 7.4. The updated molar flows F_{ij} from the ARM model were converted to the updated probabilities P_{ij} for each PDF:

$$P_{ij} = \frac{F_{ij}}{\sum F_{ij}} \quad (7.4)$$

Where i designates the attribute type and j, the attribute value.

After renormalizing the reactive attribute PDFs, it was necessary to regenerate the structural sidechain PDFs for each core series. The side chain PDFs for non-ring core series (e.g., paraffin) were directly parsed from the corresponding reactive attribute PDF. Since the ring structures were kept intact during the reaction, the side chain PDFs for each aromatic core series for each given ring number were recalculated using a uniform distribution based on the reactive attribute PDFs for the aromatic alkyl sidechain with that given ring number. The side chain PDFs for each naphthenic core series were calculated similarly.

In addition to updating the structural attribute PDFs, for side chains, it was also necessary to recalculate the global fractions for each series. This calculation was accomplished by determining the total values of paraffins, and olefins from the reaction model results. The aromatic and naphthenic absolute values did not change. The updated global fractions were then calculated by renormalizing the new values. As a result, all elemental structural attribute PDFs including ring attributes were calculated by renormalizing those new values.

7.3.6 Product resampling

The last step of the ARM resid pyrolysis model was the conversion of the updated information in the PDFs back into molecular information. The mapping relationship between those structural attribute PDFs and their molecular information was uniquely predefined during the CME composition modeling. Therefore, a straightforward mathematical sampling process was used to assemble various combinations of structural attributes into molecules as defined by their juxtaposition.

For this resid model, eight products were defined. One was light gases (C1-C5). The other seven products were liquid fractions based on the boiling point cut: gasoline (35-91°C), naphtha (91-197°C), kerosene (197-284°C), light gasoil (284-312°C), heavy gasoil (312-368°C), vacuum gasoil (368-433°C) and residua (433+°C). Through a basic separation calculation as discussed in Chapter 3, the product molecules were separated into the above desired product stream. Thus the yields of those products were obtained. Consequentially, the desired properties of each separated product were calculated based on a molecular basis using the structure property correlations discussed in Chapter 4.

7.3.7 Model predictions

The resid pyrolysis model was used to predict the conversion and product distribution of the DaQing and ShengLi resids. These two models were simulated under typical reaction conditions (temperature: 500 C, pressure: 10MPa) and the representative results are shown in Figure 7.4, Figure 7.5, Table 7.4 and Table 7.5. The total conversion for the DaQing model was higher than that of the ShengLi model under the same conditions because the saturates and non-sulfured contents in the DaQing feed were higher than these in the ShengLi feed. These types of compounds were more easily converted when compared to the alkyl side chains attached to aggregated aromatics.

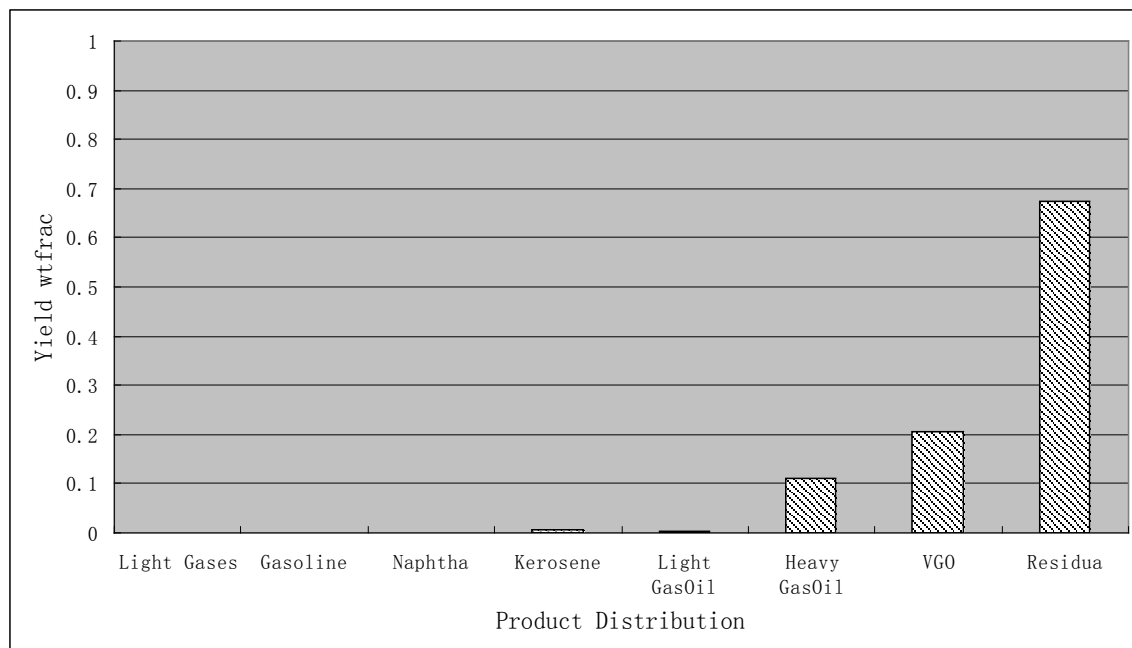


Figure 7.4 The product distribution of DaQing resid model.

Table 7.4 Product properties of DaQing resid model

	Gasoline	Naphtha	Kerosene	LGO	HGO	VGO	Residua
MW(g/mol)	83.2	130.5	168.5	211.5	345.1	469.0	775.1
H_C	1.97	1.46	1.62	1.64	1.77	1.63	1.39
C_A	0.05	0.54	0.29	0.29	0.19	0.25	0.41
C_N	0.07	0.03	0.24	0.16	0.09	0.15	0.08
WTPONA_P	0.04	0.00	0.00	0.12	0.17	0.03	0.00
WTPONA_I	0.00	0.00	0.00	0.00	0.00	0.00	0.00
WTPONA_O	0.85	0.06	0.35	0.23	0.11	0.04	0.00
WTPONA_N	0.07	0.05	0.28	0.02	0.15	0.28	0.14
WTPONA_A	0.04	0.89	0.37	0.62	0.57	0.65	0.86
IBP(K)	297	376	476	545	594	696	853
5BP(K)	300	434	478	545	617	698	937
10BP(K)	302	437	479	546	633	705	951
20BP(K)	306	447	497	546	644	717	951
30BP(K)	310	460	498	546	653	723	951
40BP(K)	314	460	498	559	661	734	951
50BP(K)	318	460	517	559	669	743	951
60BP(K)	322	461	517	559	674	755	951
70BP(K)	326	461	517	572	677	791	951
80BP(K)	330	461	517	583	683	802	951
90BP(K)	334	461	532	584	690	805	951
95BP(K)	350	461	533	585	692	808	951
EBP(K)	358	461	544	588	693	810	955

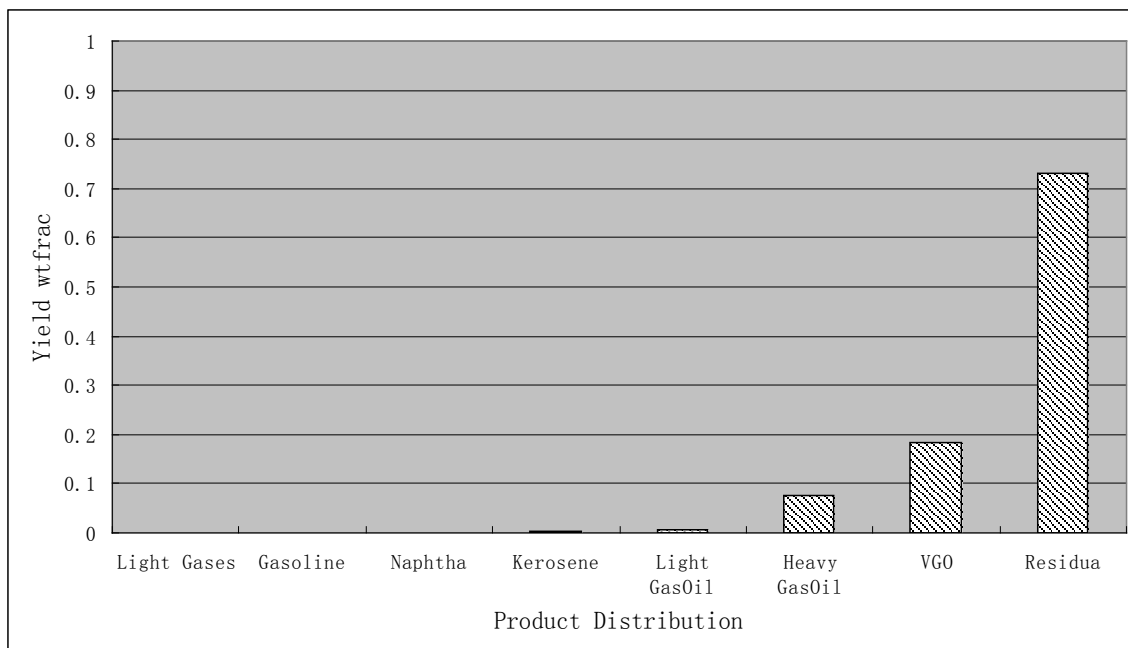


Figure 7.5 The product distribution of ShengLi resid model.

Table 7.5 Product properties of ShengLi resid model

	Gasoline	Naphtha	Kerosene	LGO	HGO	VGO	Residua
MW(g/mol)	83.6	122.1	195.7	246.8	381.2	487.8	851.5
H_C	2.02	2.00	2.00	1.99	2.02	1.83	1.55
C_A	0.00	0.00	0.00	0.00	0.00	0.13	0.32
C_N	0.00	0.50	0.04	0.18	0.08	0.11	0.01
WTPONA_P	0.05	0.02	0.05	0.03	0.36	0.19	0.00
WTPONA_I	0.00	0.00	0.00	0.00	0.00	0.00	0.00
WTPONA_O	0.95	0.32	0.87	0.47	0.36	0.04	0.01
WTPONA_N	0.00	0.66	0.08	0.50	0.28	0.58	0.04
WTPONA_A	0.00	0.00	0.00	0.00	0.00	0.19	0.95
IBP(K)	261	370	481	546	600	698	812
5BP(K)	300	398	483	546	610	699	840
10BP(K)	302	399	500	546	620	704	872
20BP(K)	306	401	501	546	645	711	888
30BP(K)	310	402	517	559	661	715	893
40BP(K)	314	404	517	560	669	721	899
50BP(K)	317	406	517	572	677	727	903
60BP(K)	321	407	520	577	677	735	907
70BP(K)	325	464	532	588	684	745	911
80BP(K)	329	465	532	589	687	759	916
90BP(K)	333	465	532	589	692	798	923
95BP(K)	335	465	532	592	694	804	928
EBP(K)	337	466	538	593	695	809	936

7.4 Summary and discussion

Thermal cracking is a classic process for converting heavy oils to lighter products. Although the mechanism of pyrolysis has been well studied in the past, it is still a challenge to model petroleum resid at molecular level due to the complexity of heavy feedstocks and the associated number of potential reactions. This chapter developed a detailed kinetic model for petroleum resid pyrolysis using the methodologies addressed in Chapter 2, 3, and 4.

Two samples (as discussed in Chapter 6) were picked up as the examples for this study: DaQing and ShengLi resid. At first, the resid feedstocks were statistically represented to obtain detailed molecular information. Next, in order to create a detailed kinetic model for resids with a practical computational burden, the ARM approach was used. Based on the reactive characteristics of resid pyrolysis, the reactive attributes were determined from the results of feedstock characterization.

In order to generate the attribute reaction network by INGen, the reactive attributes were specified as self contained molecular chemical structures. Based on a free radical mechanism, the pyrolysis reaction criteria was selected in INGen. In addition, in order to simplify the model calculation, only main reaction pathways were applied to this model. As a result, 2417 reactions for the DaQing model and 1769 reactions for the ShengLi model were generated by INGen.

Those reactions were imported to KME which generated the ARM reactions with reaction families automatically. By using the rate constants obtained as found in the literature, KME was able to run the resid model.

After simulation, the post-reaction attribute PDF results from KME were renormalized and resampled to obtain the molecules in the product stream. By a basic separation calculation and structure-property correlation, the desired properties and yields of the products were finally obtained.

The development of this resid model represents a comprehensive collaboration of the contributions addressed in Chapter 2, 3, 4, and INGen. The determination of species in this resid model was performed by CME's feed representation. Based on this information, INGen gave the attribute reaction network. Later, KME used INGen's results to perform the ARM model execution with QSRR and reaction family concepts. The logic in this chapter is particular useful for modeling the kinetics of heavier feedstocks such as complex resids, coal etc. When incorporated with other chemistries such as acid catalytic cracking, this method can be applied to more practical process chemistries.

Reference

-
- ¹ D.M. Campbell, D.M., Stochastic Modeling of Structure and Reaction in Hydrocarbon Conversion, Doctoral Dissertation, University of Delaware (1998)
- ² Craig, Bennett, “User-controlled kinetic network generation with INGen”, Doctoral Dissertation, Rutgers University(2009)
- ³ Savage, P.E., Chemical and mathematical modeling of asphaltene reaction pathways, Ph.D. Dissertation, University of Delaware (1986).
- ⁴ Hou, Z.; Bennett, C.A.; Klein, M.T. “Attribute Reaction Model (ARM) Approach for Heavy Hydrocarbon Reaction Modeling “, 235th ACS meeting, New Orleans 2007
- ⁵ Nigam, A. and M.T. Klein, A Mechanism-Oriented Lumping Strategy for Heavy Hydrocarbon Pyrolysis: Imposition of Quantitative Structure-Reactivity Relationships for Pure Components. Industrial and Engineering Chemistry Research, 1993. 32: p. 1297-1303.
- ⁶ Savage, P.E. and M.T. Klein, Asphaltene Reaction Pathways. 4. Pyrolysis of Tridecylcyclohexane and 2-Ethyltetralin. Ind. Eng. Chem. Res., 1988. 27(8): p. 1348-1356.
- ⁷ Savage, P.E. and M.T. Klein, Asphaltene Reaction Pathways. 2. Pyrolysis of n-Pentadecylbenzene. Chem. Eng. Sci., 1989. 44(4): p. 985-991.
- ⁸ Savage, P.E. and M.T. Klein, Kinetics of Coupled Reactions: Lumping Pentadecylbenzene Pyrolysis into Three Parallel Chains. Ind. Eng. Chem. Res., 1989. 44(2): p. 393-404.
- ⁹ Voge, H.H. and G.M. Good, *Thermal Cracking of Higher Paraffins*. J. Am. Chem. Soc., 1949

Chapter 8 Summary and discussions

8.1 Summary

In order to address the detailed kinetic modeling for the complex feedstocks and chemistries, this thesis developed a set of approaches and automated software tools (CME, KME and SPCM). In addition, the methodologies were successfully applied to following applications: composition model of selected complex feedstocks, ARM model for lignin pyrolysis, and resid pyrolysis.

8.1.1 Automating the modeling of momposition (CME)

A hybrid statistical approach for the modeling of a complex feedstock was developed in this thesis by optimizing a representation based on structural attributes to a set of analytical characterizations.

In most cases, the entire footprint of a complex feedstock could be straightforwardly determined via a set of finite molecular identities. A qualitative composition for the complex feedstock can be derived explicitly or implicitly from several aspects, such as: core series determination, literature, automated reaction network generation etc.

A functional module, CoreGen, has been developed to provide an interactive interface for the determination of the core series. In addition to core series determination based on fixed identities, a stochastic sampling program, MolGen, can be used as a supplemental computer aid for cases where the identities are uncertain. Non-deterministic methods will only be used for the calculation and

optimization of quantitative information, and the necessity of their use will vary from feed to feed. Thus, a huge computational burden can be avoided, thereby making this approach more feasible for practical applications.

Another functional module, FootGen, has been developed to recognize the core series identities and generate the entire footprint of this feedstock via a statistical sampling protocol based on structural attributes. As a result, FootGen uniquely maps the relationship between the probabilities of multiple attributes and the mol fractions for each molecule in this feedstock.

A third and final functional module, CompGen, has been developed in order to obtain the MCA0 (the initial identities and concentrations) of a complex feedstock. First, the appropriate forms of the structural attribute PDFs are selected, and the conditional probability constraints due to physical limitations are imposed. An optimization loop is then employed based on the juxtapositions of the mapping between molecules and multiple attributes; the objective function is given in terms of available analytical information. A structure property calculation program, PropGen, has been implemented to evaluate this objective function. Finally, the MCA0 (containing both quantitative and qualitative information of the complex feedstock) is obtained and will later serve as the initial concentration values during the kinetic modeling. After reaction modeling, the MCA0 of this feedstock will expand to the full MCA including both feed and product identities and concentrations.

A user-friendly program with an Excel-VBA interface called CME was developed to fulfill this new approach. CME automates the composition modeling for complex light to extremely heavy feedstocks by integrating the above functional modules.

In addition to applying CME to traditional petroleum feedstocks, biomass can also be modeled in CME via imposing new attribute definition and sampling protocols, altering the PDFs, and creating an appropriate objective function.

8.1.2 Automation of reaction equation generation and calculation

In addition to the automated modeling of complex feedstocks, this thesis developed new automated methodologies for handling reaction equation generation and solution for detailed kinetic models for complex systems.

To address the conflict between molecularly representing complex species and the computationally impractical number of equations thereby necessary for complex process chemistries, a hybrid statistical approach was developed in this thesis. The Attribute Reaction Model (ARM) uses independent reactive attributes to transform a set of multiplicative combinational equation into separate sets of additive equations without losing molecular detail. This process can turn the ten of thousands of ODE equations for some traditional deterministic models into a smaller, more computationally practical set of ODE equations.

Besides the ARM approach for the reduction of the number of equations, a methodology for reducing the number of parameters for a complex model was also addressed. The LFER concept is a way of organizing kinetic data based on transition state theory. Reaction families capitalize on the LFER concept by using LFER correlations for each type of reaction. Some representative results developed by Klein's group are shown in the thesis, and based on those results, an automated LFER/Reaction family algorithm was developed for KME based on the Evans-Polanyi principle.

When dealing with heterogeneous systems (e.g. all the catalyzed processes) the LHHW formalism is still the best rate law in most cases. The LHHW formalism has been extended to handle the multi-species, multi-site cases prevalent in complex process chemistries. An automated technique for using LHHW kinetics in KME was developed. By default, the rate are surface controlled, but the equations can easily be manually set for other rate-determining steps in this technique.

Complex process configurations were also considered in this thesis. A deactivation model was developed for KME that can satisfy the simulation and tuning requirements at both a laboratory and industrial level. In addition, balance equations of multiple reactor types were implemented in KME. Finally, a simple separation model was developed for KME to help serve practical industrial demand.

Finally, analytical methods were used to provide auxiliary tools for model tuning in KME. A statistical package was developed, and a model sensitivity analysis and reduction algorithm was addressed.

In summary, this thesis attempts to solve the dilemmas that arise from the complexity of real feedstocks, process chemistries and practical process configurations by implementing solutions within the KME framework. Unlike the initial version of KME¹, this new version of KME, enhanced with the above contributions, shows a much stronger capability to handle the complex model systems found in the real world.

8.1.3 Structure property correlation module

In order to provide the necessary property data for composition and reaction modeling, a structure property correlation module (SPCM) was developed in this thesis.

The first goal of the SPCM is to estimate the properties of pure components. In this thesis, every molecule in the system is represented by a Bond Electron (BE) matrix. A BE matrix is an explicit 2D representation of a molecule's structure. Structural related properties can therefore be calculated directly in a straightforward process. Thermodynamic properties are evaluated by group contribution methods and quantum chemistry software.

Two group contribution methods were applied in this thesis. The Joback method² is applicable to light feedstocks. The Gani method³, which uses totally 370 detailed structural groups, achieves better precision for heavier complex feedstocks.

The quantum chemistry tool MOPAC provides high precision for critical reactive properties (e.g., heat of formation) for both molecules and immediate species (ions, radicals). To use MOPAC's calculations, the 2D structure must first be converted to a 3D structure.

Besides these chemically relevant properties, the IUPAC name and a 2D image of each species can be generated. Finally, the properties for the pure components are entered into a database.

Finally, the thesis discussed the calculation of bulk properties based on the estimation of pure components' properties. In addition to the list of built-in properties, the SPCM allows the modeler to specify into which lump a component should be included.

8.1.4 Integration of KMT components

In addition to bringing these three major elements (composition modeling, reaction modeling and structure property correlation) to KMT, work was performed towards the integration of CME, INGen, KME and SPCM into a single kinetic modeling package. The identities of species in the system will be determined by an interaction between CME and INGen. Based on this qualitative information, INGen can then build a reaction network for the model. Meanwhile, CME with an integrated instance of the SPCM will determine an optimal set of molecular concentrations (MCA0) that will be used as the initial values during reaction modeling. KME can then generate the reaction equations based on the reaction network received from INGen. The MCA0 from CME will serve as the initial values in KME, and the SPCM provides the properties. After setting up a model using all

this information, KME will be able to compile and run the model. The new framework of KMT is shown in Figure 8.1. Except for the development of INGen, all the other elements were either developed or expanded upon in this thesis.

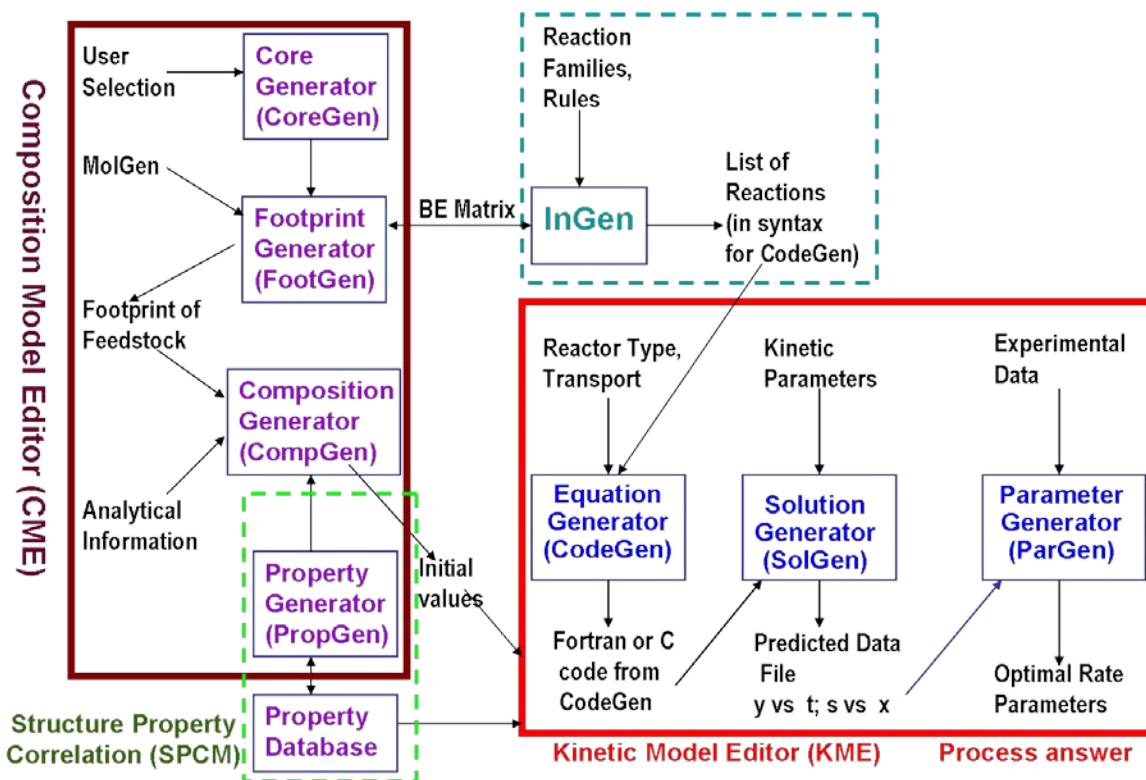


Figure 8.1 Elements of KMT

8.1.5 Composition modeling of selected complex feedstocks

Three feedstocks were used to validate the approach of automated composition modeling: petroleum middle distillates, natural gas condensates and petroleum heavy gas oil. All of those feeds were modeled based on their bulk properties, and the results were compared to the analytical compositional data. Each feedstock was evaluated using one or both of two optimization methods:

1. the feedstock was defined by a set of discrete global fractions for the homologous core series and

continuous attribute PDFs for side chain information; 2. the feedstock was defined by continuous elemental structural attribute PDFs for both core structures and side chain information.

The overall good results for the petroleum middle distillates provided validation for each optimization method and the approach in general. In addition, the results showed that the method of case 2 required less analytical information than the method of case 1. Therefore case 2 would be more useful for complex feedstocks.

The feedstocks of natural gas condensates and heavy gas oil also provided good validation for the methodology. The results and conclusions matched those of the middle distillates and also provided the first level of verification that the method could work on heavier feeds.

Two heavier feedstocks were also modeled: VGO and resid. The VGO result showed that CME could handle modeling such a complex feedstock. For the resid model, a simplified set of identities was employed. CME was aptly capable of modeling the light resid with low sulfur content. More complex resids, such as those with high-sulfur content, could be modeled by adding more heteroatom capability to CME in the future.

In summary, the CME approach was verified by analytical data, and it was able to effectively model heavier complex feedstocks with only a limited amount of analytical information.

8.1.6 Modeling lignin pyrolysis with ARM

Lignin pyrolysis was modeled as an example of the ARM approach for biomass. The statistical approach ARM paradigm provided a feasible solution for maintaining molecular level detail without the burden of one ODE per species in the model. By integrating the ARM approach with the user-friendly KME and CME tools, the complex lignin pyrolysis model was easily created, solved and edited.

8.1.7 Automated modeling of resid pyrolysis with ARM

Another conceptual modeling example was developed in this thesis in the form of a petroleum resid pyrolysis model using the ARM approach. The species' identities for this resid model were determined by CME. Based on these identities, INGen was able to generate the attribute reaction network. KME then used INGen's network to construct and execute the ARM model with the QSRR and reaction family concepts. This logic is particularly useful for modeling the kinetics of heavier feedstocks such as complex resids, coal, etc. when incorporated with other chemistries such as acid catalytic cracking, this ARM methodology can be applied to more practical process chemistries.

8.2 Suggestions for future

8.2.1 Automating the modeling of composition (CME)

Towards the goal of making CME a generic program for modeling the compositions of complex feedstocks, here are some suggestions for further study.

First, more information needs to be added for heavy feeds such as heavy resid, coal etc. Heteroatom (e.g., sulfur) should be added into the possibilities for side chains (thiols, sulfides. etc.). Because the multi-core (archipelago) model could be a better approximation than the single-core (island) model for extremely complex feedstocks (e.g., asphaltene, coal), the algorithm to sample multiple unit sheets should be added to the current sampling protocol. For example, an attribute such as “asphaltene polymerization degree” can be implemented by following Campbell⁴.

Moreover, more key structures should be incorporated for biomass. Unlike petroleum feedstocks, oxygen containing compounds such as alcohols, phenols, etc. are involved in these biofuels. Exploring new oxygen-related chemical attributes would be a key outstanding task.

In addition, other complex feedstocks such as coal, shale oil, etc. should also be addressed in future.

Finally, in order to facilitate the above extensions, several fundamental hard coding problems need to be addressed (such as a memory leak associated with the BE matrix algorithms). As it stands, the current graphical algorithm can only construct a complex core with a maximum of 16 rings. In addition, the construction of an entire footprint of a complex feed containing $O(100)$ cores (and more relevantly their BE matrix representations) may cause current computer hardware to run out of memory. To the same point, the search algorithm to parse the functional group from the BE matrix representation for an extremely aggregated multi-ring structure is also limited by memory problems.

The cause of these problems lies in the nature of a BE matrix. Since a BE matrix is a $n \times n$ matrix for a molecule with n atoms, the size of a molecule dramatically influenced the necessary computer memory. To address this issue, two strategies may work. One is to improve the BE matrix operation algorithms by optimizing computer memory in a more efficient way. Another is to change the way extremely complex molecule (e.g., asphaltene) are described digitally. Since the kinetic modeling of these feedstocks will usually be developed by ARM, the molecule could be represented by its reactive or structural attributes. Each attribute could be decomposed independently to a smaller BE matrix. Since the algorithm for parsing functional groups is based on a single BE matrix, the code would need to be able to search multiple attribute BE matrices.

8.2.2 Automation of reaction equation generation and calculation

Although the improvements to KME developed in this thesis can handle complex process chemistries by ARM, LFER, process configurations and so on, further work can still be performed to make it better. Improvements to the efficiency of KME could be performed through the redesign of the KME interface. All modes of deactivation need to be implemented at a C code level rather than VBA in order to maintain model independence and execution efficiency. Improvements need to be made such that models with $O(1000)$ ODEs are handled faster. More LFER formulations should be incorporated in the KME since they can be expressed by functions other than Evans-Polanyi correlation.

8.2.3 Structure property correlation module

The structure property correlation model could also use a number of improvements. The prediction of pure component density should be brought in line with the accuracy of the other physical

properties. This shortcoming can be reconciled by adding other advanced methods or requiring user correction of the database level. For the properties calculated from MOPAC, only the heat of formation is currently transferred automatically. Other properties (such as Gibbs energy, heat capacity, and entropy) should be integrated automatically.

Moreover, significant work can be performed to the SPCM database. Currently, the SPCM database only houses pure components' properties. It is worthwhile to study extending those properties to include reactivity data. This extension could include species (e.g., ions and radicals) seen during reactions, but also be used to illustrate the possible reactive criteria of each molecule for a certain reaction mechanism. For example, a set of reactivity data for an iso-octane could be recorded in the database for a pathCCracking (beta scission) reaction. This information could include: possible protonation sites, possible bonds to break, possible product information, related reactive properties (e.g., heat of formation) and so on.

A feasible procedure for this proposed contribution can be analogous to the algorithm for searching for multiple functional groups from a BE matrix. The lists of reactive sites and corresponding bonds can be regarded as "reactive" functional groups. The only difference between searching for such a "reactive" group and a traditional property estimation functional group is that the former requires additional information for the site position and bond information based on the canonical numbering scheme for the molecule. This work could represent another interaction between INGen and SPCM as implemented within CME. Although it entails tedious hard-coding, the upgrade would be a useful

for INGen because it would allow a user to configure the reaction through the static data manipulation rather than real time hard coding for every reaction type.

Last but not least, the SPCM could use an expanded functionality for use-supplied mixture property calculations rather than just a simple lumping function. The current algorithm of mixed property calculation is an independent c-coded functional module, and therefore provides a good basis for such a feature.

Reference

- ¹ Wei, W “The Interface of Chemical Engineering and IT in Kinetics Models”, Doctoral Dissertation, Rutgers University(2004)
- ² K.G. Joback, R.C. Reid, Chem. Eng. Comm. 57 (1987) 233–243
- ³ Marrero and R. Gani, Fluid Phase Equil. 183/184 (2001), pp. 183–208.
- ⁴ Campbell, D.M., Stochastic Modeling of Structure and Reaction in Hydrocarbon Conversion, Doctoral Dissertation, University of Delaware (1998)

