

**DESIGN AND DEVELOPMENT OF COMPUTATIONAL
INTELLIGENCE BASED METHODOLOGIES FOR CHEMICAL
ENGINEERING AND TECHNOLOGY APPLICATION**

A THESIS SUBMITTED TO
SAVITRIBAI PHULE PUNE UNIVERSITY (SPPU)

FOR AWARD OF DEGREE OF
DOCTOR OF PHILOSOPHY (Ph.D.)
IN THE FACULTY OF **CHEMICAL ENGINEERING**

SUBMITTED BY
Mr. Suhas B. Ghugare

UNDER THE GUIDANCE OF
Dr. Sanjeev S. Tambe

**CHEMICAL ENGINEERING AND PROCESS DEVELOPMENT DIVISION
CSIR-NATIONAL CHEMICAL LABORATORY
DR. HOMI BHABHA ROAD PUNE-411008, INDIA**

September 2017



राष्ट्रीय रासायनिक प्रयोगशाला
(वैज्ञानिक तथा औद्योगिक अनुसंधान परिषद)
डॉ. होमी भाभा मार्ग पुणे - 411 008. भारत
NATIONAL CHEMICAL LABORATORY
(Council of Scientific & Industrial Research)
Dr. Homi Bhabha Road, Pune - 411 008. India.



Certificate of the Guide

This is to certify that the work incorporated in the thesis titled “**Design and Development of Computational Intelligence Based Methodologies for Chemical Engineering and Technology Application**” submitted by **Mr. Suhas B. Ghugare**, for the Degree of Doctor of Philosophy, in Chemical Engineering at Savitribai Phule Pune University (SPPU), Pune, was carried out by him under my supervision in the Chemical Engineering and Process Development Division, National Chemical Laboratory, Pune – 411 008, India. Such material as has been obtained from other sources has been duly acknowledged in the thesis.

Date: September 29, 2017

Dr. Sanjeev S. Tambe

(Research Guide)

Declaration by the Candidate

I, hereby declare that the thesis entitled “**Design and Development of Computational Intelligence Based Methodologies for Chemical Engineering and Technology Application**” submitted by me for the degree of Doctor of Philosophy, in Chemical Engineering at Savitribai Phule Pune University (SPPU), Pune, is the record of work carried out by me during the period from October-2010 to September-2017 at the Chemical Engineering and Process Development Division of CSIR-National Chemical Laboratory, Pune-411008, India, under the guidance of Dr. Sanjeev S. Tambe. The thesis has not formed the basis for the award of any degree, diploma, associateship, and fellowship, titles in this or any other University or other institutions of Higher Learning. I further declare that the material obtained from other sources has been duly acknowledged in the thesis.

Date: September 29, 2017

Mr. Suhas B. Ghugare

Dedicated to
Artificial Intelligence Systems Group
NCL, Pune.

ACKNOWLEDGEMENT

I take this opportunity with deep sense of gratitude to record my sincere thanks to my research guide **Dr. Sanjeev S. Tambe**, for introducing me to the fascinating world of computational intelligence. His outlook for latest technology, accurate observations and an open mind towards novel research has driven me to achieve my goals with perfection. His command on technical writing is unparalleled and his association has created a life long learning experience for me.

I feel privileged to express my thanks to **Dr. Ashwinikumar Nangia**, Director, CSIR-National Chemical Laboratory for allowing me to carry out research work and extending all the required infrastructural facilities during my research work in National Chemical Laboratory. I also thank **Dr. B. D. Kulkarni**, Distinguished Scientist, CSIR-National Chemical Laboratory for his humble support.

Colleagues and Staff at National Chemical Laboratory (NCL), Pune were quite supportive throughout my research work. Particularly I am indebted to **Shishir Tiwary**, who was my closest colleague in most of my research work. Special thanks also to **Dr. N. N. Chavan, Dr. Vinadevi V. Patil, Dr. Renu Vyas, Vinod, Jayaram, Rupali, Aditi, Purva, Sanket** and **Mr. Makarand Y. Naniwadekar** for their valuable help as and when required.

I would like to express my gratitude towards my parent organization's **Management of All India Shri Shivaji Memorial Society**, for permitting me to pursue my higher studies, and the entire staff of Department of Chemical Engineering, for their support.

I would like to highly acknowledge the help rendered to me at various stages by my family members, especially my wife **Sujata** and daughter **Kshitija**. Their support assisted me to confront the distressful periods of this journey with courage.

Thank you all once again.

Mr. Suhas B. Ghugare

TABLE OF CONTENTS

ACKNOWLEDGEMENT	v
LIST OF FIGURES	xii
LIST OF TABLES	xvi
LIST OF APPENDICES.....	xx
ABSTRACT.....	xxi
CHAPTER 1.	
INTRODUCTION TO COMPUTATIONAL INTELLIGENCE BASED METHODOLOGIES	1
1.0. COMPUTATIONAL INTELLIGENCE (CI) AND MACHINE LEARNING	3
1.1. PROCESS MODELING.....	12
1.2. COMPUTATIONAL INTELLIGENCE BASED MODELING METHODOLOGIES	15
1.2.1. Artificial Neural Network (ANN)	15
1.2.2. Genetic Programming (GP).....	26
1.2.3. Support Vector Regression (SVR)	32
1.3. PROCESS OPTIMIZATION	38
1.3.1. Conventional Optimization Approaches	38
1.4. COMPUTATIONAL INTELLIGENCE BASED OPTIMIZATION METHODOLOGIES	40
1.4.1. Genetic Algorithm (GA)	40
1.4.2. Artificial Immune System (AIS)	48
1.5. PROCESS FAULT DETECTION AND DIAGNOSIS APPROACHES	54
1.5.1. Conventional Clustering Methodologies.....	56
1.6. COMPUTATIONAL INTELLIGENCE BASED CLUSTERING METHODOLOGIES	56
1.6.1. Fuzzy c-Means Clustering (FCC).....	57

1.7. DIMENSIONALITY REDUCTION METHODOLOGY: Principal Component Analysis (PCA)	60
1.8. STATISTICAL SIGNIFICANCE TEST METHODOLOGY: Steiger's z-Test	63
1.9. CONCLUSION	64
NOMENCLATURE	67
REFERENCES	69

CHAPTER 2.

COMPUTATIONAL INTELLIGENCE BASED MODELING OF HIGHER HEATING VALUE OF SOLID BIOMASS FUELS.....85

2.0. INTRODUCTION	87
2.0.1. Higher Heating Value.....	87
2.1. SURVEY OF BIOMASS HHV MODELS	89
2.1.1. Need for Nonlinear Models	89
2.2. CI-BASED BIOMASS HHV MODEL DEVELOPMENT.....	92
2.2.1. GPSR-Based Model Development.....	92
2.2.2. MLPNN-Based Model Development.....	94
2.3. RESULTS AND DISCUSSION.....	94
2.3.1. Biomass HHV Models Based on the Proximate Analysis	95
2.3.2. Biomass HHV Models Based on the Ultimate Analysis	97
2.4. CONCLUSION.....	101
NOMENCLATURE	102
REFERENCES	102

CHAPTER 3.

PREDICTION OF ELEMENTAL COMPOSITION OF SOLID BIOMASS FUELS FROM PROXIMATE ANALYSIS USING COMPUTATIONAL INTELLIGENCE BASED MODELS105

3.0. INTRODUCTION	107
3.0.1. Need for Non-Linear Elemental Composition Models	109
3.1. CI-BASED MODELS FOR THE PREDICTION OF ELEMENTAL COMPOSITION OF BIOMASS FUELS	112

3.1.1.	Development of GPSR-Based Biomass Elemental Composition Predicting Models	113
3.1.2.	Development of MLPNN-Based Biomass Elemental Composition Predicting Models	114
3.1.3.	Development of SVR-Based Biomass Elemental Composition Predicting Models	114
3.2.	RESULTS AND DISCUSSION	115
3.2.1.	Comparison of Models Predicting Carbon (wt%)	116
3.2.2.	Comparison of Models Predicting Hydrogen (wt%)	120
3.2.3.	Comparison of Models Predicting Oxygen (wt%)	124
3.3.	CONCLUSION	127
	NOMENCLATURE	128
	REFERENCES	129

CHAPTER 4.

DEVELOPMENT OF HIGH PERFORMING MODELS FOR THE PREDICTION OF HIGHER HEATING VALUE OF COALS OF DIFFERENT RANKS AND FROM DIVERSE GEOGRAPHIES USING COMPUTATIONAL INTELLIGENCE BASED METHODOLOGY...131

4.0.	INTRODUCTION	133
4.1.	SURVEY OF COAL HHV MODELS	133
4.1.1.	Need for Nonlinear Coal HHV Models	136
4.2.	GPSR-BASED COAL HHV MODEL DEVELOPMENT	137
4.3.	RESULTS AND DISCUSSION	138
4.3.1.	Comparison of Proximate Analysis Based Models (“as received” basis)	139
4.3.2.	Comparison of Ultimate Analysis Based Models (“as received” basis)	140
4.3.3.	Comparison of Proximate Analysis Based Models (“dry” basis)	141
4.3.4.	Comparison of Ultimate Analysis Based Models (“dry” basis)	142

4.3.5.	Comparison of Ultimate Analysis Based Models (“dry ash-free” basis).....	143
4.3.6.	Statistical Comparison of Coal HHV Models	144
4.4.	CONCLUSION.....	145
	NOMENCLATURE	146
	REFERENCES	146

CHAPTER 5.

DEVELOPMENT OF COMPUTATIONAL INTELLIGENCE BASED SOFT-SENSOR MODEL FOR STYRENE POLYMERIZATION PROCESS AND ITS APPLICATION IN A MODEL PREDICTIVE CONTROL SCHEME.....151

5.0.	INTRODUCTION	153
5.1.	PROCESS IDENTIFICATION OF POLYMERIZATION PROCESSES	154
5.2.	MODEL PREDICTIVE CONTROL OF POLYMERIZATION PROCESSES	155
5.3.	STYRENE POLYMERIZATION PROCESS.....	158
5.3.1.	Process Identification of Styrene Polymerization Process Using GPSR Methodology.....	161
5.4.	RESULTS AND DISCUSSION.....	164
5.4.1.	Performance Analysis of the GPSR Based Soft Sensor Models.....	164
5.4.2.	Model Predictive Control of Styrene Polymerization Process Using GPSR-Based Soft-Sensor Model.....	165
5.5.	CONCLUSION.....	167
	NOMENCLATURE	168
	REFERENCES	169

CHAPTER 6.

COMPUTATIONAL INTELLIGENCE BASED PERFORMANCE MODELING OF CO-GASIFICATION PROCESS USING DIFFERENT COAL-BIOMASS BLENDS173

6.0.	INTRODUCTION	175
6.0.1.	Co-gasification (COG)	176
6.0.2.	Co-gasification Modeling for Process Performance Evaluation: Need for CI-Based Models	178
6.1.	EXPERIMENTAL	185
6.1.1.	Equipments and Materials	185
6.1.2.	Experimental Procedure	186
6.2.	RESULTS AND DISCUSSION	188
6.2.1.	Effect of Biomass Types on the FBCOG Performance.....	188
6.2.2.	CI-Based FBCOG Process Modelling.....	189
6.2.3.	Statistical Comparison of the CI-Based Models	197
6.3.	CONCLUSION.....	199
	NOMENCLATURE	205
	REFERENCES	206

CHAPTER 7.

DESIGN AND DEVELOPMENT OF COMPUTATIONAL INTELLIGENCE BASED METHODOLOGIES FOR MODELING AND OPTIMIZATION OF ADSORPTIVE REMOVAL OF CHROMIUM USING SYNTHETIC POLYMER RESINS212

7.0.	INTRODUCTION	214
7.0.1.	Need of CI-Based Modeling and Optimization for Heavy Metal Ion Adsorption on Synthetic Polymer Resins	215
7.1.	EXPERIMENTAL SECTION	220
7.1.1.	Preparation of Gallic Acid-Formaldehyde-Ammonia (GFA) Resin	220
7.1.2.	Equilibrium Adsorption Experiments of Cr(VI) on GFA Resins	220
7.2.	RESULTS AND DISCUSSION	222
7.2.1.	MLPNN Based Modeling of Adsorption of Cr(VI) on GFA Resins	222
7.2.2.	CI-Based Stochastic Optimization of Adsorption of Cr(VI) on GFA Resins	223

7.3. CONCLUSION.....	225
NOMENCLATURE	236
REFERENCES	237

CHAPTER 8.

DEVELOPMENT OF COMPUTATIONAL INTELLIGENCE BASED CLUSTERING MODEL FOR FAULT DETECTION AND DIAGNOSIS IN BIOCHEMICAL PROCESS.....242

8.0. INTRODUCTION	244
8.0.1. Pattern Recognition/Clustering for Process FDD	245
8.0.2. Fault Detection and Diagnosis of Biochemical Process.....	246
8.1. BIOCHEMICAL PROCESS FOR BIOMASS GENERATION	249
8.1.1. Simulation of the Phenomenological Model	252
8.2. RESULTS AND DISCUSSION	255
8.2.1. c-means Based FDD Modeling of the Biochemical Process.....	255
8.2.2. FCC Based FDD Modeling of the Biochemical Process	256
8.3. CONCLUSION	258
NOMENCLATURE	259
REFERENCES	261

CHAPTER 9.

THESIS CONCLUSION.....264

9.0 INTRODUCTION	265
9.1 OVERALL CONCLUSION	267
9.2 FUTURE RESEARCH SCOPE.....	273
List of Publications	276

LIST OF FIGURES

Figure 1.1	A CI-based system imitating biological intelligence in supervised learning mode.....	3
Figure 1.2	Fuzzification of temperature variable using triangular membership function.....	8
Figure 1.3	Information flow in a fuzzy logic system.....	9
Figure 1.4	A biological neuron.....	16
Figure 1.5	A computational neuron.....	16
Figure 1.6	Architecture of a typical multi-layer perceptron neural network trained for a MIMO dataset.....	19
Figure 1.7	Pseudo-code of error-back-propagation (EBP) algorithm for a single hidden layer MLPNN.....	22
Figure 1.8	The symbolic regression concept of GP.....	26
Figure 1.9	Schematic of genetic programming: (a) Basic tree structure, (b) Crossover operation, (c) Mutation operation.....	28
Figure 1.10	Pseudo-code of genetic programming algorithm for symbolic regression (GPSR).....	30
Figure 1.11	Schematic of SVR using the ϵ -insensitive loss function (ϵ : tube radius, ξ, ξ' : soft margins).....	34
Figure 1.12	Schematic showing the Roulette wheel selection operation.....	44
Figure 1.13	Schematic showing the one-point crossover operation.....	45
Figure 1.14	Schematic showing the mutation operation.....	45
Figure 1.15	Pseudo-code of genetic algorithm (GA) for optimization.....	46
Figure 1.16	Types of cells generated by the immune response system.....	49
Figure 1.17	Mechanism of the clonal selection process.....	51
Figure 1.18	Pseudo-code of the clonal selection algorithm (CLONALG).....	53
Figure 1.19	Schematic of clustering of a 2-dimensional input space.....	55
Figure 1.20	Schematic showing clustering types in a 2-dimensional space: (a) hard clustering, (b) soft clustering.....	57
Figure 1.21	Pseudo-code of the fuzzy c-means (FCC) clustering algorithm.....	59
Figure 2.1	Scatter-plots of biomass HHV versus weight percentages of various components of the proximate and ultimate analysis on dry basis.....	91
Figure 2.2	Parity plots of the experimental HHVs versus those predicted by the proximate analysis based following models: (a) GPSR-Model 1, (b) MLPNN-Model 1 and (c) Parikh et al. (2005).....	96

Figure 2.3	Parity plots of the experimental HHVs versus those predicted by the ultimate analysis based following models: (a) GPSR-Model 2, (b) MLPNN-Model 2, (c) Channiwala and Parikh (2002), (d) Milne et al. (1990) and (e) Friedl et al. (2005).	99
Figure 3.1	Scatter-plots of components of the elemental composition versus components of the proximate analysis of biomass fuels; C: wt.% carbon, H: wt.% hydrogen, O: wt.% oxygen, FC: wt.% fixed carbon, VM: wt.% volatile matter and ASH: wt.% ash in biomass fuels.....	111
Figure 3.2	Parity-plots of model-predicted and the corresponding experimental values pertaining to the carbon (wt.%) predicting following models: (a) GPSR-Model 1, (b) MLPNN-Model 1, (c) SVR-Model 1, and (d) Parikh et al. (Parikh et al., 2007).	119
Figure 3.3	Parity-plots of model-predicted and the corresponding experimental values pertaining to the hydrogen (wt.%) predicting following models: (a) GPSR-Model 2, (b) MLPNN-Model 2, (c) SVR-Model 2, and (d) Parikh et al. (Parikh et al., 2007).....	123
Figure 3.4	Parity-plots of model-predicted and the corresponding experimental values pertaining to the oxygen (wt.%) predicting following models: (a) GPSR-Model 3, (b) MLPNN-Model 3, (c) SVR-Model 3, and (d) Parikh et al. (Parikh et al., 2007).	126
Figure 4.1	Comparative parity plots of model-predicted and the corresponding experimental values pertaining to coal HHVs based on proximate analysis on an as-received basis (a) GPSR-Model 1, (b) Majumder et al. (2008).	140
Figure 4.2	Comparative parity plots of model-predicted and the corresponding experimental values pertaining to coal HHVs based on ultimate analysis on an as-received basis (a) GPSR-Model 2, (b) Dulong (Selvig and Gibson, 1945).....	141
Figure 4.3	Comparative parity plots of model-predicted and the corresponding experimental values pertaining to coal HHVs based on proximate analysis on dry basis (a) GPSR-Model 3, (b) Cordero et. al. (2001).....	142
Figure 4.4	Comparative parity plots of model-predicted and the corresponding experimental values pertaining to coal HHVs based on ultimate analysis on dry basis (a) GPSR-Model 4, (b) Neveal et al. (1986).	143
Figure 4.5	Comparative parity plots of model-predicted and the corresponding experimental values pertaining to coal HHVs	

	based on ultimate analysis on dry ash-free basis (a) GPSR-Model 5, (b) Boie (1953).....	144
Figure 5.1	Conceptual block-diagram of an MPC scheme.....	157
Figure 5.2	Schematic of the continuous polystyrene reactor process.....	159
Figure 5.3	Stimulus-response profiles of the polystyrene reactor process, (a) random sequence in input (coolant feed-flowrate), (b) response of effluent polymer molecular weight and the GPSR-Model 1 predicted polymer molecular weight (inset: enlarged view).....	162
Figure 5.4	Controlled variable profiles of the simulated MPC and PID control schemes for the polystyrene reactor process, (a) Set-point tracking, (b) Disturbance rejection.....	166
Figure 6.1	Scatter-plots of FBCOG performance variables (<i>TGY</i> , <i>CCE</i> , <i>HV</i> and <i>CGE</i>) versus important feed/process operating parameters (temperature, air/fuel ratio, steam/fuel ratio and fuel feed rate).....	182
Figure 6.2	Schematic of the FBCOG pilot plant setup.....	185
Figure 6.3	Effect of biomass type and its composition in coal-biomass feed blends on the co-gasification process, (a) <i>TGY</i> versus biomass percentages, (b) <i>CCE</i> versus biomass percentages, (c) <i>HV</i> versus biomass percentages, and (d) <i>CGE</i> versus biomass percentages.....	188
Figure 6.4	Parity plots of experimental versus GPSR-based model-predicted values of the FBCOG performance variables, namely <i>TGY</i> (kg/kg fuel) (panel a), <i>CCE</i> (%) (panel b), <i>HV</i> (MJ/Nm ³) (panel c) and <i>CGE</i> (%) (panel d).....	194
Figure 6.5	Parity plots of experimental versus MLPNN-based model-predicted values of the FBCOG performance variables, namely <i>TGY</i> (kg/kg fuel) (panel a), <i>CCE</i> (%) (panel b), <i>HV</i> (MJ/Nm ³) (panel c) and <i>CGE</i> (%) (panel d).....	196
Figure 6.6	Parity plots of experimental versus SVR-based model-predicted values of the FBCOG performance variables, namely <i>TGY</i> (kg/kg fuel) (panel a), <i>CCE</i> (%) (panel b), <i>HV</i> (MJ/Nm ³) (panel c) and <i>CGE</i> (%) (panel d).....	197
Figure 7.1	Variation of percent adsorption of Cr(VI) ions on various grades of GFA reins with pH of the solution.....	221
Figure 7.2	Parity plots of experimental versus MLPNN model predicted values of the percent adsorption of Cr(VI)ions on various grades of the GFA resin at different pH.....	223

Figure 8.1	Block diagram of the continuous bioreactor generating biomass with controllers.....	250
Figure 8.2	Nonlinear projection of 3-dimensional steady-state bioreactor data in 2-dimensions: (a) Biomass concentration (X) versus Jacket temperature (T_J), (b) Bioreactor temperature (T) versus Jacket temperature (T_J), (c) Biomass concentration (X) versus Bioreactor temperature (T).	255
Figure 8.3	Plots of membership values of the U-matrix of the developed FCC-based bioreactor FDD model: (a) Training-set data, (b) Test-set data.....	258

LIST OF TABLES

Table 1.1	Recent chemical engineering/technology applications of fuzzy logic	10
Table 1.2	Classification of commonly used ANN architectures	17
Table 1.3	Recent chemical engineering/technology applications of multi-layered perceptron neural network	23
Table 1.4	Recent chemical engineering/technology applications of genetic programming	31
Table 1.5	Recent chemical engineering/technology applications of support vector regression.....	36
Table 1.6	Recent chemical engineering/technology applications of genetic algorithms	47
Table 1.7	Recent chemical engineering/technology applications of clonal selection algorithm	53
Table 1.8	Recent chemical engineering/technology applications of fuzzy clustering	59
Table 1.9	Recent chemical engineering/technology applications of principal component analysis	62
Table 2.1	Architectural details of developed MLPNN-based biomass HHV models.....	94
Table 2.2	Comparison of biomass HHV prediction performances of GPSR-Model 1, MLPNN-Model 1 and Parikh et al's model based on the proximate analysis as input	95
Table 2.3	Range-wise comparison of biomass HHV prediction accuracies and generalization performances of proximate analysis based models	97
Table 2.4	Comparison of biomass HHV prediction accuracies and generalization performances of the GPSR-Model 2 and MLPNN-Model 2 with the models of Channiwala and Parikh (2002), Milne et al. (1990) and Friedl et al. (2005) based on ultimate analysis	98
Table 2.5	Range-wise statistical analysis of HHV prediction accuracy and generalization performance of ultimate analysis based models	100
Table 3.1	Details of the proximate and ultimate analyses dataset of biomass fuels	109
Table 3.2	Dataset partitioning and architectural details of the developed MLPNN-based biomass elemental composition models.....	114

Table 3.3	Dataset partitioning and parameter details of the developed SVR-based biomass elemental composition models.....	115
Table 3.4	Comparison of model performances of CI-based and existing linear-regression models for the prediction of carbon (wt%).....	117
Table 3.5	Range-wise comparison of the performance of CI-based and existing linear-regression models for prediction of carbon (wt%).....	118
Table 3.6	Results of the statistical comparison of the correlation coefficients pertaining to CI-based models predicting the weight percent of carbon	120
Table 3.7	Comparison of model performances of CI-based and existing linear-regression models for prediction of hydrogen (wt%)	121
Table 3.8	Range-wise comparison of the performance of CI-based and existing linear-regression models for prediction of hydrogen (wt%).....	122
Table 3.9	Results of the statistical comparison of the correlation coefficients pertaining to CI-based models predicting the weight percentage of hydrogen	124
Table 3.10	Comparison of model performances of CI-based and existing linear-regression models for prediction of oxygen (wt%).....	125
Table 3.11	Range-wise comparison of the performance of CI-based and existing linear-regression models for prediction of oxygen (wt%).....	125
Table 3.12	Results of the statistical comparison of the correlation coefficients pertaining to CI-based models predicting the weight percent of oxygen	127
Table 4.1	A comprehensive list of the correlations for predicting HHV of coal.....	135
Table 4.2	Comparison of prediction performances of proximate analysis based coal HHV models using “as-received” basis data.....	139
Table 4.3	Comparison of prediction performances of ultimate analysis based coal HHV models using “as-received” basis data.....	140
Table 4.4	Comparison of prediction performances of proximate analysis based coal HHV models using “dry” basis data.....	142
Table 4.5	Comparison of prediction performances of ultimate analysis based coal HHV models using “dry” basis data.....	143
Table 4.6	Comparison of prediction performances of ultimate analysis based coal HHV models using “dry ash-free” basis data.....	144

Table 4.7	Results of the statistical comparison of the prediction performances pertaining to GPSR-based coal HHV models.....	145
Table 5.1	Process parameters/variables of the phenomenological styrene polymerization reactor model.....	160
Table 5.2	Prediction performance of the GPSR-based soft-sensor models for styrene polymerization process.....	164
Table 6.1	Existing phenomenological models of co-gasification processes.....	180
Table 6.2	Operating parameters of the FBCOG process.....	187
Table 6.3	Analysis of the solid fuel feed samples.....	187
Table 6.4	Mean and standard deviation of thirteen inputs and four outputs.....	190
Table 6.5	Comparison of CI-based models for the prediction of four FBCOG performance variables.....	193
Table 6.6	Architectural details of the MLPNN-based FBCOG models.....	195
Table 6.7	Details of SVR-based FBCOG models.....	197
Table 6.8	Results of Steiger's z-test comparing correlation coefficients pertaining to CI-based models for co-gasification performance variables.....	199
Table 7.1	Existing ANN-based models for predicting the amount of Cr(VI) ions adsorption on different adsorbent materials.....	219
Table 7.2	Structural details of the developed MLPNN-based model predicting the percent adsorption of Cr(VI) ions on the GFA resin.....	222
Table 7.3	Statistical performances of MLPNN-based model predicting the percent adsorption of Cr(VI) on the GFA resin.....	223
Table 7.4	Selected parameters of AIS-CLONALG implemented for the grade optimization for Cr(VI) ions adsorption on GFA reins.....	225
Table 7.5	Selected parameters of GA implemented for the grade optimization for Cr(VI) ions adsorption on GFA reins.....	225
Table 7.6	Optimized grade conditions for the GFA resin synthesis and adsorption process pH for total adsorption of the Cr(VI) ions by the resin.....	225
Table 8.1	Existing fault detection and diagnosis studies pertaining to bioprocesses.....	248
Table 8.2	Process parameters and variables of the phenomenological bioreactor model.....	251
Table 8.3	Data structure for the training and test sets.....	254

Table 8.4	Magnitudes of state variables in faulty and normal conditions....	254
Table 8.5	Parameters and settings used for the c-means clustering of the bioprocess data-set.	256
Table 8.6	Parameters used for the FCC-based clustering of the bioprocess data-set.	256
Table 8.7	Comparative results of c-means and FCC-based clustering of the bioprocess data-set.....	257

LIST OF APPENDICES

Appendix 6.A	Experimental dataset of the pilot plant fluidized bed co-gasification process	201
--------------	---	-----

Appendix 7.A

Table 7.A.1	Experimental dataset of Cr(VI) adsorption on different grades of GFA resins	227
Code 7.A.1	MATLAB code for CLONALG-based optimization of percent adsorption of Cr(VI) by GFA resin	229

ABSTRACT

The chemical industry is facing daunting challenges arising from factors such as global competition, strict environment protection laws, complexity of modern processes, and maintaining high and consistent product quality. This has compelled chemical engineers to design and operate chemical processes in an efficient, safe, environment-friendly and optimal manner. Modeling, simulation, monitoring, and optimization of chemical reactions/processes/systems assist the chemical engineers and plant personnel in this difficult task. Accordingly, a number of strategies and tools are being continuously developed in chemical process engineering for optimal design and operation of chemical processes.

Modeling of a chemical process/reaction is essential for a number of tasks such as control, monitoring, optimization and, fault detection and diagnosis. Commonly, reaction/reactor/process models are constructed using *phenomenological* and *empirical* approaches. The first approach requires complete understanding of the thermodynamics, heat and mass transport phenomena, and kinetics underlying the chemical process. Often, the said knowledge is difficult, time-consuming, costly and tedious, to acquire and invariably needs extensive experimentation. The second i.e., empirical approach, entails development of linear/nonlinear regression models and requires extensive guesswork of the structure (form) of the data-fitting function and its parameters, especially for nonlinear systems.

Optimization is another important task in process engineering. It is conventionally performed using deterministic gradient-based methodologies. These have a prerequisite that the objective function to be maximized/minimized must be smooth, continuous and differentiable. Another major shortcoming of the deterministic gradient-based methods is that they invariably get stuck in to a local optimum, thus producing a sub-optimal solution.

The limitations of phenomenological and conventional empirical modeling and deterministic gradient-based optimization methods necessitated a paradigm shift in the approach to modeling and optimization of chemical

reactions/processes/systems. In recent years, various *Computational Intelligence* (CI) based methodologies have offered attractive avenues for modeling and optimization of chemical reactions/reactors/processes as well as other process engineering tasks such as process monitoring, control and fault detection and diagnosis. Often, the CI-based modeling and optimization methods have been found to be more efficient, cost-effective, and easier to implement than their conventional counterparts. Accordingly, the primary focus of the work presented in this thesis is the design and development of computational intelligence based modeling and optimization methodologies for chemical engineering applications. The specific CI methodologies used in conducting the modeling studies are *artificial neural networks* (ANNs), *genetic programming* (GP), *fuzzy logic* (FL), and *support vector regression* (SVR), while the CI-based optimization methods employed are *artificial immune systems* (AIS) and *genetic algorithms* (GA).

The thesis is organized in nine chapters and the studies presented in each chapter are described below in brief.

Chapter 1 first introduces the concept of computational intelligence along with its importance and applications to chemical engineering and technology systems. The main categories of CI with methodologies included in each category are discussed in brief along with their applications. The specific chemical engineering tasks performed using CI-based methods along with the corresponding conventional approaches are discussed next. The following CI-based methodologies employed in this thesis are next elaborated along with their pertinent applications.

1. Modeling Methodologies: *Multilayered perceptron neural network* (MLPNN), *Genetic programming based symbolic regression* (GPSR), *Support vector regression* (SVR) and *Fuzzy c-means clustering* (FCC).
2. Optimization Methodologies: *Clonal selection algorithm* (CLONALG) based on *Artificial immune systems* (AIS) strategy, and *Genetic algorithm* (GA).

Lastly, this chapter describes a dimensionality reduction methodology, namely *principal component analysis* (PCA) as also a statistical test namely *Steiger's z-test* used in comparing the models.

The higher heating value (HHV) is an important property of solid fuels such as coal and biomasses. It is used extensively in designing and developing combustion and gasification processes using coal and/or biomass as fuels. In Chapter 2, we introduce development of CI-based methodologies for modeling the HHV of solid biomass fuels from their proximate or ultimate analyses. The HHV of a biomass fuel is determined experimentally using a high pressure bomb calorimeter. This involves tedious calorimetric experiments. A detailed literature survey indicates that commonly, linear regression based models have been developed for the prediction of HHV of biomass fuels from their proximate or ultimate analyses, although in reality nonlinear dependencies are witnessed between the HHV and a few components of the corresponding proximate and ultimate analyses. Accordingly, novel CI-based methodologies namely, *genetic programming* (GP) and ANNs has been introduced for the nonlinear modelling of the HHV of solid biomass fuels using a large database of the proximate and ultimate analyses of biomasses and their corresponding HHVs. A total of four CI-based biomass HHV models—two each using GP and ANN methods—were constructed. All the four CI-based biomass HHV prediction models produced higher prediction and generalization performance than their existing counterparts as observed from high correlation coefficients (*CCs*) (>0.96) and low mean absolute percent errors (*MAPEs*) ($<4.5\%$).

Chapter 3 presents design and development of computational intelligence based modeling methodologies for predicting the elemental composition of solid biomass fuels from their proximate analysis. The said elemental composition comprises carbon (*C*), hydrogen (*H*) and oxygen (*O*) as major components and is an important measure of the biomass fuel's energy content. Its knowledge is important in the efficient design and operation of combustion and gasification systems using these fuels, and pollution control thereof. As done conventionally, determining the elemental composition of a biomass fuel through its ultimate analysis is time-consuming and expensive. In comparison, proximate analysis,

which determines fixed carbon, ash, volatile matter and moisture content of a biomass fuel is easier to perform and can be related to the ultimate analysis of the fuel. A few linear models predicting the said elemental composition are available in the literature. However, often the weight percentages of C , H , and O as a function of the weight percentages of the constituents of the proximate analysis, indicate nonlinear relationships between them. Accordingly, this study utilizes three computational intelligence (CI) formalisms, namely, GP, ANN and SVR, for developing nonlinear models for the prediction of C , H , and O percentages of solid biomass fuels. All the developed nonlinear CI-based models predicting the elemental compositions of biomass significantly and consistently outperformed their linear counterparts in terms of higher prediction accuracies and generalization capabilities.

In Chapter 4, genetic programming has been introduced for the development of high performing models for the prediction of higher heating value of coals of different ranks, and from diverse geographies of the world. There exist mostly linear regression based models correlating the constituents of a coal's proximate and/or ultimate analysis to its HHV as also a few CI-based models based on the ANNs, *co-active neuro-fuzzy adaptive networks* (CANFIS), *alternating conditional expectation* (ACE), and SVR formalisms. A literature search indicates a continuing effort in developing HHV correlations for coals, with its emphasis on coals of a specific rank and/or from specific geographical regions; mostly these efforts were based on limited amounts of data. Towards developing generalized correlations encompassing several ranks of coal, the *genetic programming based symbolic regression* (GPSR) formalism has been introduced for modeling HHV of coals based on a large-sized dataset comprising coals of various ranks and from diverse geographies in the world. Specifically, five GPSR-based coal HHV predicting nonlinear correlations have been developed and each one of them possesses following notable features: (i) higher prediction accuracy and generalization performance with high correlation coefficients ($CCs > 0.99$) and root mean squared error ($RMSEs < 0.8$) as compared to the existing HHV models based on proximate or ultimate analyses (on as-received basis), (ii) much wider applicability for different rank coals from diverse

geographies, and (iii) due to their lower complexity, GPSR-based models are easier to understand and deploy in a practical setting than the other CI-based models.

In Chapter 5, a GPSR-based approach has been proposed for the development of a soft-sensor model for styrene polymerization process and its application in a model predictive control scheme. A soft-sensor essentially is a mathematical model that relates the key process variables, which are difficult to measure online, to those process variables whose continuous online measurement is easy and accurate. In general, polymerization reactions are difficult to model as they involve complex nonlinear reaction phenomena with numerous kinetic parameters, multiple phases, nonlinear thermodynamics and heat and mass transfer mechanisms, possible non-homogeneity of the reaction mass due to non-ideal mixing, and non-uniform polymer particle size and molecular weight distribution. The styrene polymerization reaction exhibits all these traits and, therefore, poses difficulties to model accurately using conventional phenomenological approach. Generally, the final polymer product quality is evaluated in terms of its average molecular weight based on the molecular weight distribution (MWD). The molecular weight and consequently the yield and quality of the polymer are highly sensitive to the reaction conditions. Determination of MWD is commonly done offline in a quality control laboratory and is costly, and time consuming. Development of a soft-sensor model using adequate historical process data is expected to overcome the said difficulties since the prediction of the difficult to measure variable/parameter is almost instantaneous. Accordingly, two GPSR-based soft-sensor models were developed for identification of an industrial continuous styrene polymerization process. These models dynamically predict the molecular weight of the effluent polystyrene product and possess very high prediction and generalization capability. Next, the best performing GPSR model was successfully utilized in designing a model predictive control (MPC) scheme for the styrene polymerization process with an objective to accurately control the effluent polymer product quality. When compared with an appropriately tuned proportional-integral-derivative (PID) controller, the MPC

scheme utilizing the GPSR-based model was found to yield improved performance for set-point tracking and disturbance rejection tasks.

Chapter 6 reports a study, wherein experimental data, from the co-gasification (COG) experiments conducted in a pilot-plant scale fluidized bed gasifier located at Council of Scientific and Industrial Research's (CSIR), *Central Institute for Mining and Fuel Research* (CIMFR) Laboratory, Dhanbad, India, were utilized for the design and development of CI-based models for the COG process. Specifically, co-gasification experiments were conducted using a number of blends of high ash Indian coals and biomasses (rice husk, saw dust and press mud) separately. The CI-based models utilizing the ANN, GP and SVR methodologies were developed to predict the effect of thirteen variables describing COG process conditions and feed characteristics on the four key process performance variables, namely *total gas yield* (TGY) (kg/kg fuel), *carbon conversion efficiency* (CCE) (%), *heating value of product gas* (HV) (MJ/Nm³), and *cold gas efficiency* (CGE) (%). The original experimental dataset consisting of thirteen input variables was compressed to obtain four linearly uncorrelated inputs using principal component analysis (PCA) and this PCA transformed dataset was used for developing the models. All the developed CI-based COG process models exhibit good prediction accuracy and generalization performances. Finally, a statistical test termed "*Steiger's z-test*" was performed to compare CI-based models and identifying the best ones predicting the four COG performance variables.

In Chapter 7, a CI-based hybrid approach for modeling and optimization of a reaction involving adsorptive removal of chromium ions from waste water using gallic acid-formaldehyde-ammonia (GFA) synthetic polymer resins has been proposed. In the first step, the experimental data consisting of the resin synthesis reaction conditions, adsorption pH, and the corresponding percent adsorption of Cr(VI) were used to develop an optimal multilayer perceptron neural network (MLPNN) model; this model shows an excellent prediction and generalization performance. Next, the *artificial immune systems* (AIS)-inspired *clonal selection algorithm* (CLONALG) was used for optimizing the 3-dimensional input space of the above-stated MLPNN model with the objective of maximizing the Cr(VI)

adsorption on the resin. For the purpose of comparison, optimized resin composition and reaction conditions were also obtained using a widely used CI-based stochastic optimization method, namely *genetic algorithm* (GA). Finally, the optimized resin composition and reaction conditions obtained using the MLPNN-AIS hybrid modeling-optimization strategy were validated experimentally; the results obtained thereof matched closely with the optimized solution predicted by the hybrid strategy. The MLPNN-AIS hybrid optimization strategy presented in this chapter can be gainfully utilized for modeling and optimization of similar waste-water treatment reactions/processes.

Chapter 8 presents a study, wherein the CI-based *fuzzy c-means clustering* (FCC) algorithm has been successfully used for fault detection of a biochemical process. Specifically, occurrences of various types of faults in a controlled continuous bioreactor generating biomass were accurately detected by the FCC methodology. Specifically, four types of faults in the bioreactor were simulated, namely increase/decrease in the bioreactor feed and jacket fluid flow rates. The resultant data consisting of bioreactor operating conditions and the related state variables, namely the bioreactor biomass concentration, bioreactor temperature and the jacket fluid temperature, were used to develop a model using *fuzzy c-means clustering* (FCC) algorithm. This method uses an unsupervised learning algorithm and depending upon the type of the occurred fault, groups the corresponding data into one of four possible clusters (one each for four types of faults). The results of the application of the developed FCC model clearly indicate that it is capable of accurately identifying all four types of the faults as also the normal process behavior.

Finally Chapter 9 concludes and summarizes the research studies conducted and presented in the thesis and provides suggestions for the future work.

Notable Features of Studies Presented in the Thesis

- Computational intelligence-based methodologies, namely ANN, GP and SVR have been comprehensively applied for modeling of important fuel

properties. Specifically, high performing CI-based models were designed and developed for predicting the higher heating value (HHV) and elemental composition of a wide variety of biomass fuels. Also, GP-based highly accurate models have been developed for predicting the HHV of coals of different ranks and from diverse geographies.

- Genetic programming based soft sensors were designed and developed for process identification of styrene polymerization process, and the resulting best model was successfully implemented in a model predictive control (MPC) scheme for controlling the polymer quality.
- Computational intelligence-based methodologies have been introduced for modeling of a fluidized bed co-gasification (COG) pilot plant. The COG process uses blends of high ash coals and three types of biomasses from India. The CI-based models that predict four COG performance variables possess excellent prediction and generalization performance.
- A hybrid CI-based strategy, integrating artificial neural networks and an artificial immune system (AIS) has been introduced for modeling-optimization of the reaction conditions of a synthesis reaction involving polymer resin based adsorptive removal of hexavalent chromium ions from waste water.
- A novel CI-based clustering methodology, namely *fuzzy c-means clustering* (FCC) has been successfully applied for fault detection of a biochemical process.

CHAPTER 1.

INTRODUCTION TO COMPUTATIONAL INTELLIGENCE BASED METHODOLOGIES

INTRODUCTION TO COMPUTATIONAL INTELLIGENCE BASED METHODOLOGIES

ABSTRACT

Due to factors such as global competition, strict environmental protection laws, and customer-demand for high and consistent product quality, it has become a must for process engineers to design and operate chemical processes in an efficient, safe, environment-friendly, and optimal manner. Modeling, simulation, and optimization of chemical reactions/reactors/processes assist the process engineers in this difficult task. Conventional modeling and optimization approaches have several limitations, which can be overcome by utilizing the contemporary *computational intelligence* (CI) based approaches for the mentioned process engineering tasks. Accordingly, in the present thesis several CI-based methodologies such as *artificial neural networks* (ANNs), *fuzzy logic* (FL), *genetic programming* (GP), *support vector regression* (SVR), *genetic algorithms* (GA) and *artificial immune system* (AIS) have been successfully employed for developing various chemical engineering/technology applications. This chapter gives a brief overview of computational intelligence along with a detailed discussion of the above mentioned CI-based formalisms thus forming a conceptual framework for the subsequent chapters. Additionally, the dimensionality reduction method namely *principal component analysis* (PCA), which has been often used in the CI-based modeling studies reported in this thesis, is discussed. Finally, a statistical method for identification of the best model from a set of the competing models is discussed.

1.0. COMPUTATIONAL INTELLIGENCE (CI) AND MACHINE LEARNING

Computational intelligence (CI) is a branch of computer science and comprises a collection of mathematical algorithms based on various kinds of intelligent behaviour exhibited by biologically evolving species (nature-inspired); it aims at providing novel and efficient solutions to complex real-life problems in science, engineering, technology, and even finance and economics. The most common intelligent behaviour exhibited by biologically evolving species involves learning, perception, logical reasoning, and decision making in uncertain and complex environments (Engelbrecht, 2007).

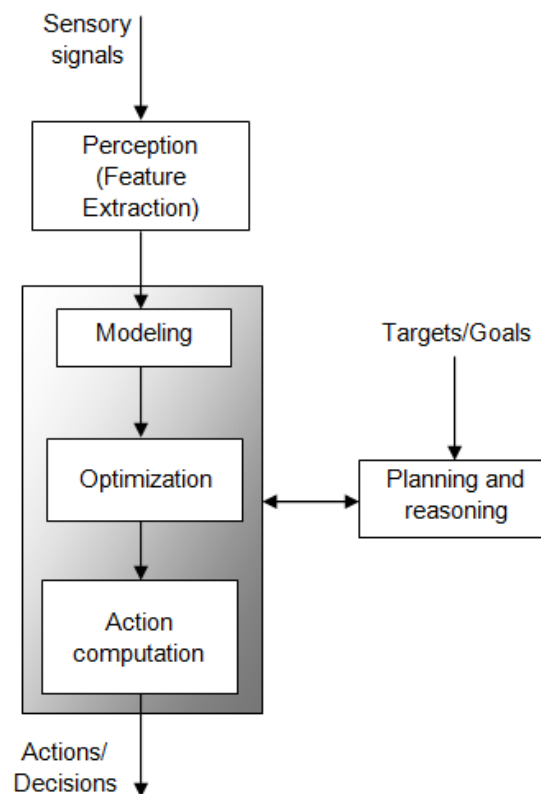


Figure 1.1: A CI-based system imitating biological intelligence in supervised learning mode.

Figure (1.1) shows a general architecture and working of a typical CI-based system (commonly known as “intelligent agent”) imitating biological intelligence. The intelligent agent perceives and models its environment reasonably well even in the presence of uncertainties and computes appropriate

(optimal) actions by anticipating their effects. For doing this, the targets/goals acting as set-points towards which the system is working for, are compared with the outputs given by the intelligent agent. This is known as supervised learning in the CI paradigm. The CI-based formalisms also encompass a range of machine learning methodologies based on the intelligence exhibited through other mechanisms such as statistical learning.

Around 1940s Alan Turing first devised the theory of computation suggesting a machine (now known as *computer*) that exhibits logical operations (Berlinski, 2000). The concurrent advancements in neurology and information theory made researchers to think on the possibility of the concept of an electronic brain (McCorduck, 2004). So the first work that came up and was recognized as exhibiting intelligence was the McCulloch's formal design for Turing-complete-artificial neurons in the year 1943. *Artificial intelligence* was a term coined by John McCarty at the Dartmouth conference in 1956 (Russell and Norvig, 2009), where the first AI program was presented. Although, the notion of *computational intelligence* was introduced by the IEEE Neural Networks Council in 1990, Bezdek (1994) (Siddique and Adeli, 2013), gave the formal definition of Computational Intelligence and defined CI as a subset of AI. Prior to 1970's the AI/CI research was slow. However, thereafter due to the tremendous advancements in computing power and availability of large datasets, the CI research gained momentum by way of advancements in CI-based generalized methods and those for solving specific problems. Over the years, CI-based methodologies have become an essential part of the engineering and technology industry, helping to solve many challenging problems in the field of engineering and other domains (Russell and Norvig, 2009).

The computational intelligence based methodologies address complex real-world problems, where traditional approaches may become useless for either or all of the following reasons for a system/process under study (Siddique and Adeli, 2013).

- The system/process might be too complex for traditional mathematical reasoning or the traditional approaches may prove to be difficult and time consuming to model the system/process.

- The system/process might contain various uncertainties.
- The system/process might be stochastic in nature.

A wide variety of engineering problems face such difficulties, wherein the CI based approaches can deliver speedy and efficient solutions since (i) they can model any nonlinear complex process with reasonable accuracies, (ii) being robust they can handle the system/process uncertainties (noise) robustly, and (iii) being nature-inspired these have some portion of stochastic behavior in-built which enables them to efficiently handle systems possessing stochastic nature.

The CI-based paradigms used in this thesis are:

- *Artificial neural networks* (ANN)
- *Evolutionary computation* (EC)
- *Artificial immune systems* (AIS)
- *Fuzzy logic* (FL)

There exists a much broader range of methodologies under the CI umbrella, and encompass theories based on *statistical learning* (support vector machines (SVM)), *probabilistic reasoning* (Bayesian networks) and intelligent agents (Engelbrecht, 2007) sometimes also termed as machine learning theories. Each of these CI-based formalisms used in the studies reported in this thesis is described in sufficient detail in the forthcoming paragraphs.

(A) Artificial Neural Networks (ANNs)

Artificial neural networks (ANNs) are highly parallel information-processing formalisms based on the mechanisms mimicking the working of the highly interconnected cellular structure of human brain (Freeman and Skapura, 1991; Zurada, 1992; Tambe et al., 1996). ANNs, with their diverse architectures form the most widely utilized CI-based methodologies for various engineering tasks such as modeling of steady state and dynamical systems, classification, clustering, feature extraction, noise reduction, nonlinear principal component analysis, and pattern recognition. The most popular ANN architecture, namely *multi-layer perceptron* (MLP) used in nonlinear function approximation utilizes a generic nonlinear transfer function in its building block and thus the troublesome

task of pre-specifying the form of the fitting function gets eliminated (Tambe et al., 1996). In this thesis we have utilized the multi-layer perceptron neural network (MLPNN) exhaustively for modeling of chemical engineering and technology systems. Accordingly, MLPNN is described in sufficient depth in the forthcoming section 1.2.1 (A).

(B) Evolutionary Computation (EC)

The evolutionary computational methodologies are stochastic iterative algorithms mimicking processes from natural evolution, wherein the main objective is based on the Darwinian's theory of the "*survival of the fittest*". In natural evolution, survival of a species occurs through reproduction. Offspring, containing the genetic material of the parents survives only if it can successfully face the challenges posed by the environment and produces a new generation of offspring. This way nature creates generations that are fitter than their parent generation. EC algorithms employ these principles of evolution. Generally they begin with a randomly created population of individuals (referred to as *chromosomes*) and iteratively evolve to the final generation composed of the fittest individuals by performing stochastic genetic operations on the individuals such as fitness evaluation, selection, reproduction and mutation. Several classes of EC algorithms are in use:

- Genetic algorithms (GA) (Holland, 1975): This is a popular stochastic optimization technique.
- Genetic programming (GP) (Koza, 1992): It performs symbolic regression whereby the form and parameters of a data-fitting function are automatically searched and optimized.
- Differential evolution (DE) (Storn and Price, 1997): DE is similar to GA with a difference in the reproduction mechanism.
- Cultural evolution (CE) (Campbell, 1960): CE models the evolution of culture of a population of individuals.

EC methodologies have been used (with GA being used most commonly used) successfully in diverse engineering applications, for example, numerical and combinatorial optimization, scheduling and planning, and in fault detection and diagnosis. In this thesis work, we have extensively used the genetic programming

method for modeling of chemical engineering and technology systems. In spite of its several attractive features, GP which has been elaborated in Section (1.2.2) has rarely been used in chemical engineering/technology applications. In this thesis, GA has been applied for the optimization of chemical engineering system and the detailed working of GA is given in Section (1.4.1).

(C) Artificial Immune Systems (AIS)

Artificial immune systems (AIS) is a relatively new CI-based paradigm inspired from various mechanisms exhibited by the *biological immune system* (BIS) of the vertebrates to counter with the foreign pathogen cells entering the body (referred to as antigen) (Farmer et al., 1986). BIS creates antibody cells specific to the antigen to counter the antigen and, thus, the infection as a whole. BIS exhibits several attractive features such as parallel processing, noise tolerance, robustness, diversity, memory, multi-layered architecture and adaptive nature. As the BIS encounters an antigen, its adaptive nature starts creating antigen-specific antibodies. The antibodies evolve into better antigen-specific antibodies through a parallel multilayered mechanism. Once the antigen attack is successfully counteracted upon, few of the antibodies remain in the system as memory cells for faster future response of the same or a similar antigen. The *clonal selection algorithm* (CLONALG) of the AIS paradigm has been utilized in the thesis work for optimization of a reaction for maximizing the adsorptive removal of chromium ions using synthetic polymer resins. The CLONALG algorithm is described in sufficient details in the Section (1.4.2 (A)).

(D) Fuzzy Logic (FL)

Fuzzy Logic (FL) is a systematic mathematical formulation based on the human reasoning (intelligence) for solutions of problems involving uncertainties (Zadeh, 1965). Human reasoning is almost always, not exact and the decisions taken by us under uncertain conditions are based on weighting—to a certain extent—each option available for decision making. This is mimicked by FL through a fuzzy set theory as opposed to the Boolean logic (binary-valued logic). A Boolean logic set, although precise, comprises of only two values, represented by 0 and 1 and it does not adequately represent an imprecise information. The

fuzzy set theory uses membership functions having values between 0 and 1 to represent weighing of an option under uncertainties. The degree of membership in the fuzzy set theory allows an object (option) in the fuzzy set to be anywhere in the range of 0 (completely not in the set) to 1 (completely in the set), thus permitting to deal with uncertainties in a natural way (Bose, 1994). The values of fuzzy variables are expressed in terms of linguistic objects, e.g. based on the height, a person may be graded as *short*, *medium* or *tall*, with each of them defined by a suitable membership function. The common membership functions are, *triangular*, *trapezoidal* and *Gaussian*, which allow for gradual variations in the variables as opposed to the abrupt changes in the case of Boolean logic. For example, the temperature control of a domestic heater will have the following fuzzy set,

$$T = \{VC, C, W, VW, H\} \quad (1.1)$$

where, T = temperature (variable to be fuzzified), and the linguistic terms applicable to the set are, VC = very cold, C = cold, W = warm, VW = very warm, H = hot. Membership function $\mu(T)$ quantify the input variables in the range 0 to 1 to represent them in a fuzzy set graphically. Figure (1.2) shows the possible fuzzy membership function distribution of temperature using the triangular membership function. The variable to be fuzzified on the x-axis is called the universe of discourse (in the FL terminology) and y-axis depicts the degree of membership, one for each linguistic term defined for the temperature variable.

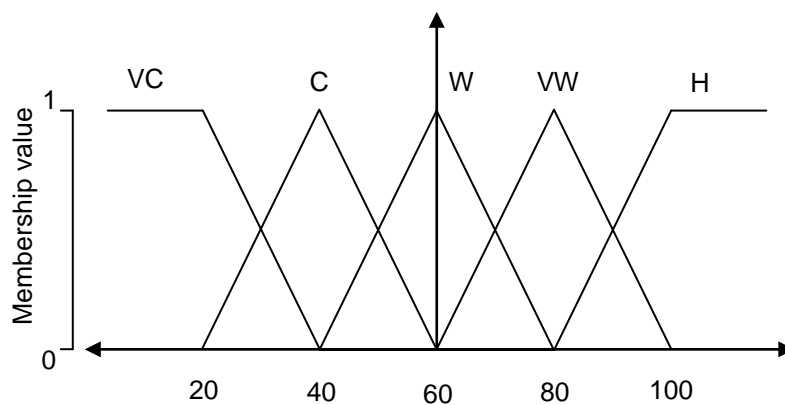


Figure 1.2: Fuzzification of temperature variable using triangular membership function.

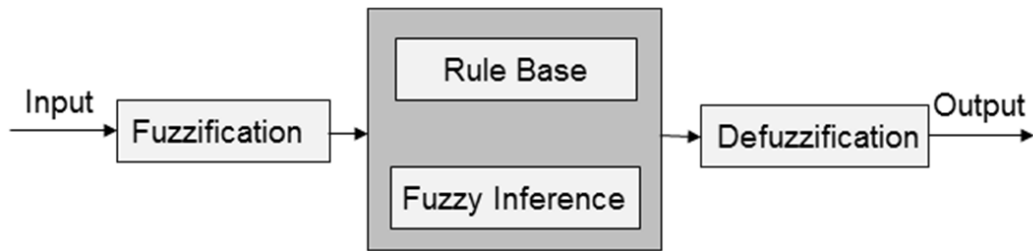


Figure 1.3: Information flow in a fuzzy logic system.

Figure (1.3) shows the working of a typical fuzzy logic system, wherein the task of FL is to map an input space to an output space, with the mechanisms of fuzzification, rule-base scrutiny, fuzzy inference and defuzzification. The fuzzification module transforms the inputs, which are crisp numbers, into fuzzy sets using relative measures of the variable to be fuzzified. The rule-base is a set of if-then rules provided by expert knowledge about the system. The fuzzified inputs are scrutinized by these rules and the results are inferred by the fuzzy inference engine. Finally the defuzzification module transforms the resulting fuzzy set obtained from the inference engine into a crisp value as required by the system output.

Fuzzy logic is most suitable to model real life problems where the process knowledge/data is imprecise and noisy. FL provides an easy and faster way to arrive at a definite conclusion based upon vague, imprecise, noisy or incomplete input information.

Following are the attractive features of fuzzy logic:

- FL is conceptually easy to understand and apply as it is based on human reasoning.
- FL is tolerant of imprecise and noisy data.
- FL is able to model nonlinear functions of arbitrary complexity.
- FL can be built using the experience of experts.
- FL can be applied in combination with traditional control techniques.

Due to these multiple attractive features, FL has been increasingly applied in various fields of engineering for tasks such as modeling, fault detection and

diagnosis and designing robust control systems and the table below, lists the most recent chemical engineering/technology applications of fuzzy logic.

Table 1.1: Recent chemical engineering/technology applications of fuzzy logic

Sr. No.	Application area	Specific study	Reference
1.	Process modeling	Process identification of reactive distillation by fuzzy-logic modeling for acetic acid recovery	Araromi et al. (2014)
		Neuro-fuzzy modeling of microbial fuel cell for the investigation of power density and columbic efficiency	Esfandyari et al. (2016)
		Predictive model for biomass pyrolysis based on fuzzy logic	Lerkkasemsa n (2017)
2.	Process fault detection/diagnosis	Fault tree analysis study of a distillation tower unit in a oil refinery process using fuzzy logic	Omidvari et al. (2014)
		Investigation of fault diagnostic capabilities of fuzzy clustering of the Tennessee Eastman process	Tóth and Hangos (2016)
		Fuzzy logic method used for the estimation of corrosion failure probability of oil and gas pipeline	Zhou et al. (2016)
		Failure mode and effect analysis of supercritical water gasification system for sewage sludge treatment using fuzzy logic method	Adar et al. (2017)
		Fuzzy logic used for fault diagnosis of a proton exchange membrane fuel cell	Benmouna et al. (2017)
3.	Process control	Distillation column control design using fuzzy logic method	Vasičkaninová et al. (2016)
		Design of fuzzy logic controller for freeze drying system	Fissore (2016)
4.	Process monitoring	Process monitoring for the improvement of data reliability using fuzzy logic method for outlier detection	Tanatavikorn and Yamashita (2015)

The *fuzzy c-means clustering* (FCC) algorithm, based on the concepts of FL has been used in this thesis for fault detection and diagnosis application of a continuous bioreactor generating biomass.

Today's technology is competent enough to collect and store large amounts of data, from peta-bytes to zetta-bytes; however today's (or conventional) analytical approaches have limitations of data-size. With the generation of an astounding amount of data in industries, customer services, R&D institutes, and government departments and agencies, the main problem is making sense of these data and using it for prediction. This is in the realm of "big data," and new and novel approaches are being devised to compress and analyze (extract important knowledge from) the monitored and archived huge data. Newer CI-based techniques such as *deep learning neural networks* are proposed to classify/model big data emanating from images and speech. Such upcoming applications range from healthcare to astronomical surveys. As an example in an upcoming application of astronomical surveys, wherein the "Large Synoptic Survey Robotic Telescope" will be operational in Chile from 2022 and will scan the entire southern hemisphere sky every three days, taking images of 50 billion objects, generating 2 tera-bytes of data per hour, and in-effect accumulating a 150 peta-byte imaging dataset in the next 10 years, the newer CI-based approaches will prove useful for feature extraction, feature selection, clustering, visualization, and classification for discovering important knowledge from the huge data. Another newer arena requiring intelligent tools is the Internet of Things (IoT), a network of physical objects (intelligent devices and smart objects) provided with sensors and actuators wherein, again the CI-based methods will prove their potential (Estevez, 2016).

While the individual CI-based methodologies have been successfully applied to solve real-world problems, there also exist numerous cases of successful solutions obtained by applying hybrid CI-based techniques. Using the hybrid approach we gain the advantages of multiple CI systems.

In commercial applications, the CI-based algorithms are now commonly used in video games, character recognition, image recognition, speech recognition, and "data-mining." Several leading software/information technology

such as Google, Facebook and Microsoft, are investing heavily in developing AI/CI based technologies. Although it's hard to predict the exact path that AI/CI will take in the future, however as seen from the increasing number of AI/CI based startups and the increasing amounts of funding for such firms it is clear that AI/CI will lead the technology industry of the future (Fortune, 2017). Following are the few commercial areas in which the AI/CI methods will dominate and make their mark in the future.

- Business Intelligence and Analytics
- Security: Cyber security and biometrics
- Automobile Technology: Autonomous driving automobiles
- Internet of Things
- Health care industry

An increasing use of AI/CI methodologies is also seen in chemical engineering and technology practice, which mainly include tasks such as process modeling and optimization, process fault detection and diagnosis, nonlinear process identification and control, QSPR (Quantitative Structure-Property Relationships) and QSAR (Quantitative Structure-Activity Relationships). The CI-based methods offer several attractive features over the conventional methods, while performing these tasks. The forthcoming sections describe these tasks in context with the conventional and CI-based methodologies applied for such process tasks.

1.1. PROCESS MODELING

Modeling is a process of establishing relations between the set of system variables to describe the behavior of the system. It essentially consists of development of mathematical relation(s) between the dependent (also termed “output” or “response”) and independent (also termed “input” or “predictor”) variables in a manner such that these are capable of predicting accurately the value(s) of the former. Chemical process models are of two types: *steady-state* and *dynamic* models. The former describe the time-invariant behavior of a chemical reaction/reactor/process. Steady state process models are important during the process design and operation phases of the plant as they form the basis for the normal operation of the plant. On the other hand, dynamic models describe

the time-dependent behavior of the process and are more complex than the steady-state models. These models are used in the control system design for analyzing the process response in case of disturbances, set-point changes, etc. as also for tuning the controllers. Commonly, reaction/reactor/process models are constructed using *phenomenological* and *empirical* approaches.

(A) Phenomenological Modeling Approach

The *phenomenological* modeling (also termed “first principles” modeling) modeling rigorously describes the physico-chemical mechanisms underlying a reaction/process and thus requires a complete understanding of the heat and mass transport phenomena, kinetics, and thermodynamics thereof. First principles models have several advantages such as,

- They provide an insight into the system/process behavior.
- They are capable of both interpolation and extrapolation of the system/process behavior.
- They are useful in scale-up of processes/equipment.

Acquiring the required information for developing these phenomenological models is in general time-consuming, costly and tedious due to the following reasons,

- Most real-life chemical systems/processes are inherently nonlinear and complex and thus difficult to model phenomenologically.
- Often, multiple nonlinear interactive dependencies exist between system/process variables and parameters in a chemical process.
- Modern day processes are complex and, therefore, there exists insufficient knowledge of the underlying physicochemical phenomena (e.g., heat and mass transfer mechanisms, reaction mechanisms and kinetics and the thermodynamics) thus requiring extensive efforts to arrive at a reasonable model.

(B) Empirical Modeling Approach

The empirical approach to process modeling entails development of single or multi-variable linear/nonlinear regression models. In here, a random or an informed guess is made regarding the structure of the data-fitting function and its

parameters are determined using process data and a suitable linear/nonlinear parameter estimation method. This approach is relatively less tedious than the phenomenological modeling although guessing an appropriate fitting function and subsequent parameter estimation does involve a trial and error procedure. Among these, the first task of specifying the fitting function requires multiple trials especially for nonlinear systems since a large number of fitting functions compete for fitting the data. Also, the domain over which an empirical model is valid may be limited since it is determined by the ranges of the data used in building the model. The other disadvantages of empirical modeling are that a statistically well distributed data is required to develop a good empirical model of the process, and empirical models are in general poor at extrapolation. The principal advantage of empirical modeling is that a detailed knowledge of the underlying physicochemical phenomena of the process is not required.

The above-stated limitations of phenomenological and empirical modeling approaches created a need for a paradigm shift in the approach towards modeling of chemical processes. Specifically, an approach, with following attributes is well-suited when phenomenological and empirical modeling are infeasible.

- A method that doesn't require an in-depth knowledge of the underlying physicochemical phenomena of a process (i.e., it should be data-driven).
- The modeling formalism doesn't require pre-specifying the form of the data-fitting function.
- The method should be capable of developing models possessing good prediction accuracy and equally important an excellent generalization capability.

In recent years, *Computational Intelligence* (CI) based approaches possessing above-stated attributes have offered a number of attractive avenues for modeling and optimization of chemical systems/processes. The studies presented in this thesis have employed these CI-based approaches extensively.

1.2. COMPUTATIONAL INTELLIGENCE BASED MODELING METHODOLOGIES

The CI-based methodologies have found a large number of modeling applications in every branch of science and engineering. The most commonly used CI-based modeling methodologies are *artificial neural networks* (ANNs), *fuzzy logic* (FL), and *support vector regression* (SVR), while *genetic programming* (GP) despite possessing some attractive properties has been rarely used method for modeling of chemical systems/processes.

1.2.1. Artificial Neural Network (ANN)

Artificial neural network (ANN) is a information-processing formalism based on the mechanisms followed by highly interconnected networks of the cellular structure of a human brain. The billions of cells forming the network in the human brain are known as *neurons*. Figure (1.4) shows the basic structure of a biological neuron. It is well known that the complex network of the neurons gives humans their intelligent characteristics. Individually, a neuron performs simple signal processing, however the neural network formed by billions of neurons possesses immense information processing capability which forms the basis of human intelligence. This intelligence imparting ability of the neural network is utilized in an artificial neural network, which is basically a computational algorithm mimicking the behavior of a biological neural network for accomplishing various tasks such as classification, regression, clustering, pattern recognition, memory associations etc.

The artificial neuron works in an analogues way as does a biological neuron. As shown in Figure (1.4), signals coming in through various inlet connectors known as *dendrites*, are summed in the central nucleus of the cell. The summed signal is further propagated through the *axon* connectors to downstream neurons. The axons of a neuron are connected to downstream neurons through a separating nerve cell gap called the *synapse*. A synapse acts as an activation function for the electrochemical signals transferred. Information processing/transfer occurs in the network through electrochemical signals. The signal is then received by the connected downstream neuron and processed in a

similar fashion. These numerous neurons working in conglomeration in the network finally generate the processed signal at the output nodes of the network.

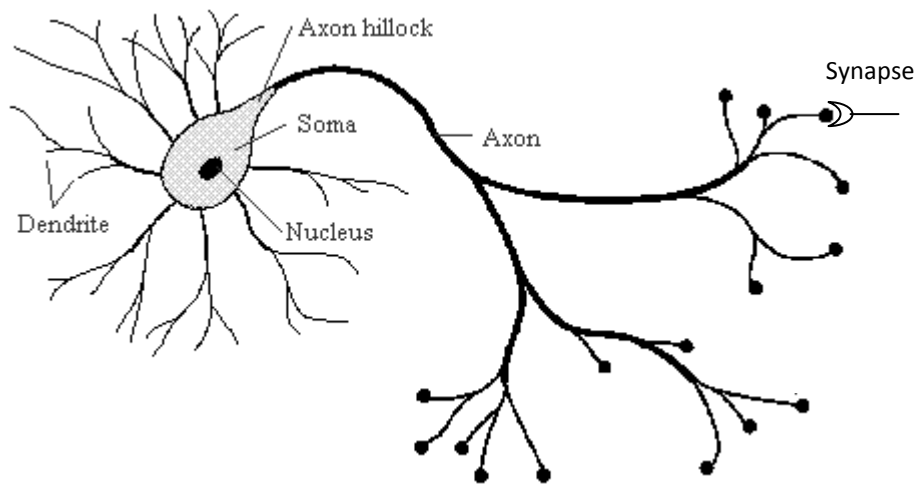


Figure 1.4: A biological neuron.

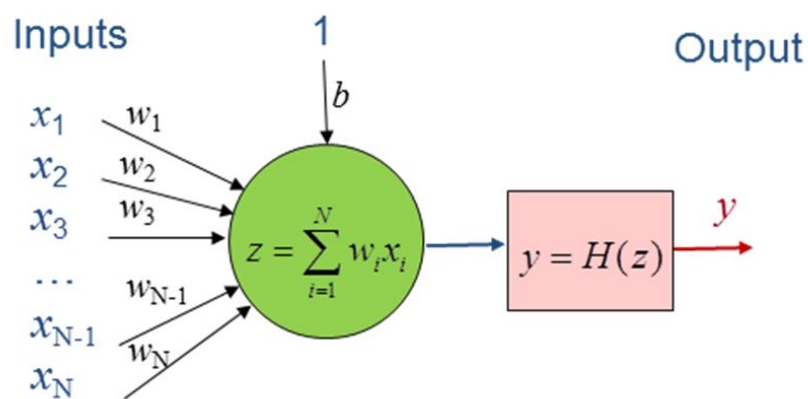


Figure 1.5: A computational neuron.

The information flow (via signals) in a computational (artificial) neuron, commonly known as a ‘node’, is shown in Figure (1.5). As represented in the figure, if an N-dimensional input vector, $\mathbf{x} = [x_1, x_2, x_3, \dots, x_{N-1}, x_N]^T$, is presented to the neuron, each of the input x_i is weighed by a weight w_i and is passed to the summing junction. A unity input with a weight ‘b’, known as bias is also fed to the summing junction. Here, all the weighed inputs are summed according to Eqn. (1.2) and the resulting output ‘z’ is transformed by an appropriate transfer function (known as an activation function in the ANN terminology) to produce the desired output (target) ‘y’.

$$z = \sum_{i=1}^N w_i x_i + b \quad (1.2)$$

It is the use of nonlinear transfer functions that impart ANNs their ability of nonlinear classification/regression. The commonly used transfer functions are:

- i. Logistic Sigmoid Function:

$$f(z) = \frac{1}{1 + \exp(-z)} \quad (1.3)$$

- ii. Hyperboloid Tangent Function:

$$f(z) = \frac{\exp(z) - \exp(-z)}{\exp(z) + \exp(-z)} \quad (1.4)$$

- iii. Linear Function: $f(z) = z$ (1.5)

- iv. Gaussian Function: $f(z) = \exp(-z^2)$ (1.6)

This nonlinear transformation and multivariable problem handling capabilities of ANNs has been extensively utilized by various researchers over the years by modifying the basic architecture of the ANNs for solution of specific problems under consideration. Thus, there exist various types of ANNs suitable for various tasks such as regression, classification, clustering, pattern recognition etc. Table (1.2) lists various types of ANNs along with their information flow and learning modes (Tambe et al., 1996).

Table 1.2: Classification of commonly used ANN architectures

Feed-forward information flow		Feed-backward information flow	
Learning Approach			
<i>Supervised</i>	<i>Unsupervised</i>	<i>Supervised</i>	<i>Unsupervised</i>
Perceptron	Kohonen Self-organizing Feature Map (SOFM)	Fuzzy Cognitive Map (FCM)	Discrete Bidirectional Associative Memory (BAM)
Multilayer Perceptron (MLP)	Linear Associative Memory (LAM)	Brain-State in-a-Box (BSB)	Additive Grossberg (AG)
Counter-	Learning Vector		Analog Adaptive

Feed-forward information flow		Feed-backward information flow	
Learning Approach			
<i>Supervised</i>	<i>Unsupervised</i>	<i>Supervised</i>	<i>Unsupervised</i>
propagation (CP)	Quantizer (LVQ)		Resonance Theory (ART2)
Radial Basis Function1 (RBF)			Discrete Hopfield (DH)
Boltzmann Machine (BM)			Continuous Hopfield (CH)
Radial Basis Function1 (RBF)			
Cauchy Machine (CM)			

(A) Multi-Layer Perceptron Neural Network (MLPNN)

The most widely used ANN architecture for steady-state process modeling is the fully interconnected feed-forward *multi-layer perceptron* (MLP) hereafter termed as the multi-layer perceptron neural network (MLPNN) architecture. A MLPNN has proven capabilities for nonlinear modeling of a full multiple input-multiple output (MIMO) system (Tambe et al., 1996). Given an example MIMO data set consisting of independent (input), and the corresponding dependent (output) variables of a system/process, an MLPNN has the ability of learning and generalizing the nonlinear relations that exist between the dependent and the independent variables. The attractive features of MLPNN based modeling are (Tambe et al., 1996):

- MLPNNs can approximate any complex, nonlinear MIMO relations in a given process/system dataset by performing supervised regression.
- MLPNNs use a generic nonlinear function for regression/classification and thus it is not necessary to pre-specify (as required to be pre-specified

for empirical modeling) the exact form of the data-fitting function explicitly.

- A well-trained MLPNN process/system model possesses good prediction accuracy and generalization performance, due to which it performs well, while predicting a test example set that, does not belong to the training dataset.

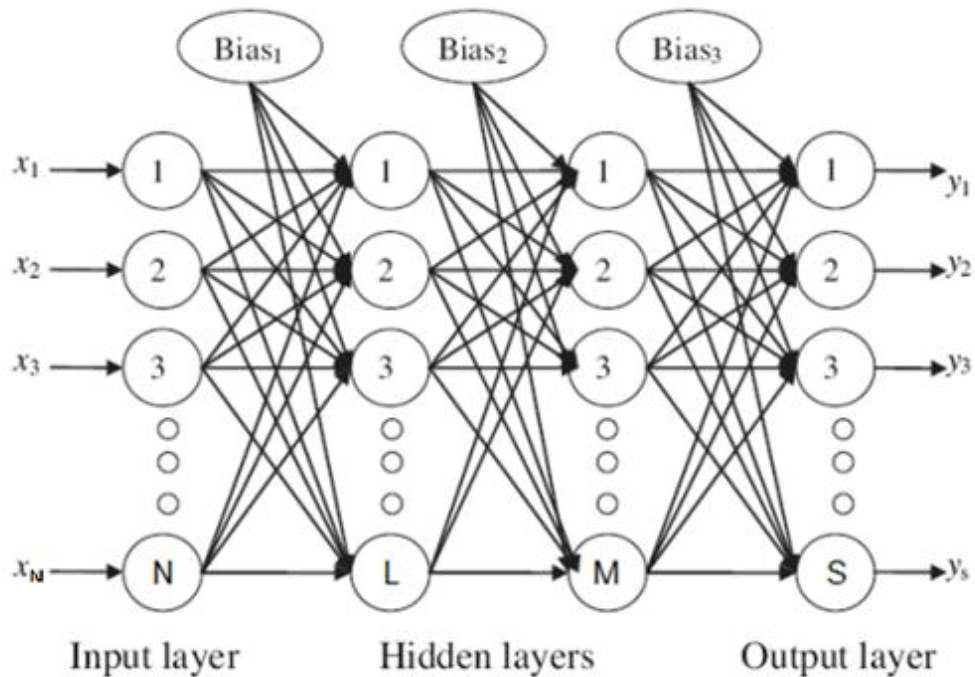


Figure 1.6: Architecture of a typical multi-layer perceptron neural network trained for a MIMO dataset.

The MLPNN's architecture consists of an input layer, one or more intermediate layers known as hidden layers and an output layer. The nodes in a layer are fully interconnected with nodes in the consecutive layers, as shown in Figure (1.6). The interconnections are weighed by weights $\{w_{ij}\}$, where 'i' represent the i^{th} node in previous layer and 'j' represents the j^{th} node in the current layer. Bias nodes are present (with a unity output) in its input and hidden layers. A bias node is connected to all the nodes in the subsequent layer through appropriate weights. The output of a MLPNN modeling a multiple input-single output (MISO) system is given by,

$$\hat{y} = f_o(z_o) \quad (1.7)$$

$$z_o = \sum_{i=1}^n w_{io}x_{io} + b_o \quad (1.8)$$

where, \hat{y} = predicted output, f_o = output transfer function, z_o = output of the summing junction, w_{io} = weight between the node 'i' in the hidden layer and the output node 'o', x_{io} = input from node 'i' to the output node 'o', b_o = bias weight for the output node 'o' and n = number of nodes in the hidden layer. Thus, we can express the output of a single hidden layer MLPNN as,

$$\hat{y} = f_o\left(\sum_{i=1}^n w_{io}x_{io} + b_o\right) = f_o\left(\sum_{i=1}^n w_{io}f_i(z_H) + b_o\right) = f_o\left(\sum_{i=1}^n w_{io}f_i\left(\sum_{j=1}^N w_{ij}x_{ij} + b_{Hj}\right) + b_o\right) \quad (1.9)$$

where, f_i = transfer function of the i^{th} node in the hidden layer, subscript 'H' indicates the hidden layer and N = number of input layer nodes (equal to the number of inputs).

The MLPNN consisting of two hidden layers for modeling a full MIMO system is shown in Figure (1.6). In general, the number of input layer nodes is equal to the number of independent variables (inputs: x_1, x_2, \dots, x_N) in the N -dimensional input space while the number of nodes in the output layer is equal to the number of 'S' dependent variables (outputs: y_1, y_2, \dots, y_S). The number of intermediate hidden layers and the number of nodes in each hidden layer are the parameters of the MLPNN architecture and these are determined heuristically based on the model's desired prediction accuracy and generalization performance.

Any ANN and so the MLPNN works in the following two modes:

1. *Training Mode*: In the first step, known as the training or learning mode of a MLPNN, the dataset (inputs and the corresponding outputs/targets) to be modeled is presented to the MLPNN in either batch mode or on-line mode. Training involves adjustment of the interconnecting weights in the MLPNN to drive the network towards desired regression abilities with good prediction accuracy and generalization capability. Typically in the on-line training mode a single data-point is presented to the MLPNN at a time; then the errors are calculated and weights are updated and this is done repeatedly for all the data-points till the termination criteria is met. In the batch mode all the data-points are presented sequentially to the MLPNN, an average error is calculated and weights are updated. This is

done repeatedly till the termination criteria is met. The popular training algorithm for a MLPNN is the error-back-propagation (EBP) algorithm as devised by Rumelhart et al. (1986).

2. *Prediction Mode*: After successful training of the MLPNN, the developed MLPNN model can then be used to predict the outputs for a given input set, which is not a part of the original training set. Thus now the developed MLPNN model can be used for prediction.

Error-back-propagation (EBP) Training

As mentioned previously the *error back propagation* (EBP) algorithm by Rumelhart et al. (1986), has been the most popular MLPNN training algorithm. Consider a given MISO example dataset,

$$\mathbf{D} = \{(\mathbf{x}_j, y_j)\} \quad (1.10)$$

consisting of P number of patterns ($j = 1, 2, \dots, P$) with each input pattern \mathbf{x}_j containing N inputs (x_1, x_2, \dots, x_N) and the corresponding single output (y_j). When an input pattern from this dataset is presented to the MLPNN, the EBP algorithm calculates the prediction error and minimizes it by adjusting the interconnecting weights of the network. A pseudo-code of the EBP algorithm for a single hidden layer MLPNN architecture is listed in Figure (1.7). The error minimization is done in two passes. Starting with randomly initialized weights, in the forward pass the MLPNN predicts the output \hat{y}_j for a presented input pattern ' \mathbf{x}_j '. This predicted output is then compared with the actual or target output to determine the prediction error as,

$$e_j = y_j - \hat{y}_j \quad (1.11)$$

In the backward pass this error is minimized in steps. In the first step the weights connecting the nodes in the output and the hidden (previous to output) layer are updated according to the EBP algorithm using the gradient-descent technique known as 'delta rule' as,

$$\Delta w_k = \eta \Delta w_{k-1} + (1 - \eta) \mu e_j x_j \quad (1.12)$$

where, Δw_k = weight change in current iteration 'k', Δw_{k-1} = weight change in the previous iteration, η = learning rate, μ = momentum coefficient, and x_j = input to the node 'j', e_j = error for the j^{th} node, x_j = input to the j^{th} node. Next the weights interconnecting the input and the hidden layer are updated using the same delta

rule (see Eqn. 1.12), using the error that is propagated back from the next layer. The *learning rate*, η ($0 < \eta < 1$) and the *momentum coefficient*, μ ($0 < \mu < 1$) are EBP specific parameters that are tuned heuristically to obtain the best performance. Once the whole dataset is presented to the network an average error metric such as the *root mean squared error (RMSE)* as given in Eqn. (1.13) is computed, which is minimized iteratively to determine the optimal network weights.

$$RMSE = \sqrt{\frac{\sum_{j=1}^P (y_j - \hat{y}_j)^2}{P}} \quad (1.13)$$

where, y_j denotes the desired (or target) values and \hat{y}_j refers to the MLPNN model predicted values corresponding to the j^{th} pattern. A pseudo-code of the EBP algorithm for the training of a single hidden layer MLPNN model is listed in Figure (1.7).

```

Initialize MLPNN weights randomly
do
  for: each training example
    Predicted output = MLPNN output in forward path
    Error = predicted output - actual output
    Compute all weights between the hidden and output layer
    according to Eqn. (1.12) in backward path
    Compute all weights between the input and hidden layer
    according to Eqn. (1.12) in backward path
  end for
until the stopping criterion satisfied
return final weights of the MLPNN

```

Figure 1.7: Pseudo-code of error-back-propagation (EBP) algorithm for a single hidden layer MLPNN.

ANN Training Issues

To develop an optimal MLPNN model, following factors are critical,

- Architectural factors: the number of hidden layers, the number of nodes in each hidden layer, and the transfer function at hidden layer nodes.
- EBP algorithm-specific factors: learning rate (η) and momentum coefficient (μ).

For design and development of an optimal MLPNN model the effect of these parameters must be rigorously investigated. This is commonly done using heuristics; the details of which can be found in, for example, Freeman and Skapura (1991), Zurada (1992), Bishop (1994) and Tambe et al. (1996).

During the training of an MLPNN, its, output prediction accuracy is monitored in terms of two statistical quantities, namely, *coefficient of correlation (CC)* and *root mean squared error (RMSE)* between the model predicted output and its target (desired) value from the training dataset; in general, as *CC* increases, *RMSE* decreases as the number of training cycles increases. Although, in general as training cycles increase the prediction accuracy of the model increases, above a certain number of cycles, the generalization ability of the model starts decreasing. Such a model is known to be *over-trained* or *over-fitted*. An over-trained ANN model captures the micro details of the dataset such as noise, at the expense of learning smooth trends in the dataset. Such an over-trained model is useless since it makes inaccurate predictions for a new set of inputs resulting in poor generalization. To avoid the problem of over-training of an ANN, the dataset is divided into *training* and *test* subsets. While the training set is used for training the network, the test set is used for validating the generalization ability of the network undergoing the training. Training is stopped at the point, when the network starts over-fitting of the dataset. A complete procedure for determining an optimal MLPNN architecture and the training (EBP) parameters is summarized for example in Bishop (1994); and Tambe et al. (1996). MLPNNs have been used successfully in several chemical engineering and technology applications and a partial list of such most recent applications is given in the following table.

Table 1.3: Recent chemical engineering/technology applications of multi-layered perceptron neural network

Sr. No.	Application area	Specific study	Reference
1.	Process Modeling	A modified multilayer perceptron model for gas mixture analysis	Moore et al. (1993)
		Modelling of an Industrial Fluid Catalytic Cracking Unit Using Neural Networks	Michalopoulos et al. (2001)
		Hybrid process modeling and	Nandi et al.

Sr. No.	Application area	Specific study	Reference
		optimization strategies integrating neural networks/support vector regression and genetic algorithms: study of benzene isopropylation on Hbeta catalyst	(2004)
		Artificial neural network based modeling of activated sludge process	Moral et al. (2008)
		Estimation of thermal conductivity of ionic liquids	Hezave et al. (2012)
		Experimental investigation, modeling and optimization of membrane separation using artificial neural network and multi-objective optimization using genetic algorithm	Soleimani et al. (2013)
		Prediction of carbon dioxide solubility in ionic liquids using MLP and radial basis function (RBF) neural networks	Tatar et al. (2016)
		Correlating thermal conductivity of pure hydrocarbons and aromatics via perceptron artificial neural network (PANN) method	Lashkarbолоki et al. (2017)
2.	Process fault detection and diagnosis	Optimum parameters for fault detection and diagnosis system of batch reaction using multiple neural networks	Tan et al. (2012)
		Designing supervised local neural network classifiers based on EM clustering for fault diagnosis of Tennessee Eastman process	Rad and Yazdanpanah (2015)
3.	Soft sensor development	Soft sensor for continuous product quality estimation (in crude distillation unit)	Rogina et al. (2011)
		Development of soft sensor with neural network and nonlinear variable selection for crude distillation unit process	Sun et al. (2016)
4.	Data classification	Designing supervised classifiers for multiphase flow data classification	Tarca et al. (2004)
		Comparison of Fisher's linear discriminant to multilayer perceptron networks in the classification of vapors using sensor array data	Pardo et al. (2006)
		Luciferase-based bioassay for rapid	Denisov

Sr. No.	Application area	Specific study	Reference
		pollutants detection and classification by means of multilayer artificial neural networks	(2017)
5.	Model based process control	Neural modelling, control and optimization of an industrial grinding process	Govindhasamy et al. (2005)
		System identification and model predictive control for a flotation column	Mohanty (2009)
6.	Quantitative Structure-Activity/Property Relationships (QSAR/QSPR)	Elicitation of the most important structural properties of ionic liquids affecting ecotoxicity in limnic green algae; a QSAR approach	Izadiyan et al. (2013)
		Development of QSAR model to predict the ecotoxicity of <i>Vibrio fischeri</i> using COSMO-RS descriptors	Ghanem et al. (2017)

The ANN algorithms are provided by various software packages, out of which some are open source packages. Following are the popular ANN software packages.

1. *RapidMiner* (Mierswa et al., 2006): It is a freeware from “Technical University of Dortmund”, consisting of various CI-based algorithms such as ANNs, SVR, GA, etc. for engineering and business applications involving machine learning, data mining, text mining and predictive analytics.
2. *IBM-SPSS®* (2011): IBM-SPSS Statistics is mainly a statistical analysis software, that also includes MLPNN based ANN learner with various training facilities (batch/online) and algorithms (conjugate gradient/Levenberg Marquardt).
3. *MATLAB Neural Network Toolbox™* (MATLAB, 2000): MATLAB is a commercial *mathematical/numerical* computational software, widely used in research and industries. Built on top of MATLAB are various toolboxes, specially developed for specific domains. The neural network toolbox™ is such a commercial toolbox providing algorithms for different ANN architectures for accomplishing tasks such as regression, classification,

clustering, dimensionality reduction, time-series forecasting, and modeling of dynamical systems.

1.2.2. Genetic Programming (GP)

Genetic programming (Koza, 1992) is an evolutionary, population-based methodology possessing stochastic search and optimization features. It is inspired from the Darwinian theory of biological evolution as an extension of genetic algorithm (GA) (Holland, 1975) methodology. It was originally proposed to write computer programs doing pre-specified tasks. Additionally, GP is also capable of symbolic regression. Given a data-driven modeling problem, GP-based symbolic regression (GPSR) can automatically search and optimize the form and parameters of an optimally fitting linear/nonlinear function. (see Figure (1.8)).

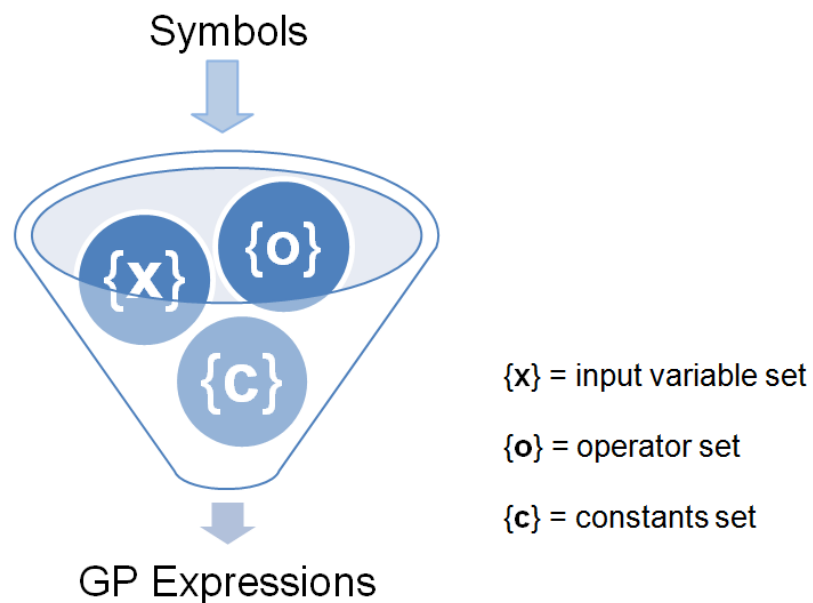


Figure 1.8: The symbolic regression concept of GP.

In the GPSR procedure, random expressions are generated initially and these are evolved over multiple generations according to the principles of genetic evolution. Although novel and potentially attractive, the GP methodology has been rarely used to solve chemical engineering problems as compared to the use of other CI-based methodologies.

GP Implementation

For a given multiple input-single output (MISO) example dataset (see Eqn. (1.10)) consisting of P number of patterns ($k = 1, 2, \dots, P$), where each input pattern \mathbf{x}_k contains N inputs (x_1, x_2, \dots, x_N) and the corresponding single output (y), the task of the GPSR is to determine the best combinations of the library symbols $\mathbf{l} = \{\mathbf{x}, \mathbf{c}, \mathbf{o}\}$, so as to obtain the best-fit function ' f ' for the given dataset,

$$y = f(x_1, x_2, \dots, x_N, c_1, c_2, \dots, c_R) \quad (1.14)$$

The library symbols consist of the following sets, $\mathbf{x} = [x_1, x_2, \dots, x_N]$ is the vector of independent variables, $\mathbf{c} = [c_1, c_2, \dots, c_R]^T$ is the vector of constants/parameters and $\mathbf{o} = [o_1, o_2, \dots, o_S]^T$ is the vector of operators, e.g. $\{+, -, *, /, ^, \text{exponentiation, logarithm, square root}\}$.

The objective function that GPSR minimizes for arriving at the best expression ' i ' is of a form,

$$F_i = RMSE_i + \alpha * n_{di} + \beta_i \quad (1.15)$$

where, $RMSE_i$ = root mean squared error for the candidate expression ' i ' evaluated over the entire dataset, α = parsimony pressure, n_{di} = number of nodes in the candidate expression ' i ', β_i = penalty coefficient for expressions without variables. To achieve the above-stated modeling task, the GPSR iterates over a population of candidate expressions to evolve into best-fitting expressions through a sequential process of genetic operations such as selection, reproduction by crossover and mutation, over the generations.

For an efficient operation, GPSR requires specific data structures to express the candidate expressions/models such as the tree-structured GP, linear GP, Cartesian GP and stack-based GP. The most popular data structure commonly employed by GPSR is the tree structure. It is a hierarchical structure with a root node and branches extending from the root, forming different sub-branches, joining the leaves (nodes) at different levels. A leaf (node) is either of the two types, namely "operator/function" and "terminal" node. The operator node consists of a mathematical operator from the library set ' \mathbf{o} ' as mentioned

previously. A terminal (operand) node consists of either a constant (c_i) from the constants library set ‘c’ or an input variable (x_i) from the input variable library set ‘x’. A tree is traversed from the left-bottom to right-top direction for a sub-branch, to get the represented expression, as shown in the basic tree structure in Figure (1.9). Figure (1.9) also shows the common genetic operations GPSR performs on such tree structures.

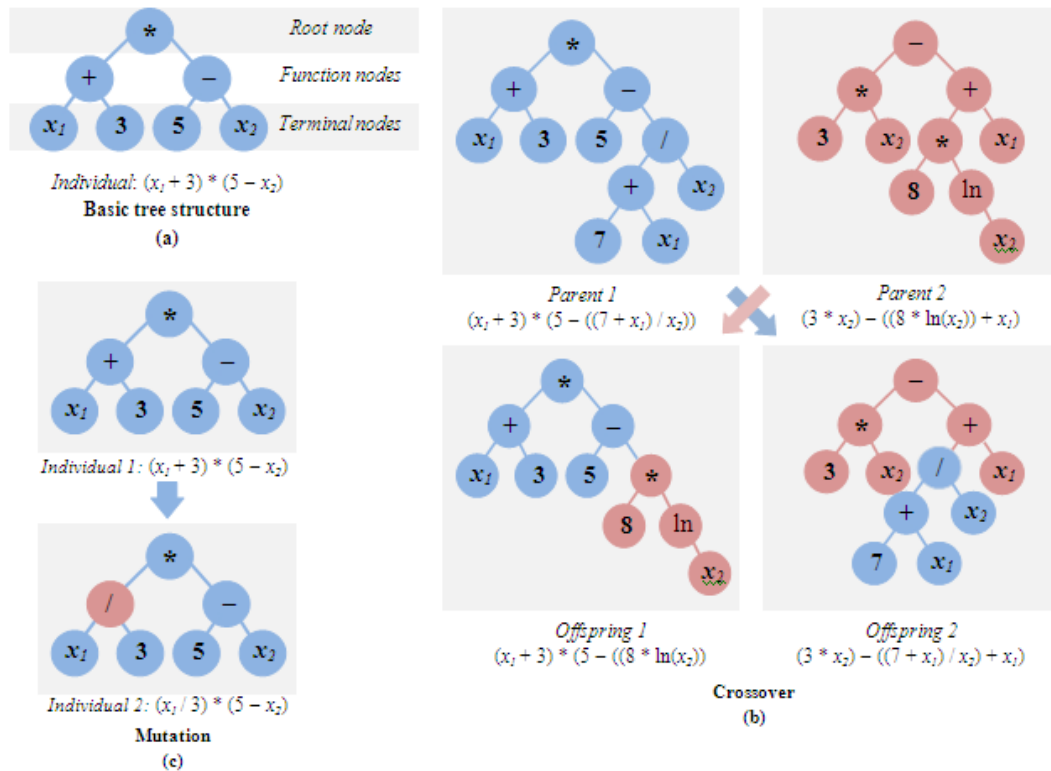


Figure 1.9: Schematic of genetic programming: (a) Basic tree structure, (b) Crossover operation, (c) Mutation operation.

The trees of variable depths allow construction of expressions of varying lengths and complexity. GPSR being a stochastic procedure, it is likely that repeated GPSR runs—for instance, using different random initializations—will lead to different solutions. So to arrive at an acceptable model a set of several GPSR runs is required. A GPSR pseudo-code is given in Figure (1.10) and the steps are explained next. The stepwise procedure for the implementation of a typical GPSR run is given below.

Initialization (Step 1): Initially the user has to specify the library sets of input variables, operators and constants for the GPSR to select randomly from the

specified set. An initial population of candidate trees (expressions) is generated randomly.

Fitness evaluation (Step 2): In this step the fitness of each candidate tree is determined, e.g. according to Eqn. (1.15), by determining the *RMSE* over the entire dataset. The fitness function assigns a higher fitness to expressions making better predictions.

Selection (Step 3): In this step a “mating” pool of candidate expressions is created by selecting the best candidate expressions, based on their fitness. The occupants of this mating pool forming the parents, take part in the most important mating operation by crossover (step 4), to produce the offspring candidate tree expressions. The selection step allows the best candidate expressions to form the mating pool to undergo crossover. Several selection procedures exist, such as the Roulette-wheel selection (De Jong, 1975), tournament selection (Goldberg, 1990), truncation selection (Crow and Kimura, 1979) as used in GA are also used for GPSR.

Crossover (Step 4): Here, two candidate expressions (parents) from the mating pool are randomly selected to undergo crossover. The crossover operation between a parent pair occurs by exchanging of sub-branches between the parent tree expressions to create off-spring expressions. Figure (1.9(b)) shows an illustration of such a crossover operation on the tree structures.

Mutation (Step 5): In this step, the population of the selected parent and generated offspring tree expressions are subjected to random mutation. In mutation, the content (symbol) of a node is replaced with other content from the same library set. The mutation is carried out randomly with a small probability to maintain diversity in the population and prevent the trees from getting stuck in a local minimum. Figure (1.9(c)), shows a mutation operation on a tree structure wherein a randomly selected ‘addition’ operator node is replaced randomly to a ‘division’ node. The mutated population representing a new generation is now ready for the next iteration cycle and thus the generation index is incremented.

Termination (Step 6): At this stage the termination criteria is checked to ascertain if the stopping criteria has met. The stopping criterion as specified by the user

could be either a specified number of generations or no significant change in the fitness value over the successive generations. If the set termination criteria is not met, steps 2 to 5 are repeated till the convergence is reached.

The whole GP based symbolic regression procedure is summarized in the form of a pseudo-code below.

```
Randomly initialize a population of N candidate expressions
do
  Select best n candidate expressions according to fitness
  criteria (see Eqn. (1.15)) to form the parents in the mating
  pool

  for i = 1 to (N-n)/2
    Select a parent pair probabilistically from the mating
    pool
    Perform reproduction by crossover to generate offspring
    pair
  end for

  Replace the lower-fit (N-n) candidate expressions with the
  offspring population
  Mutate the population probabilistically
  generation = generation + 1
until the stopping criteria is met
return best candidate solution from the last generation
```

Figure 1.10: Pseudo-code of genetic programming algorithm for symbolic regression (GPSR).

In this way, GPSR works on the given dataset without making any assumptions about the form and coefficients of the fitting-function. Despite of its several attractive features, GP has been rarely utilized in chemical engineering/technology applications. The following Table (1.4) lists a few recent GP applications in chemical engineering/technology domain. Further details of GPSR can be found in, e.g., Vyas et al. (2015).

Table 1.4: Recent chemical engineering/technology applications of genetic programming

Sr. No.	Application area	Specific study	Reference
1.	Process modeling	Prediction of the minimum spouting velocity in a conical base spouted bed	Hosseini et al. (2014)
		Genetic Programming method used for modeling and optimization of surfactant/polymer flooding	Bahrami et al. (2016)
		Modeling of supercritical CO ₂ thermal conductivity using genetic programming	Rostami et al. (2017a)
		Genetic programming based modeling of adsorption of phenols and nitrophenols by activated carbon	Emigdio et al. (2017)
		Prediction of hydrocarbon/water interfacial tension using genetic programming	Rostami et al. (2017b)
2.	Process fault detection/diagnosis	Fault diagnosis of centrifugal pump using genetic programming	Sakthivel et al. (2012)
		Fault propagation analysis of a simulated hydrogen production plant using genetic programming	Gabbar et al. (2014)
3.	Quantitative Structure-Activity/Property Relationships (QSAR/QSPR)	Quantitative structure–retention relations for the prediction of Kovats retention indices based on Genetic programming symbolic regression	Goel et al. (2015)
		Development of QSPR model for predicting solubility parameters of polymers using genetic programming	Koç and Koç (2015)
4.	Soft sensor development	Soft sensor based process identification of analyte-specificity using a non-specific whole-cell biosensor using genetic programming	Podola and Melkonian (2012)
		Development of soft sensor models for biochemical processes using genetic programming	Sharma and Tambe (2014)
5.	Process control	Genetic Programming as a tool for control system design	Balandina (2017)

A number of GP-based software tools are currently available; a few prominent ones are described below in brief.

1. *Eureqa Formulize* (Schmidt and Lipson, 2009): This commercial software package is optimized to provide “parsimonious” (low complexity) solutions to a given regression problem. It has the facilities to pre-specify a seed function, adjust the amount of split between the training and test datasets, pre-specify the mathematical/relational/logical operators and pre-specify the error metric to be minimized for the GP run.
2. *pySTEP* (Khoury, 2009): This is an open source Python based GP software package from Massachusetts Institute of Technology (MIT) and its abbreviation stands for ‘Python Strongly Typed gEnetic Programming’. It can easily evolve populations of tree structures with precise grammatical and structural constraints. To run the pySTEP application the user needs to pre-specify the population size, minimum tree depth, maximum tree depth, maximum number of generations, crossover probability, mutation probability and reproduction probability.

1.2.3. Support Vector Regression (SVR)

Support vector machine (SVM) is a statistical/machine learning methodology, originally devised for performing supervised linear classification (Vapnik, 1995; Burges, 1998). By introducing nonlinear feature or kernel functions in SVM, it is possible to perform nonlinear classification as well. Support vector regression (SVR) is an adaption of SVM that performs regression. To do this, SVR, first nonlinearly maps the inputs into a high dimensional “feature/kernel” space (Φ), wherein these are correlated linearly with the output. Support vector regression follows the structural risk minimization (SRM) principle, as opposed to the empirical risk minimization (ERM) commonly employed by the standard learning methods, such as MLP neural network. The SRM approach creates an optimized SVR model such that, both the prediction errors and model complexity are concurrently minimized. This is done by considering both the prediction error and model complexity in the loss function to be minimized. Owing to this attribute an SVR model possesses a higher potential to generalize the learned input-output relations.

To solve a nonlinear regression problem for a given dataset (see Eqn. (1.10)), the SVR methodology considers the following linear estimation function,

$$y = f(\mathbf{x}) = (\mathbf{w} \cdot \Phi(\mathbf{x})) + \mathbf{b} \quad (1.16)$$

where, \mathbf{w} denotes the weight (parameter) vector; \mathbf{x} is the input vector, \mathbf{b} is the bias (constant) vector; $\Phi(\mathbf{x})$ is the kernel function and $(\mathbf{w} \cdot \Phi(\mathbf{x}))$ describes the dot product in the feature/kernel space. Although, Eqn. (1.16), is linear, nonlinear regression is performed by first nonlinearly mapping the input data vector, \mathbf{x} into a high-dimensional feature space, Φ and performing linear regression in that space. Thus, the problem of nonlinear regression in the lower dimensional input-space is transformed into a linear regression problem in a higher dimensional feature/kernel space.

Following kernel functions are commonly used with SVM/SVR formalisms,

- i. Simple dot product kernel function:

$$\Phi(x_i, x_j) = (x_i \cdot x_j) \quad (1.17)$$

where, x_i and x_j are the variables on which the operation (here dot product) is performed.

- ii. Simple polynomial kernel function:

$$\Phi(x_i, x_j) = (1 + (x_i \cdot x_j))^d \quad (1.18)$$

where, $d = \text{degree of the polynomial}$.

- iii. Radial basis kernel function:

$$\Phi(x_i, x_j) = \exp(-\gamma |x_i - x_j|^2) \quad (1.19)$$

where, $\gamma = \text{kernel gamma}$.

- iv. Anova kernel function:

$$\Phi(x_i, x_j) = \sum_{k=1}^N \exp(-\sigma(x_i^k - x_j^k)^2)^d \quad (1.20)$$

where, σ = scale parameter, N = number of examples.

There are a number of loss functions available, such as the Laplacian, Huber's, Gaussian, and ε -insensitive for the SVR formulation. Among these, the most common is the robust ε -insensitive loss function (Vapnik, 1995) (see Figure (1.11)) given as,

$$L(f(x) - y) = \begin{cases} |f(x) - y| - \varepsilon & \text{for } |f(x) - y| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases} \quad (1.21)$$

where ε , commonly known as epsilon is a precision parameter representing the radius of the tube located around the nonlinear regression function, $f(x)$ (see Figure (1.11)).

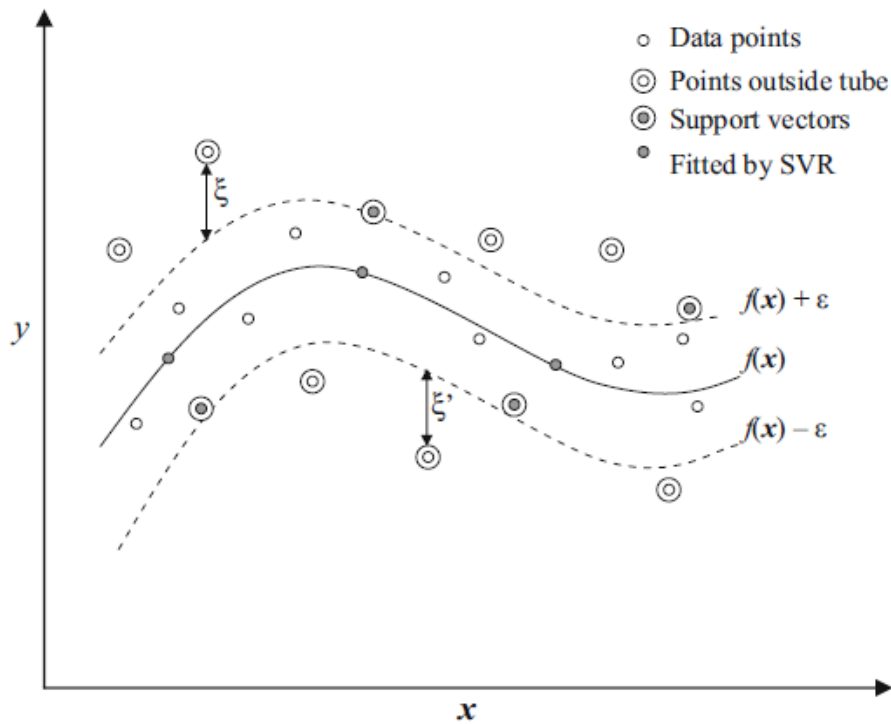


Figure 1.11: Schematic of SVR using the ε -insensitive loss function (ε : tube radius, ξ, ξ' : soft margins).

The region within this tube is known as 'ε-insensitive' zone, since the loss function 'L' is formulated such that it assumes a zero value in this zone and as a result it does not penalize the prediction error with magnitudes smaller than 'ε'. Ideally, the tube diameter depicts the amount of noise in the data. The optimization criterion in SVR penalizes those data points, whose 'y' values lie

more than ‘ ε ’ distance away from the regression function $f(x)$. The basic symmetric loss function (defined in Eqn. (2.21)) is modified for minimization of the empirical risk by adding the slack variables, ξ_j , and ξ_j' , $j = 1, 2, \dots, P$, into the function for a set of linear constraints. The slack variables, ξ_j , and ξ_j' are a measure of the deviation ($y_j - f(x_j)$) from the boundaries of the ε -insensitive zone. Thus, the overall loss function using the ε -insensitive loss function and introducing the regularization constant, ‘ C ’, the objective function to be minimized becomes,

$$\min L(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{j=1}^P (\xi_j + \xi_j') \quad (1.22)$$

$$\text{subject to, } \begin{cases} (w \cdot \Phi(x_j)) + b - y_j \leq \varepsilon + \xi_j' \\ y_j - (w \cdot \Phi(x_j)) - b \leq \varepsilon + \xi_j \\ \xi_j, \xi_j' \geq 0 \end{cases} \quad (1.23)$$

The SVR optimizes the position of the ε -insensitive tube around the data so as to get the best regression fit (see Figure (1.11)). Specifically, the optimization criterion in Eqn. (1.23) penalizes those training data points whose output values ‘ y ’ lie more than ‘ ε ’ distance away from the fitted function, $f(x)$. The stated excess positive (ξ) and negative (ξ') deviations are represented in terms of the slack variables in Figure (1.11). These slack variables assume non-zero values outside the ε -insensitive region. While fitting the function $f(x)$ to the training data, the SVR minimizes the training set prediction error by minimizing both the slack variables (ξ_j and ξ_j') and also $\|w\|^2$ for increasing the flatness of the function or penalizing its over-complexity. This avoids the problem of under-fitting and over-fitting of the training data to the fitting function and thus increases the generalization capability of the SVR model. Finally the dot product in Eqn. (1.16) is replaced with an appropriate kernel function to write the SVR-based regression function in its general form as,

$$f(\mathbf{x}, w) = f(\mathbf{x}, \lambda, \lambda^*) = \sum_{i=1}^P (\lambda - \lambda^*) K(\mathbf{x}, \mathbf{x}') + b \quad (1.24)$$

where, the weight vector ' w ' is expressed in terms of the Lagrange multipliers λ and λ^* . The values of these multipliers are obtained by solving a convex quadratic programming problem generated by combining the constraints with the objective function as in Eqn. (1.24). On solution of the optimization problems in terms of the Lagrange multipliers, we get only few of the coefficients, $(\lambda - \lambda^*)$, to be non-zero and the corresponding input vectors, x_j , are known as "support vectors (SVs)". The SVs are the most informative data points, which compress the information content of the training set spanning the whole training dataset to only a few example data-points in the training set. Cherkassky and Ma (2004) suggest important guidelines for the judicious selection of SVR parameters. A more detailed description of SVR is provided, for example by, Vapnik (1995) and Nandi et al. (2004).

Distinctive features of SVR are:

- SVR uses the principles of structural risk minimization, which penalizes the model complexity while minimizing the training dataset error, thus resulting in parsimonious models with lesser complexity and good generalization capability.
- The loss function minimized by SVR is quadratic function, which possesses a single minimum and, thus, it provides a globally optimum solution.
- SVR allows for a robust regression function with sparse solutions.

The SVM/SVR methodologies, due to their robustness and automatic control of the solution complexity, have been used in various process engineering tasks as listed in the table below.

Table 1.5: Recent chemical engineering/technology applications of support vector regression

Sr. No.	Application area	Specific study	Reference
1.	Process modeling	A bayesian inference based two-stage support vector regression framework for soft sensor development in batch bioprocesses	Yu (2012)
		Modelling and predictive control of a	Ławryńczuk

Sr. No.	Application area	Specific study	Reference
		neutralisation reactor using sparse support vector machine wiener models	(2016)
		Predictive model of chemical flooding for enhanced oil recovery purposes: application of least square support vector machine	Ahmadi and Pournik (2016)
2.	Process fault detection/diagnosis	Simultaneous fault diagnosis using multi class support vector machine in a dew point process	Pooyan et al. (2015)
		SVM and PCA based fault classification approaches for complicated industrial process	Jing and Hou (2015)
		Fed-batch fermentation penicillin process fault diagnosis and detection based on support vector machine	Yang and Hou (2016)
3.	Soft sensor development	Soft-sensor development for fed-batch bioreactors using support vector regression	Desai et al. (2006)
		A soft sensor based on adaptive fuzzy neural network and support vector regression for industrial melt index prediction	Zhang and Liu (2013)
4.	Quantitative Structure-Activity/property Relationships (QSAR/QSPR)	A new kernel function of support vector regression combined with probability distribution and its application in chemometrics and the QSAR modeling	Xue and Yan (2017)
5.	Data classification	Application of support vector machine to rapid classification of uranium waste drums using low-resolution γ -ray spectra	Hata et al. (2015)

The SVM/SVR algorithms are a part of various software packages. Following are the popular SVM/SVR based software packages,

1. *RapidMiner* (Mierswa et al., 2006): Along with numerous other data-driven modeling and optimization routines, RapidMiner Studio includes

SVM and SVR modules for conducting classification and modeling, respectively.

2. *LIBSVM* (Chang and Lin, 2001): LIBSVM is a free open source software with codes available in both C++ and Java and has facilities to perform support vector classification, (C-based and nu-based), regression (epsilon-based and nu-based) and distribution estimation (one class-based SVM).

1.3. PROCESS OPTIMIZATION

Optimization is the task of obtaining the best solution(s) from among multiple possible ones while satisfying pre-defined constraints. A typical optimization goal consists of maximization of profits expressed, for instance, as minimization of raw materials and/or resources used, or maximization of the product quantity along with the quality. For example, a typical reactor operation optimization involves determination of the optimum reactor operating and feed conditions so as to get maximum product yield from minimum amount of reactants, energy, catalyst, etc. Optimization requires an accurate model of the system under study. The objective function to be maximized/minimized using an optimization algorithm pertaining to a real life industrial optimization problem involves multiple variables and constraints.

1.3.1. Conventional Optimization Approaches

The conventional optimization approaches are based on *deterministic* methods with main emphasis on the gradient ascent/descent methods. Following conventional optimizations approaches are commonly used for unconstrained multi-variable optimization problems (Edger et al., 2001; Deb, 2012),

- Direct search methods: Pattern search, Conjugate direction, Simplex search.
- Gradient based methods: Cauchy's steepest descent method, Newton's method, Levenberg Marquardt (LM) method, Generalized Reduced Gradient (GRG) method, Nelder-Mead algorithm simplex method, Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, sequential linear programming (SLP), sequential quadratic programming (SQP).

The most commonly used gradient based methods are iterative methods in which first order (and second order) gradients of the objective function (to be minimized/maximized) with respect to the decision (independent) variables are determined at successive steps. For example, the Newton's method requires computing the first order as well as the second order gradients (in the Hessian matrix form: a matrix of second order partial derivatives) of the objective function with respect to the individual decision variables involved in the objective of the optimization problem. Other gradient based methods are variations of the Newton's method, with main focus on elimination of the computation of the second order gradients (Hessian matrix) at each step. The above stated conventional optimization approaches, commonly exhibit following limitations/disadvantages,

- The objective function needs to be smooth, continuous, and differentiable over the feasible region.
- The feasible region of a constrained optimization problem must form a convex region.
- The gradient based methods require determination of the second and/or first order derivatives, which involve a large amount of computations at each step.
- A major shortcoming of the conventional methods is that they invariably get stuck in a local optimum, thus leading to a sub-optimal solution.
- The convergence of conventional optimization depends on the chosen initial conditions (guess).
- The gradient based optimization methods are inefficient in handling discrete search space problems.

In most of the real-life industrial optimization problems, the objective function is non-smooth, noisy and discontinuous and, thus, conventional deterministic optimization approaches are not able to solve the problem efficiently.

1.4. COMPUTATIONAL INTELLIGENCE BASED OPTIMIZATION METHODOLOGIES

Computational intelligence includes a class of population based stochastic optimization methodologies such as genetic algorithms (GA), swarm intelligence (SI), and the newer artificial immune systems (AIS) (Engelbrecht, 2007), possessing some unique advantages over the conventional gradient based deterministic optimization methodologies. Specifically the CI-based optimization methodologies offer the following advantages,

- Being stochastic and population based, the CI-based methods generally don't get stuck in local optimum.
- The computationally intensive gradient calculations are not required in the CI-based methods.
- They can operate on any non-smooth, non-differentiable and noisy functions successfully.
- They can work with highly multi-variable and also multi-objective systems (pareto optimization) with ease.
- They can work with discrete search space problems, for e.g. using the binary versions.

In this thesis we have utilized the GA and AIS optimization methodologies, which are explained in detail in the next sections.

1.4.1. Genetic Algorithm (GA)

Genetic algorithms (GAs) (Holland, 1975; Goldberg, 1989) are the basic evolutionary computing algorithms used for stochastic optimization. Like GP, GAs are also based on the Darwin's principle of the "survival of the fittest" followed in biological evolution and the genetic propagation of characteristics over multiple generations. GAs are the most popular population based stochastic optimization formalisms used with a great success in solving problems involving very large search spaces (Goldberg, 1989). The process of natural evolution of biological species involves survival of the most adaptive organisms to changes in their environment and propagation of their genetic characteristics to the next generation by reproduction. In reproduction, genetic material of parents is passed

over to offspring via crossover, recombination and mutation operations. *Mutation* leads to combinatorial diverse characteristics in the subsequent generations, thus resulting in new (novel) characteristics in the offspring. The combined effects of selection, crossover, recombination and mutation lead to fitter individuals in the next generation, while also creating diversity of the population characteristics. Diversity is necessary for the generation of novel characteristics that may lead to even better generations. GA broadly works in this framework, wherein it considers the environment as an objective function to be maximized/minimized and the individuals in the environment as *candidate solutions* in the search space.

The CI-based stochastic optimization methodologies including GAs exhibit a remarkable capability of handling nonlinear and noisy objective functions as also solving multi-objective functions. Currently, several variations of GAs such as the parallel-GA (Pettey et al., 1987) and the adaptive-GA (Srinivas and Patnaik, 1994) are available for specific problems. In the parallel-GA several GA runs for a problem are implemented in parallel and migration of the best solutions among the individual parallel runs are allowed for better performance. While the adaptive-GA works using adaptive parameters in which the probabilities of crossover (p_c) and mutation (p_m) are adaptively adjusted during each iteration (generation) utilizing the population information in each generation in order maintain the population diversity as well as to sustain the convergence capacity.

GA Implementation

A typical multi-variable objective function for process optimization is given as,

$$\max/ \min y(\mathbf{x}) = f(\mathbf{x}, \mathbf{c}); \quad (1.25)$$

$$\text{subject to: } x_i^L < x_i < x_i^U \quad (1.26)$$


where, y is the process output (dependent) variable to be optimized (minimized or maximized), $\mathbf{x} = [x_1, x_2, \dots, x_i, \dots, x_N]^T$ is the vector of process input (independent) variables and \mathbf{c} is the vector of constants/parameters of the function ' f ' relating the input variables to the output variable of the process. The optimization problem is subject to variable constraints with a variable x_i limited to a range between its

lower limit x_i^L and upper limit x_i^U . The task of an optimizer is to determine the best value (maximum or minimum) of the process output ‘y’ by searching the optimal values for the input (decision) variables.

GA encodes the input variable values (commonly known as candidate solutions or chromosomes in the GA parlance) in different ways, according to which we have *binary encoded GA*, *gray encoded GA* or *real value encoded GA*.

Binary Encoded GA: In a binary valued GA, the candidate solutions (chromosomes) are represented by binary strings comprising of bits of 0 or 1. Shown below is a binary encoding of the input vector with two variables, wherein each variable is encoded as an 8-bit string and the chromosome becomes a 16-bit string.

Variable	x_1	x_2
Genes (Bit strings)	01101101	10101011



Chromosome

Thus a variable, x_i can be encoded into binary strings by first normalizing according to,


$$x_n = \frac{x - x^L}{x^U - x^L} \quad (1.27)$$

where, x_n = normalized value of variable ‘x’, x^L and x^U = lower and upper value of the variable x respectively, and then converting to binary digits. Binary GA works on these binary strings to give the optimized results in binary form, which are then quantized to decimal values. These decimal values are finally de-normalized to get the optimized input variable values.

Real Value Encoded GA: The encoding of the variables in this type of GA is as usual done using real floating point numbers as shown below. The advantage of using real value encoded GA is that it is possible to use large domains for the decision variables, which is difficult to achieve in the binary-coded GA

implementation where increasing the domain means sacrificing precision for a fixed length of the chromosome.

Variable	x_1	x_2
Genes (Real valued numbers)	18.34	0.28


 Chromosome

The general step-wise procedure for a typical GA run is explained below.

Initialization (Step 1): To begin, the user has to specify the variable encoding, the population size (N) and the fitness function according to the minimization/maximization criteria and the objective function. An initial population of candidate solutions is either generated randomly or is based on careful choosing of candidates based on user's experience.

Fitness evaluation (Step 2): In this step, the fitness of each candidate solution is determined based on the optimization criteria. Various forms of fitness functions are in use depending on the problem under study, with the well-known being the proportional fitness function, wherein the fitness F_i (in percentage) of a candidate solution (individual) for the maximization of an objective function f_i is calculated as:

$$F_i = \frac{f_i}{\sum_{i=1}^{N_p} f_i} \times 100 \quad (1.28)$$

where, N_p is the number of parent pairs in the mating pool. The fitness function assigns a higher fitness to better solutions. The handling of constraints in GA is accomplished by augmenting the objective function with a penalty term that gives more weightage to solutions in the feasible regions.

Selection (Step 3): In this step a "mating" pool of candidate solutions is created by selecting the top-ranking candidate solutions, based on their fitness. The occupants of this mating pool forming the parents, take part in the most important mating operation namely crossover (step 4), to produce the offspring candidate

solutions. The selection step allows the best candidate solutions to form the mating pool to undergo crossover. Several selection procedures exist, such as the *Roulette-wheel* selection (De Jong, 1975), *tournament* selection (Goldberg, 1990), *truncation* selection (Crow and Kimura, 1979), *elitist* selection, etc. Figure (1.12) represents an example of selection operation by the Roulette-wheel method, wherein four individuals with their objective function and fitness values are indicated in the upper table. It shows the Roulette wheel representation of the example problem, which if rotated at a random speed, will stop with the pointer selecting the most probabilistic value (the 55% region in the example) for the selection of individuals. The Roulette wheel selection method with varying selection probability over the generations is also used (Back and Schwefel, 1993).

Individual	Value	Fitness ($f_i/\sum f_i$)*100
I1	13.75	55%
I2	5	20%
I3	3.75	15%
I4	2.5	10%

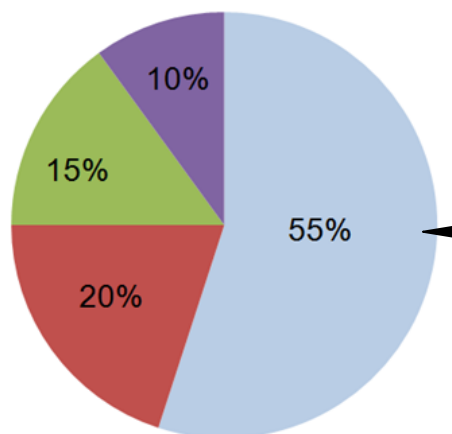


Figure 1.12: Schematic showing the Roulette wheel selection operation.

Crossover (Step 4): Here, two candidate solutions (parents) from the mating pool are randomly selected to undergo crossover. The crossover operation between a parent pair occurs by exchanging of sub-portions between the parent candidate solutions to create off-spring solutions. Figure (1.13) shows an illustration of such a one-point crossover operation on the candidate solutions represented by binary strings. The crossover point is chosen probabilistically. Two-point and multi-point crossovers are also common.

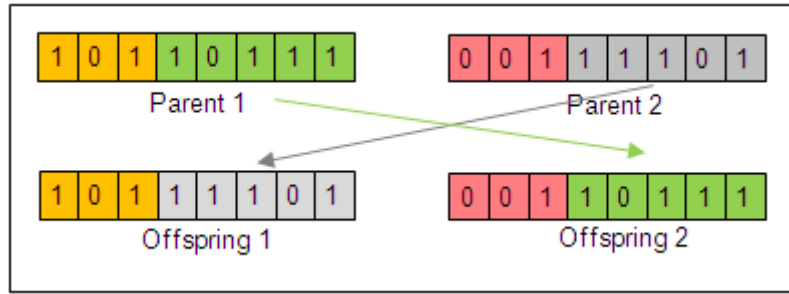


Figure 1.13: Schematic showing the one-point crossover operation.

The most common type of crossover operation for a real value encoded GA is given by Eshelman and Schaffer (1993), who introduced a blend crossover operator. For two parent candidate solutions y_{p1} and y_{p2} , the crossover operator randomly picks a real number as offspring in the given range as,

$$y_o = (1 - \alpha)y_{p1} + \alpha y_{p2} \quad (1.29)$$

where, y_o is the real valued offspring, $\alpha = (1 + 2\beta)r - \beta$, r is a random number in between 0 to 1 and β is a constant (generally 0.5).

Mutation (Step 5): In this step, the population of the selected parent and generated offspring tree solutions are subjected to probabilistic mutation. In mutation, the bits in an individual string are flipped, e.g. 0 to 1 or vice versa. The mutation is carried out randomly with a small probability, with an important motive to maintain diversity in the population and prevent the population from getting stuck in a local minimum. Figure (1.14) shows a mutation operation on an individual candidate solution represented by a binary string wherein a ‘0’ in the original individual is replaced randomly to ‘1’. The mutated population representing a new generation is now ready for the next iteration cycle and thus the generation index is incremented.

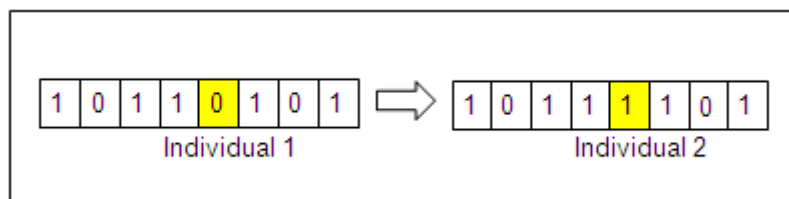


Figure 1.14: Schematic showing the mutation operation.

For the case of real valued GA the most common mutation operator is random mutation (Michalewicz, 1992), wherein the mutation operator generates a random solution within the entire variable range (x_{\min} to x_{\max}).

Termination (Step 6): At this stage the termination criteria is checked to ascertain if the stopping criteria has met. The stopping criterion as specified by the user could be either a specified number of generations or no significant change in the fitness value of the population over the successive generations. If the set termination criteria is not met, steps 2 to 5 are repeated till the convergence is reached.

Finally, the binary string coded optimum variables are decoded back to real numbers for the computation of the max/min objective function (output) value. Similar to a GP run, it is necessary for the GA to repeat several times using different random initializations to arrive at best optimum values of the objective function. A generalized pseudo-code of GA is given in Figure (1.15).

```

Randomly initialize a population of N candidate expressions
do
  Select best n candidate solutions according to fitness criteria
  (see Eqn. 1.28) to form parents in the mating pool

  for i = 1 to (N-n)/2
    Select a parent pair probabilistically from the mating
    pool
    Perform reproduction by crossover to generate offspring
    pair
  end for

  Replace the lower-fit (N-n) candidate solutions with the
  offspring population
  Mutate the population probabilistically
  generation = generation + 1
until the stopping criteria is met
return best candidate solution from the last generation

```

Figure 1.15: Pseudo-code of genetic algorithm (GA) for optimization.

In this thesis, we have applied GA in comparison with the artificial immune system (AIS) to a 3-dimensional input search space to determine the optimum reaction and process conditions for maximum percent adsorption of hexavalent chromium ions in waste water on a synthetic polymer resin. To perform the GA-based optimization an excel add-in tool, namely MendelSolve (2016) was used in excel to optimize the objective function along with the

constraints. MendelSolve offers to prespecify the following GA parameters before executing the algorithm: (i) ranges of input variables, (ii) cross-over probability, (iii) mutation probability and (iv) number of generations. The following table lists the most recent applications of GA for various process-related tasks as also for molecular modeling.

Table 1.6: Recent chemical engineering/technology applications of genetic algorithms

Sr. No.	Application area	Specific study	Reference
1.	Process Optimization	Reaction modeling and optimization of synthetic resin for the adsorptive removal of arsenite and arsenate ions from wastewater	Patil-Shinde et al. (2016)
		Genetic algorithm applied for the optimization of chemical reactors network	Leong et al. (2016)
		Genetic algorithms and simulated annealing applied for the optimization of type-4 composite pressure vessel	Alcántar et al. (2017)
2.	Process Modeling	Genetic algorithm based modeling of hot flow behavior of API-X70-micro-alloyed steel using design of experiments	Abarghooei et al. (2017)
		Hybrid approach using genetic algorithm for rate constant prediction using structures of reactants and solvent for Diels-Alder reaction	Datta et al. (2017)
3.	Process control	Genetic algorithm used for optimal control of batch cooling crystallizer	Amini et al. (2016)
4.	Process flowsheet synthesis	Genetic algorithm applied for automated process flowsheet synthesis for membrane processes	Shafiee et al. (2016)

Sr. No.	Application area	Specific study	Reference
5.	Molecular Modeling	Molecular design by combining genetic algorithms and COSMO-RS	Scheffczyk et al. (2016)

1.4.2. Artificial Immune System (AIS)

There are various AIS formalisms based on different mechanisms of the working of the biological immune system (BIS) system to attack and kill the invading pathogenic microorganisms (antigens) by producing antigen-specific antibodies entering a biological body. Currently, the following theories (mechanisms) of AIS are popularly used in the form of computational methodologies to solve the engineering tasks of optimization, clustering and pattern recognition for fault detection:

- Clonal selection theory (De Castro and Zuben, 2002): It is mainly based on the mechanism of the active B-cells producing antibodies through the cloning process which are mutated to maintain diversity.
- Negative selection theory (Forrest et al., 1994): This is based on the mechanism of the BIS, which provides for the ability of the T-cells to provide tolerance for self-cells.
- Danger theory (Matzinger, 2001): The main mechanism in this method has the ability to distinguish between the dangerous and non-dangerous antigens.
- Network theory (Jerne, 1974): In this theory it is assumed that the B-Cells form a network and when a B-cell encounters an antigen it becomes activated and stimulates all other B-cells connected in the network.

For accomplishing its objectives, the BIS system, mainly originating from the bone marrow starts generating lymphocytes as a result of an immune response. An immune response is the body's reaction to antigens to prevent them from damaging the body. Figure (1.16) indicates the important cells created in sequence in the immune response system.

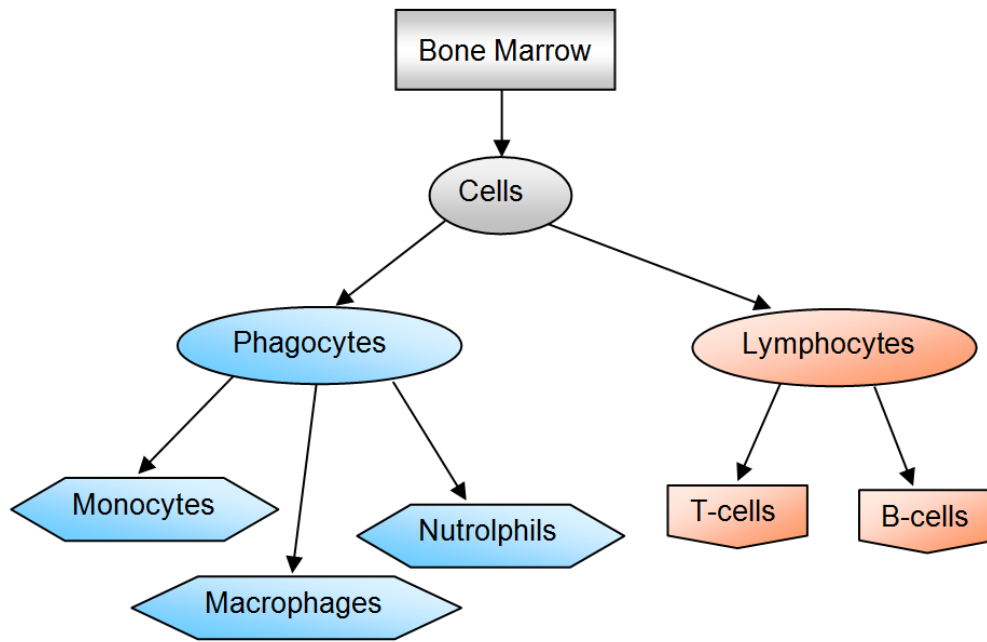


Figure 1.16: Types of cells generated by the immune response system.

Some of these cells develop into large white cells known as phagocytes (Schindler et al., 2002), which include sub-cells such as monocytes, macrophages and neutrophils. The macrophages cells secrete important chemicals that are useful in activation of T-Cells. Other cells secreted in the bone marrow develop into small white cells known as lymphocytes. Antibodies are formed when the antigen come into contact with lymphocytes. The T-cells and B-cells are the type of lymphocytes created in the bone marrow. On the invasion of a foreign substance (pathogens), delayed hypersensitivity T-cells (TDH, a type of T-cells) recognize the infection and in response start producing cytotoxic factor (protein messengers). This helps macrophages (a type of phagocytes) to reach the infection. The TDH cells then produce the migration inhibitory factor, which stop the macrophages from leaving their sites. The macrophages engulf the pathogenic microorganisms and then break into short chains of amino acids known as “peptides”. The peptides are held on the surface of the macrophages, where the helper T-cells recognize them. The T-cells then undergo maturation, wherein they undergo a selection process to ensure that they are able to recognize the pathogens (termed as non-self) held on the macrophages and also the macrophages (termed as self cells). If a T-cell recognizes itself, it is retained in the system else

discarded. This mechanism provides inspiration for the “positive selection” algorithm (Forrest et al., 1994) of the AIS class. As opposed to this the “negative selection” (Forrest et al., 1994) algorithm is based on the ability of the T-cells to provides tolerance for self-cells. Further the B-cells undergo important transformations to improvise the immune response to its final state and this forms the basis for an important class of AIS mechanism used mainly for stochastic optimization commonly known as the “clonal selection” (Burnet, 1978).

The *clonal selection algorithm* (CLONALG) of the AIS paradigm has been utilized in this work for process optimization and thus it is described in detail next.

(A) Clonal Selection Algorithm (CLONALG)

The *clonal selection algorithm* (CLONALG) as proposed by Burnet (1978) is a newer CI-based stochastic optimization methodology from the AIS paradigm. It is based on the working of a specific part of the immune response system of the BIS, known as the “clonal selection” mechanism. This mechanism starts after the recognition of the antigen, after which the antigen-activated T-cells secrete lymphokines. After the recognition of an antigen by the B-cells in presence of lymphokines secreted by the T-cells, the B-cells surface receptors bind to the antigen. These stimulated B-cells then start proliferating and differentiating. Proliferation mainly includes the generation of multiple copies of the stimulated B-cells with variations, which is known as cloning. The generated B-cells are diverted as to either memory cells or plasma cells. The plasma cells secrete a large amount of Y-shaped antigen-specific antibodies that bind themselves with the antigen, while the task of the memory cells is to ensure that if the same antigen attacks again the BIS receives a much faster response.

In the initial phase of the immune response, the affinity between the antibodies and antigens is low. However, in the intermediate phases the B-cells undergoing the process of “clonal selection”, clone and mutate repetitively to improve these binding affinities. The mutation process in which the poor antibodies (with lower affinities towards the antigens) are mutated with higher probabilities is known as “somatic hyper-mutation” and the entire process of creation of new antibodies with higher affinities to the antigens is called affinity

maturation (Dasgupta and Nino, 2009). Finally, the process of affinity maturation leads to the production of a pool of antibody-secreting B-cells, differentiated as the plasma cells and the memory cells (Dasgupta and Nino, 2009). The antigen coated with the matured antibody is ultimately killed by the Killer T-cells to combat the antigen attack. Cloning diverts CLONALG faster towards convergence, while hyper-mutation maintains diversity to achieve the global optimum and so CLONALG can achieve faster convergence towards the global optimum as compared to GAs. The working of a typical clonal selection process is shown schematically in Figure (1.17).

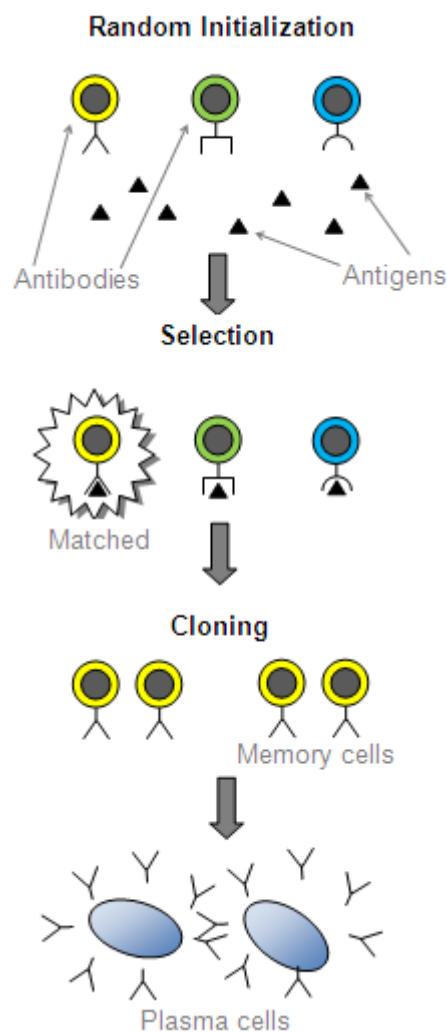


Figure 1.17: Mechanism of the clonal selection process.

CLONALG Implementation

For optimization of a pre-specified objective function, such as Eqn. (1.25), the binary version of CLONALG commonly operates over binary strings and a collection of the binary strings of all the inputs forms a candidate solution (antibody). The fitness score (affinity between the antibody and antigen) of an antibody is evaluated in terms of the objective function, e.g. as given in Eqn. (1.28). In case of optimization, the antigen forms the final desired value of the objective function. A collection of fittest (best) antibodies from the population are selected and further processed iteratively according to the algorithm shown as a pseudo-code in Figure (1.18) till a stopping criteria such as the number of generations or the error tolerance of the solutions is satisfied (whichever happens earlier). Namely after creation of the initial random population, a sub-population containing the best solutions is selected and cloned to create multiple copies, with the number of copies proportional to the affinities of individual solutions according to the equation,

$$N_c = \text{round}\left(\frac{\beta \times N}{i}\right) \quad (1.30)$$

where N_C is the number of clones generated for ' i^{th} ' candidate solution ($i = 1, 2, \dots, N_S$), N_S is the number of selected antibodies, β is the cloning factor ($\cong 1$) (Aragón et al., 2007) and N is the population size. Thus, here the selected antibodies proliferate for the overall average population to move towards the best antibodies in the selected sub-population. The resulting sub-population is then hyper-mutated to introduce diversity. The probability of hyper-mutation, ' h_P ' of a antibody is given by,

$$h_P = \exp(-\rho * f) \quad (1.31)$$

where, ρ = hyper-mutation coefficient, f = normalized fitness function (affinity) of the anti-bodies. A sub-population of the resulting population containing the best solutions is combined with a randomly generated population and forms the next generation.


```

Randomly initialize a population of N candidate solutions
(antibodies)

do
select Ns best candidate solutions (antibodies) for cloning
according to fitness criteria (see Eqn. 1.28)

for i = 1 to Ns
    Clone each selected antibody with affinity according to Eqn.
    (1.30)
    Hyper-mutate the population probabilistically according to Eqn.
    (1.31)
end for

Replace the lower-fit (N-Ns) candidate solutions (antibodies)
with newly generated random candidate solutions (antibodies)
maturity = maturity + 1
until the stopping criteria is met
return best candidate solution (antibody) from the matured pool

```

Figure 1.18: Pseudo-code of the clonal selection algorithm (CLONALG).

With the increasing use of various AIS-based formalisms for cyber security (fault detection applications in computer science/engineering), the use of CLONALG methodology is also on the rise for chemical process optimization and fault detection applications and the table below, listing the most recent applications of CLONALG also indicate the same.

Table 1.7: Recent chemical engineering/technology applications of clonal selection algorithm

Sr. No.	Application area	Specific study	Reference
1.	Process optimization	Optimization of distillation resources based on neighborhood-clonal selection learning algorithm.	Yang and Shi (2012)
		Biodegradable iron chelate for H ₂ S abatement: modeling and optimization using artificial intelligence strategies.	Hamid et al. (2014)
		Process real-time optimization using CLONALG algorithm applied to the Tennessee Eastman process.	Zhong et al. (2015)
2.	Process modelling	Evolutionary hybrid configuration applied to a polymerization process modeling.	Curteanu et al. (2015)
3.	Process fault detection/diagnosis	A new fault classification approach applied to Tennessee Eastman	D'Angelo et al. (2016)

		benchmark process based on the association of CLONALG immune systems with the kohonen neural network.	
4.	Process control	Research on tuning parameters for a model predictive controller based on CSA in CSTR process.	Jiang et al. (2015)

In this thesis work, CLONALG has been used for optimization of adsorptive removal reaction of chromium ions from waste water using synthetic polymer resin. For this, a MATLAB code was developed in-house, which is listed in Appendix 7.A of Chapter 7.

1.5. PROCESS FAULT DETECTION AND DIAGNOSIS APPROACHES

In process industries safety is of utmost importance and accidents are mostly caused due to faulty working of equipment(s)/device(s)/controller(s) and/or human error. Today's process industry is highly automated and thus equipment(s)/device(s)/controller(s) faults are the major reason behind industrial accidents and financial losses. A process fault indicates those process symptoms that result from the undesired physical changes in the process. Over the years, on-line fault detection and diagnosis have been given a serious attention by researchers and chemical engineers (Caccavale and Pierri, 2009). In a *fault detection and diagnosis* (FDD) task, the first objective is to recognize the deviations (from normal operation), disturbances or equipment/sensor/actuator malfunction followed by investigating the fault cause-effect relationship.

The FDD methods mostly rely on modeling methodologies that can accurately model the faulty and normal operations of a process. Such methods are either based on the dynamical process models or clustering/pattern recognition models. The former types of models are either based on the phenomenological modeling approach of the process in case the complete knowledge of the process (under normal and faulty operating conditions) is available or empirical (including CI-based) modeling approaches, wherever such a detailed knowledge is not available. The clustering/ pattern recognition modeling approaches work on the

fact that during a fault the important process variables start indicating abnormal readings, deviating from the normal variable values, often forming clusters/groups/patterns according to the type and magnitude of the fault. Thus cluster analysis/pattern recognition of a process dataset composed of the important process variable values as operating in the normal and various faulty conditions helps in the accurate detection and diagnosis of faults generated in the process.

Clustering

A *cluster* is a collection of similar objects and clustering involves classification of similar objects into different groups/classes/categories according to some common trait such as the “proximity” to each other assessed using some distance measure. Clustering identifies common patterns in the dataset and groups them in clusters thus generating different clusters according to the presence of different similar patterns. In this thesis, clustering methods are used for fault detection and diagnosis of a biochemical process generating biomass.

In mathematical terms, clustering tries to subdivide a data set \mathbf{X} into c subsets (clusters), with each subset containing similar objects. Figure (1.19) illustrates the concept for a 2-dimensional dataset $\mathbf{D} = \{x_1, x_2\}$, wherein three groups (marked with different shape icons) of close data-points is clearly observed.

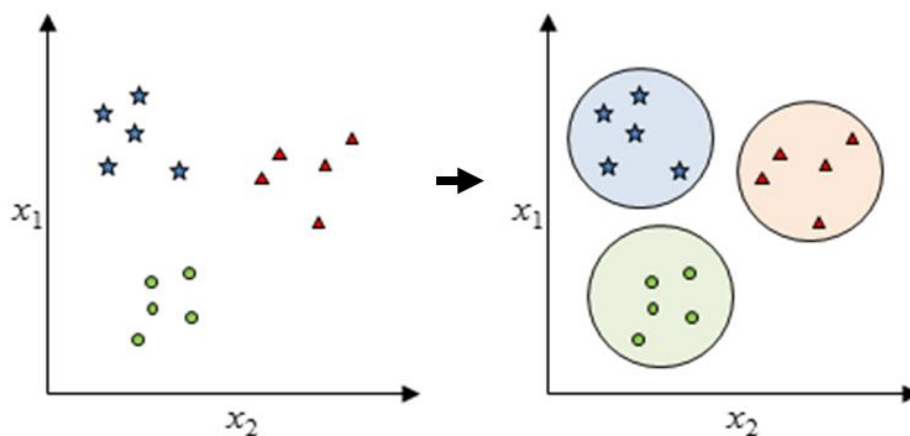


Figure 1.19: Schematic of clustering of a 2-dimensional input space.

1.5.1. Conventional Clustering Methodologies

Conventionally following are the classes of clustering methods commonly employed for various applications which mainly include fault detection and diagnosis, voice and image analysis, biomolecular recognition etc.,

- Connectivity-based clustering: this includes *hierarchical clustering* which build the clustering model using the distance connectivity.
- Centroid-based clustering: in these methods a cluster is represented by a single average vector of inputs known as the centroid. The popular *k-means algorithm* falls under this class of methods.
- Distribution-based clustering: in these methods the clusters are modeled by statistical distributions. The *expectation-maximization algorithm* lies under this class.
- Density-based clustering: as the name implies these methods form clusters based on connected dense regions in the data space. Common methods include *DBSCAN* (Ester et al., 1996) and *OPTICS* (Ankerst et al., 1999).

k-means (also known as c-means) is one of the popular conventional centroid based hard clustering algorithms (MacQueen, 1967) that follows a simple procedure to classify a given data set into a certain number of clusters (assuming k clusters) fixed *a priori*. The algorithm starts with k centroids, placed randomly, one for each cluster and iterates to refine the locations of these centroids so as to minimize the average distance between the data-points in a cluster and the centroid representing that cluster. Although the k-means algorithm can be proved to converge to appropriate k-centroids, it does not necessarily find the most optimal configuration, corresponding to the global objective function minimum. Another disadvantage of this algorithm is that it is highly sensitive to the initial randomly generated cluster centroid locations.

1.6. COMPUTATIONAL INTELLIGENCE BASED CLUSTERING METHODOLOGIES

Computational intelligence offers various clustering methodologies that can determine complex clusters in a multi-dimensional dataset. Cluster modeling

is provided by most of the regression CI-based formalisms such as ANNs (MLPNN using a hard threshold transfer function in the output layer nodes) (Charalampidis and Muldrey, 2009), self organizing maps (SOM) based on the ANNs (Pandit et al., 2011), GP (Boric and Estevez, 2007), SVMs (Ben-Hur et al., 2001) and fuzzy-logic based clustering method (Bezdek et al., 1984).

1.6.1. Fuzzy c-Means Clustering (FCC)

Fuzzy c-means clustering (FCC) (Bezdek et al., 1984) is a soft clustering methodology that uses unsupervised learning for classifying the data-points to one of the c -clusters. FCC tries to determine soft (i.e., fuzzy) c -partitions of a dataset \mathbf{X} . Figure (1.20) indicates the hard and soft clustering concepts of a dataset.

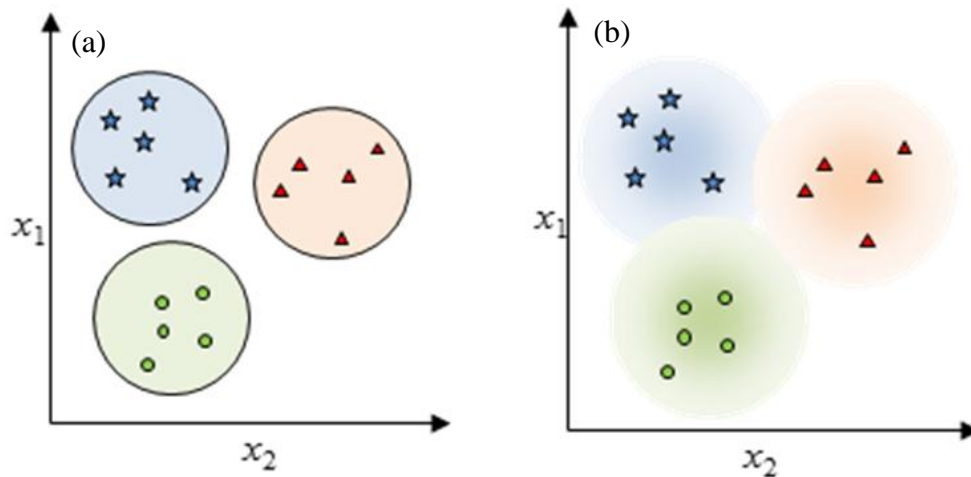


Figure 1.20: Schematic showing clustering types in a 2-dimensional space: (a) hard clustering, (b) soft clustering.

The idea behind the working of fuzzy clustering was laid by Zadeh (1965), which is to represent the similarity a point shares with each cluster with a function (known as membership function). The membership functions as discussed previously in Section (1.0 (D)) have values (called memberships) between 0 and 1. Each data-point in the dataset will have a membership in every cluster, and a data-point is assigned to a cluster having its membership value close to unity for that cluster, signifying a high degree of similarity. The membership values close to zero imply little similarity between the sample and that cluster. It was Bezdek et al. (1984) who formulated the idea to produce fuzzy c -partitions of a given dataset. A fuzzy c -partition of the dataset \mathbf{X} , characterizes membership values of each data-point in all the clusters by membership values ranging between 0 to 1.

The sum of the membership values for each data-point in the dataset must be unity. For a given dataset $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_P\}$ containing P observations in N -dimensional space, a conventional hard clustering algorithm will cluster it into c -clusters of \mathbf{X} in a c -tuple $(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_c)$ of subsets of \mathbf{X} . FCC constructs a U ($c \times N$) matrix, representing the partition $\{\mathbf{X}_i\}$, with membership function values $u_i: \mathbf{X} \rightarrow \{0,1\}$ whose values $u_i(\mathbf{x}_p) \in [0,1]$ are interpreted as the grades of membership of the individual x_k in the fuzzy subsets u_i of \mathbf{X} .

A centroid v_k is calculated from the cluster membership values u_{ik} as assigned to all the data-points (x_i) in the dataset according to the equation,

$$v_k = \frac{\sum_{i=1}^P u_{ik}^m x_i}{\sum_{i=1}^P u_{ik}^m} \quad (1.32)$$

where, $m =$ weighing exponent factor ($1 \leq m < \infty$).

Next the fuzzy membership values of each point (u_{ik}) with respect to each cluster is calculated as,

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{\|x_i - c_k\|}{\|x_i - c_j\|} \right)^{\frac{2}{m-1}}} \quad (1.33)$$

where, the subscript 'j' stands for the j^{th} cluster.

The criteria proposed to cluster the dataset is the generalized least-squared error,

$$J(U, V) = \sum_{k=1}^P \sum_{i=1}^c (u_{ik})^m \|x_k^{(j)} - v_i\|^2 \quad (1.34)$$

where, $c =$ number of clusters in \mathbf{X} , $U =$ fuzzy c -partition of \mathbf{X} , $\mathbf{V} = [v_1, v_2, \dots, v_c]^T$ is a matrix of the vector of centroids and $\mathbf{v}_i = [v_{i1}, v_{i2}, \dots, v_{iN}]^T$ is the centroid location vector of cluster 'i'. Figure (1.21) indicates the pseudo-code of the FCC algorithm for the determination of c -centroid locations representing the c -clusters of the dataset.

```

Initialize the algorithm with parameter values: c, m, U-matrix and
start with 'c' centroids with random locations
do
  for
    Compute the membership value of each point with respect
    to each centroid according to Eqn. (1.32)

    Compute the new centroid locations according to the
    minimum distance criteria in Eqn. (1.34)
  end for

until the centroids no longer move
return the U-matrix and the c-centroid locations as final
prototypes

```

Figure 1.21: Pseudo-code of the fuzzy c-means (FCC) clustering algorithm.

Since its invent, the applications of fuzzy logic and various methodologies based on FL are on the rise. Table (1.8) indicates some recent process engineering applications of fuzzy clustering.

Table 1.8: Recent chemical engineering/technology applications of fuzzy clustering

Sr. No.	Application area	Specific study	Reference
1.	Process fault detection and diagnosis	Sequential fuzzy clustering based dynamic fuzzy neural network for fault diagnosis and prognosis more efficient high speed milling (HSM) processes	Jahromi et al. (2016)
		A novel fault diagnosis scheme applying fuzzy clustering algorithms for nonlinear continuous stirred-tank reactor.	Rodriguez et al. (2017)
		Structural analysis based sensor measurement fault diagnosis in cement industries used fuzzy clustering algorithm for fault detection and isolation in cement kilns.	Gomathi et al. (2017)
		Expert diagnosis of polymer electrolyte fuel cells a fuzzy logic diagnostic algorithm for polymer electrolyte fuel cells is proposed.	Davies et al. (2017)
2.	Cluster Analysis	A modified fuzzy c-means clustering algorithm and its application on carbonate fluid identification	Liu et al. (2016)
		Hybrid chemical reaction based	Nayak et al.

		metaheuristic with fuzzy c-means algorithm for optimal cluster analysis	(2017)
3.	Quantitative Structure-Activity/Property Relationships (QSAR/QSPR)	Fuzzy clustering as rational partition method for QSAR	Pérez-Garrido et al. (2017)

The FCC algorithm for data clustering is provided by the following software packages,

1. *MATLAB Fuzzy Logic Toolbox* (MathWorks and Wang, 1998): It includes tools for modeling, classification and clustering of data in a given dataset using the fuzzy logic formalism. The toolbox includes a *fuzzy c-means clustering* facility, using which a large dataset can be clustered into a given number of ‘c’ clusters. Given a dataset the FCC algorithm returns back the U-matrix and a matrix of the optimal ‘c’ centroids.
2. *Online Fuzzy Clustering* (Aydos, 2015): This is an online webpage that offers the *fuzzy c-means clustering* facility, wherein we can upload our dataset, choose the fuzzy parameters (c, m, ε) and then the calculations are done on-line to return the results as the cluster centroids using the FCC methodology.

1.7. DIMENSIONALITY REDUCTION METHODOLOGY: Principal Component Analysis (PCA)

In a chemical process, huge amounts of data comprising values of the important process (input-output) variables and parameters are generated and recorded continuously for process analysis. These data inherently contain instrumental and measurement noise. Also, often, the variables are correlated linearly/nonlinearly giving rise to redundancy. Thus, it is necessary to preprocess the data to remove correlated variables and the noise and retain only the important features in the dataset thus reducing its dimensionality. Dimensionality reduction, aids in constructing parsimonious process models possessing good generalization capabilities.

Following are the popular linear dimensionality reduction techniques, while the nonlinear PCA can be performed, for example, using auto-associative and Sammon's neural networks (Lerner et al., 1999).

- Principal component analysis (PCA)
- Kernel PCA
- Linear discriminant analysis (LDA)
- Generalized discriminant analysis (GDA)
- Auto-associative Neural Network (AANN)
- Curvilinear Component Analysis (CCA)

Principal component analysis (PCA) (Pearson, 1901; Geladi and Kowalski, 1986) is the most widely used linear dimensionality reduction technique that extracts linear relationships existing among the variables of a dataset. PCA was first introduced by the statistician, Pearson (Pearson, 1901) for the formulation of the analysis of determining the closest fitting lines and planes for a system of points in a given space. Thereafter PCA was found to be more suitable for the analysis of variance for modeling of response data (Fisher and MacKenzie, 1923). Later the percentage concept in PCA was developed by Hotelling (1933). PCA determines the linear mapping between the variables in the original dataset in a low dimensional space in such a way that the variance of the data in the low dimensional representation is maximized. This is done by decomposing the original dataset comprising the linearly correlated variables into a PCA transformed variable set defining the eigenvectors of the covariance of the data. In practice, a correlation matrix of the dataset is constructed and the eigenvectors of this matrix are determined. The eigenvectors corresponding to the largest eigenvalues (termed as the principal components (PCs)) are used to reconstruct a large fraction of the variance of the original data. Thus PCA generates PCs as a set of pseudo-variables (also known 'latent variables'), which are linearly independent (uncorrelated) orthogonal variables. The importance of PCs is that the first few latent variables capture maximum amount of variability in the data and are thus used to represent the original dataset variables in a compressed way. Mathematically, for a two dimensional matrix, $A(R \times C)$, with R rows containing

the data-points for C columns of variables, PCA decomposes, A , into matrices of latent variables and the associated parameters (known as “loadings”) as given by:

$$A = TP' + E \quad (1.35)$$

where, matrix A is a normalized matrix with zero mean and variance scaled so that the standard deviation of elements of each column is unity. T ($R \times C$) denotes the matrix of C principal component (PC) scores with each column of matrix T signifying a principal component, while P' refers to the transpose of the loading matrix, $P(C \times C)$, and E indicates the residuals. The first ‘ G ’ principal component scores capture a large amount of variance in the dataset, and thus the original matrix can also be written as,

$$A = \sum_{k=1}^G s_k (l_k)' + E' \quad (1.36)$$

where, $s_k = R$ -dimensional k^{th} score vector, $l_k =$ transpose of the k^{th} J -dimensional loading vector l_k , and $E' =$ residual matrix. Thus the original ($R \times C$) dimensional dataset matrix, A , is now represented in terms of a smaller number ‘ J ’ of I -dimensional score vectors. The sum of squares of elements of the score vector s_k is related to the eigenvalue of that vector, which forms a measure of the variance captured by the k^{th} PC. Thus a larger eigenvalue indicates more significance of the respective PC. This feature of data reduction of the PCA has been utilized in various process engineering tasks as listed below.

Table 1.9: Recent chemical engineering/technology applications of principal component analysis

Sr. No.	Application area	Specific study	Reference
1.	Process fault detection and diagnosis	Fault detection in time-varying chemical process through incremental principal component analysis	Gao et al. (2016)
		On the structure of dynamic principal component analysis used in statistical process monitoring	Vanhatalo et al. (2017)
		Fault propagation path estimation in NGL	Ahmed et al.

Sr. No.	Application area	Specific study	Reference
2.	Quantitative Structure-Activity/Property Relationships (QSAR/QSPR)	fractionation process using principal component analysis	(2017)
		Fault discriminant enhanced kernel principal component analysis incorporating prior fault information for monitoring nonlinear processes	Deng et al. (2017)
		Combining DFT and QSAR studies for predicting psychotomimetic activity of substituted phenethylamines using statistical methods	Aouidate et al. (2016)
		Structure activity relationship modelling of milk protein-derived peptides with dipeptidyl peptidase IV (DPP-IV) inhibitory activity	Nongonierna and Fitzgerald (2016)
		Biological activities of triazine derivatives; combining DFT and QSAR results	Larif et al. (2017)

1.8. STATISTICAL SIGNIFICANCE TEST

METHODOLOGY: Steiger's z-Test

In a variety of situations multiple models are been developed for a common task for the sake of comparison, and identification of the best model unambiguously, is a difficult task as more often the models perform equally well with minor chance errors. This task of the identification of the best model in terms of statistical comparisons between correlation coefficients of the models is eased by a statistical test known as the Steiger's z-test (Steiger, 1980), wherein each *CC* is converted to a z-score using the Fisher's r-to-z transformation and then they are compared statistically. Specifically, this test while comparing the performances of two competing models, examines whether the two *CCs* corresponding to the predictions of the models are significantly different or same. It tests for the *null*

hypothesis (H_0) that statistically two correlation coefficient magnitudes are same, i.e. $CC_{AB} = CC_{AC}$, where CC_{AB} is the correlation coefficient for the model B when compared with the experimental target values (A) and so is CC_{AC} for the model C. For the comparison, initially the z-scores are computed using the Fisher's r-to-z transformation equation,

$$z_{ij} = \frac{1}{2} \ln \left(\frac{1+r_{ij}}{1-r_{ij}} \right) \quad (1.37)$$

where, r_{ij} = correlation coefficient between readings 'i' and 'j'.

$$h = \text{cov}(r_{AB}, r_{AC}) \quad (1.38)$$

$$z = \frac{(z_{AB} - z_{AC})\sqrt{N_R - 3}}{\sqrt{2 - 2hr_{BC}}} \quad (1.39)$$

where, $\text{cov}(x,y)$ is a function that determines the covariance between the readings x and y , z_{AB} and z_{AC} are the Fisher's z-transformations of r_{AB} and r_{AC} respectively, N_R is the number of readings and r_{BC} is the correlation coefficient between the model predicted values of models 'B' and 'C'. Finally the p -value is computed as,

$$p = 2(1 - \text{normdist}(\text{abs}(z))) \quad (1.40)$$

where, $\text{normdist}(x)$ is a function that determines the normal distribution function of 'x', $\text{abs}(x)$ is a function that determines the absolute value of 'x'. If the p -values obtained are less than 0.05, then it suggests for the rejection of the null hypothesis (with 95% confidence level) inferring that the models being compared are significantly distinct. Thus ultimately, from the results of the Steiger's z-test for a pair of models, it is possible to identify the model possessing higher prediction accuracy and generalization capability. This test has been utilized in chapter 3, 4 and 6 to compare the developed CI-based models with existing non-CI based models as well as among themselves to identify the best performing CI-based model.

1.9. CONCLUSION

The CI-based methods offer a range of advanced, convenient and effective tools for conducting a number of important process engineering tasks such as modeling, optimization, control, monitoring, identification, and fault detection.

When compared with the conventional approaches, the CI-based modeling methodologies such as artificial neural networks, genetic programming, support vector regression and fuzzy c-means clustering, and optimization methodologies, namely, genetic algorithm, and clonal selection algorithm among others offer several advantages. Accordingly, this chapter provides a rather detailed overview of the computational intelligence discipline, and a number of CI-based modeling and optimization paradigms that are used in conducting the studies presented in this thesis. The chapter also highlights the attractive features of CI-based methods vis a vis their conventional counterparts. The chapter also provides details of the PCA and a statistical test that have been extensively used in performing dimensionality reduction of the input space of the models and comparing model performance, respectively. The various methods detailed in this chapter and their chemical engineering applications described in subsequent chapters are as follows:

- Chapter 2: In this chapter the CI-based methodologies, namely, multilayer perceptron neural network (MLPNN) belonging to the ANN class of methods and the GP-based symbolic regression (GPSR) have been introduced to develop nonlinear models predicting the higher heating value (HHV) of solid biomass fuels from the constituents of their proximate or ultimate analysis as inputs.
- Chapter 3: The CI-based methodologies, namely, MLPNN, GPSR and SVR have been applied to model the elemental composition of solid biomass fuels using the constituents of their proximate analysis as inputs.
- Chapter 4: The HHV of coal being its most important property has been previously modeled by linear/nonlinear regression-based as well as CI-based methods (except GP). In this chapter the modeling of coal's HHV using the GPSR methodology has been introduced to develop high performing coal HHV predicting correlations that are applicable for coals of various ranks and from different geographic locations of the world.
- Chapter 5: In this chapter GPSR-based soft-sensor models are developed for the process identification of styrene polymerization process. Here, the best GPSR-based model has been shown to perform successfully in a

model predictive control scheme to accurately control quality of the effluent polystyrene product.

- Chapter 6: Co-gasification is a newer gaseous fuel producing technology that mitigates the problem of pollution and adds operational flexibility to gasification processes. The MLPNN, GPSR and SVR methodologies have been introduced for performance modeling of a pilot plant scale fluidized bed co-gasifier utilizing different coal-biomass blends as feedstock. Prior to the CI-based modeling, PCA was performed on the experimental co-gasification process dataset to reduce the dimensionality of the dataset for easing the task of CI-based models development.
- Chapter 7: In this chapter a hybrid CI-based modeling-optimization strategy has been proposed to model and optimize the chromium ion removal process from contaminated water by utilizing a synthetic polymer resin. Specifically, the CI-based stochastic optimization methodology, namely, “clonal selection algorithm” (CLONALG) belonging to the artificial immune system (AIS) class of formalisms has been utilized to optimize the process, which is modeled by another CI-based methodology, namely, MLPNN. The CI-based genetic algorithm (GA) has also been utilized for the said optimization for the purpose of comparison with CLONALG.
- Chapter 8: Biochemical processes are highly sensitive to faults and an accurate fault detection and diagnosis (FDD) model can be useful in preventing extreme faulty situations by taking immediate corrective actions once minor faults are detected by the model. Accordingly, the CI-based clustering methodology, namely fuzzy c-means clustering (FCC) has been successfully applied to accurately model faults in an appropriately controlled, continuous bioreactor generating biomass. In comparison to FCC, the conventional k-means clustering method has also been applied to the said FDD problem.

NOMENCLATURE

<i>A</i>	Normalized dataset matrix in PCA
<i>b</i>	Bias
<i>c</i>	Constant/Cluster in FCC
<i>D</i>	Data-set
<i>d</i>	Degree of polynomial
<i>e</i>	Error (difference between actual and target output value)
<i>E</i>	Residual matrix in PCA
<i>f</i>	Linear/nonlinear function
<i>F</i>	Objective function in GA-based optimization
<i>G</i>	Principal component score in PCA
<i>H</i>	Activation function in ANN/Null hypothesis in Steiger's test
<i>h</i>	Hyper-mutation parameter
<i>J</i>	Least-squared error objective function
<i>l</i>	Library symbol in GP-based regression
<i>L</i>	Number of first hidden layer nodes in MLPNN/Loss function in SVR
<i>l</i>	Loading vector in PCA
<i>M</i>	Number of second hidden layer nodes in MLPNN
<i>m</i>	Weighing exponent factor in FCC
<i>N</i>	Number of inputs/data-points/Number of input layer nodes in MLPNN/Population size in CLONALG
<i>o</i>	Operator symbol in GP-based regression
<i>P</i>	Loading matrix in PCA

p	Probability
L	Number of nodes in first hidden layer of an MLPNN
M	Number of nodes in second hidden layer of an MLPNN/Random number in crossover operator in GA/ Correlation coefficient
S	Number of nodes in output layer of MLPNN/Score vector in PCA
T	Temperature/Matrix of principal components in PCA
u	Fuzzy membership value
U	Fuzzy c-partition matrix
v	Centroid of a cluster
w	Connection weight in an ANN
x	Input/independent variable
\mathbf{x}	Input vector
X	Dataset matrix
y	Output/dependent variable
z	Output of summing junction in an ANN node/ Standardized normalized variable

Greek letters

μ	Fuzzy membership function/Momentum coefficient in the EBP algorithm
ε	Epsilon, a precision parameter in SVR
ρ	Hyper-mutation coefficient
σ	Scale parameter in SVR
α	Parsimony pressure in GPSR objective function

β	Penalty coefficient in GP/cloning factor in CLONALG
η	Learning rate in the EBP algorithm
γ	Kernel gamma of radial basis function in SVR
λ	Lagrange multiplier
ζ	Slack variable in optimization
Φ	Kernel function

REFERENCES

- Abarghooei, H., Arabi, H., Seyedein, S. H., Mirzakhani, B. (2017). Modeling of steady state hot flow behavior of API-X70 microalloyed steel using genetic algorithm and design of experiments. *Applied Soft Computing* 52:471-477.
- Adar, E., İnce, M., Karatop, B., Bilgili, M. S. (2017). The risk analysis by failure mode and effect analysis (FMEA) and fuzzy-FMEA of supercritical water gasification system used in the sewage sludge treatment. *Journal of Environmental Chemical Engineering* 5(1):1261-1268.
- Ahmadi, M. A., Pournik, M. (2016). A predictive model of chemical flooding for enhanced oil recovery purposes: application of least square support vector machine. *Petroleum* 2(2):177-182.
- Ahmed, U., Ha, D., An, J., Zahid, U., Han, C. (2017). Fault propagation path estimation in NGL fractionation process using principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 162:73-82.
- Alcántar, V., Aceves, S. M., Ledesma, E., Ledesma, S., Aguilera, E. (2017). Optimization of Type 4 composite pressure vessels using genetic algorithms and simulated annealing. *International Journal of Hydrogen Energy* 42(24):15770-15781.
- Amini, Y., Gerdroodbary, M. B., Pishvaie, M. R., Moradi, R., Monfared, S. M.

- (2016). Optimal control of batch cooling crystallizers by using genetic algorithm. *Case Studies in Thermal Engineering* 8:300-310.
- Ankerst, M., Breunig, M. M., Kriegel, H. P., Sander, J. (1999). OPTICS: ordering points to identify the clustering structure. In: *ACM Sigmod record* 28(2), pp. 49-60.
- Aouidate, A., Ghaleb, A., Ghamali, M., Chtita, S., Choukrad, M., Sbai, A., Lakhlifi, T. (2016). Combining DFT and QSAR studies for predicting psychotomimetic activity of substituted phenethylamines using statistical methods. *Journal of Taibah University for Science*, 10(6), 787-796.
- Aragón, V. S., Esquivel, S. C., Coello Coello, C. A. (2007). Artificial immune system for solving constrained optimization problems. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial* 11(35).
- Araromi, D. O., Sonibare, J. A., Emuoyibofarhe, J. O. (2014). Fuzzy identification of reactive distillation for acetic acid recovery from waste water. *Journal of Environmental Chemical Engineering* 2(3):1394-1403.
- Aydos, H. (2015). <https://aydos.com/fcm/>
- Bäck, T., Schwefel, H. P. (1993). An overview of evolutionary algorithms for parameter optimization. *Evolutionary computation* 1(1):1-23.
- Bahrami, P., Kazemi, P., Mahdavi, S., Ghobadi, H. (2016). A novel approach for modeling and optimization of surfactant/polymer flooding based on Genetic Programming evolutionary algorithm. *Fuel* 179:289-298.
- Balandina, G. I. (2017). Control System Synthesis by Means of Cartesian Genetic Programming. *Procedia Computer Science* 103:176-182.
- Ben-Hur, A., Horn, D., Siegelmann, H. T., Vapnik, V. (2001). Support vector clustering. *Journal of machine learning research* 2:125-137.
- Benmouna, A., Becherif, M., Depernet, D., Gustin, F., Ramadan, H. S., Fukuhara, S. (2017). Fault diagnosis methods for Proton Exchange Membrane Fuel Cell system. *International Journal of Hydrogen Energy* 42(2):1534-1543.

- Berlinski, D. (2000). *The Advent of the Algorithm*. Harcourt Book, San Diego, USA.
- Bezdek, J. C., Ehrlich, R., Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences* 10(2-3):191-203.
- Bishop, C. M. (1994). *Neural networks and their applications*. *Review of Scientific Instruments* 65(6):1803-1832.
- Boric, N., Estevez, P. A. (2007). Genetic programming-based clustering using an information theoretic fitness measure. In: *IEEE Congress on Evolutionary Computation, CEC 2007*, IEEE, pp. 31-38.
- Bose, B. K. (1994). Expert system, fuzzy logic, and neural network applications in power electronics and motion control. *Proceedings of the IEEE* 82(8):1303–23.
- Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* 2(2):121-167.
- Burnet, F. M. (1978). Clonal selection and after. In: *Theoretical Immunology*, Bell, G. I., Perelson, A. S., Pimbley, G. H. Eds. New York: Marcel Dekker, pp. 63–85.
- Campbell, D. T. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review* 67(6):380–400.
- Chang, C. C., Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)* 2(3):27. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Charalampidis, D., Muldrey, B. (2009). Clustering using multilayer perceptrons. *Nonlinear Analysis: Theory, Methods & Applications* 71(12):e2807-e2813.
- Cherkassky, V., Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks* 17(1):113-126.

- Crow, J. F., Kimura, M. (1979). Efficiency of truncation selection. *Proceedings of the National Academy of Sciences* 76(1):396-399.
- Curteanu, S., Dragoi, E. N., Dafinescu, V. (2015). Evolutionary Hybrid Configuration Applied to a Polymerization Process Modelling. In: *International Work-Conference on Artificial Neural Networks*, Springer, pp. 237-249.
- D'Angelo, M. F., Palhares, R. M., Camargos Filho, M. C., Maia, R. D., Mendes, J. B., Ekel, P. Y. (2016). A new fault classification approach applied to Tennessee Eastman benchmark process. *Applied Soft Computing* 49:676-686.
- Dasgupta, D., Nino, L.F. (2009). *Immunological Computation*. 1st ed., Auerbach Publications, Taylor & Francis Group, Boca Raton.
- Datta, S., Dev, V. A., Eden, M. R. (2017). Hybrid genetic algorithm-decision tree approach for rate constant prediction using structures of reactants and solvent for Diels-Alder reaction. *Computers & Chemical Engineering*. In Press, Corrected Proof, Available online 17 February 2017.
- Davies, B., Jackson, L., Dunnett, S. (2017). Expert diagnosis of polymer electrolyte fuel cells. *International Journal of Hydrogen Energy* 42(16):11724-11734.
- De Castro, L. N., Von Zuben, F. J. (2002). Learning and optimization using the clonal selection principle. *IEEE transactions on evolutionary computation* 6(3):239-251.
- De Jong, K. A. (1975). An analysis of the behavior of a class of genetic adaptive systems. (Doctoral dissertation, University of Michigan). *Dissertation Abstracts International*, 36(10), 5140B. (University Microfilms No. 76-9381)
- Deb, K. (2012). *Optimization for engineering design: Algorithms and examples*. PHI Learning Pvt. Ltd..

- Deng, X., Tian, X., Chen, S., Harris, C. J. (2017). Fault discriminant enhanced kernel principal component analysis incorporating prior fault information for monitoring nonlinear processes. *Chemometrics and Intelligent Laboratory Systems* 162:21-34.
- Denisov, I. A. (2017). Luciferase-based bioassay for rapid pollutants detection and classification by means of multilayer artificial neural networks. *Sensors and Actuators B: Chemical* 242:653-657.
- Desai, K., Badhe, Y., Tambe, S. S., Kulkarni, B. D. (2006). Soft-sensor development for fed-batch bioreactors using support vector regression. *Biochemical Engineering Journal* 27(3):225-239.
- Edgar, T. F., Himmelblau, D. M., Lasdon, L. S. (2001). *Optimization of chemical processes*. McGraw-Hill, New York.
- Emigdio, Z., Abatal, M., Bassam, A., Trujillo, L., Juárez-Smith, P., El Hamzaoui, Y. (2017). Modeling the adsorption of phenols and nitrophenols by activated carbon using genetic programming. *Journal of Cleaner Production* (Online: In Press).
- Engelbrecht, A. P. (2007). *Computational Intelligence An Introduction*. John Wiley & Sons Ltd, Southern Gate, West Sussex, England.
- Esfandyari, M., Fanaei, M. A., Gheshlaghi, R., Mahdavi, M. A. (2016). Neural network and neuro-fuzzy modeling to investigate the power density and Columbic efficiency of microbial fuel cell. *Journal of the Taiwan Institute of Chemical Engineers* 58:84-91.
- Eshelman, L. J., Schaffer, J. D. (1993). Real-coded genetic algorithms and interval-schema, *Foundations of Genetic Algorithms 2*, Morgan Kaufmann Publishers pp. 187-202.
- Ester, M., Kriegel, H. P., Sander, J., Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Kdd* 96(34), pp. 226-231.

- Estevez, P. A. (2016). IEEE CIS Presidents welcome address: President of the IEEE Computational Intelligence Society. <http://cis.ieee.org/about-cis.html>. Accessed: 10 May 2017
- Farmer, J. D., Packard, N., Perelson, A. (1986). The immune system, adaptation and machine learning. *Physica D* 2:187–204.
- Fisher, R. A., Mackenzie, W. A. (1923). Studies in crop variation. II. The manurial response of different potato varieties. *The Journal of Agricultural Science* 13(3):311-320.
- Fissore, D. (2016). On the Design of a Fuzzy Logic–Based Control System for Freeze-Drying Processes. *Journal of pharmaceutical sciences* 105(12):3562-3572.
- Forrest, S., Perelson, A. S., Allen, L., Cherukuri, R. (1994). Self-nonsel self discrimination in a computer. In *Research in Security and Privacy, 1994. Proceedings., 1994 IEEE Computer Society Symposium on*, pp. 202-212.
- Fortune, Betting on AI, March 2017.
- Freeman, J. A., Skapura, D. M. (1991). *Neural Networks: Algorithms, Applications, and Programming Techniques*. Addison-Wesley Reading.
- Gabbar, H. A., Hussain, S., Hosseini, A. H. (2014). Simulation-based fault propagation analysis—application on hydrogen production plant. *Process Safety and Environmental Protection* 92(6):723-731.
- Gao, Y., Wang, X., Wang, Z., Zhao, L. (2016). Fault detection in time-varying chemical process through incremental principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 158:102-116.
- Geladi, P., Kowalski, B.R. (1986). Partial least-squares regression: a tutorial, *Analytica Chimica Acta* 185:1–17.
- Ghanem, O. B., Mutalib, M. A., Lévêque, J. M., El-Harbawi, M. (2017). Development of QSAR model to predict the ecotoxicity of *Vibrio*

fischeri using COSMO-RS descriptors. *Chemosphere* 170:242-250.

Goel, P., Bapat, S., Vyas, R., Tambe, A., Tambe, S. S. (2015). Genetic programming based quantitative structure–retention relationships for the prediction of Kovats retention indices. *Journal of Chromatography A* 1420:98-109.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*, Reading: Addison-Wesley.

Goldberg, D. E. (1990). A note on Boltzmann tournament selection for genetic algorithms and population-oriented simulated annealing. *Complex Systems* 4(4):445-460.

Gomathi, V., Srinivasan, S., Ramkumar, K., Muralidharan, G. (2017). Structural analysis based sensor measurement fault diagnosis in cement industries. *Control Engineering Practice* 64:148-159.

Govindhasamy, J. J., McLoone, S. F., Irwin, G. W., French, J. J., Doyle, R. P. (2005). Neural modelling, control and optimisation of an industrial grinding process. *Control Engineering Practice* 13(10):1243-1258.

Hamid, A., Deshpande, A. S., Badhe, Y. P., Barve, P. P., Tambe, S. S., Kulkarni, B. D. (2014). Biodegradable iron chelate for H₂S abatement: Modeling and optimization using artificial intelligence strategies. *Chemical Engineering Research and Design* 92(6):1119-1132.

Hata, H., Yokoyama, K., Ishimori, Y., Ohara, Y., Tanaka, Y., Sugitsue, N. (2015). Application of support vector machine to rapid classification of uranium waste drums using low-resolution γ -ray spectra. *Applied Radiation and Isotopes* 104:143-146.

Hezave, A. Z., Raeissi, S., Lashkarbolooki, M. (2012). Estimation of thermal conductivity of ionic liquids using a perceptron neural network. *Industrial & Engineering Chemistry Research* 51(29):9886-9893.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. University of

Michigan Press, Ann Arbor, MI.

Hosseini, S. H., Karami, M., Olazar, M., Safabakhsh, R., Rahmati, M. (2014). Prediction of the minimum spouting velocity by genetic programming approach. *Industrial & Engineering Chemistry Research* 53(32):12639-12643.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology* 24(6):417.

IBM SPSS. (2011). *Neural Networks 20 manual* IBM: Chicago, USA.

Izadiyan, P., Fatemi, M. H., Izadiyan, M. (2013). Elicitation of the most important structural properties of ionic liquids affecting ecotoxicity in limnic green algae; a QSAR approach. *Ecotoxicology and environmental safety* 87:42-48.

Jahromi, A. T., Er, M. J., Li, X., Lim, B. S. (2016). Sequential fuzzy clustering based dynamic fuzzy neural network for fault diagnosis and prognosis. *Neurocomputing* 196: 31-41.

Jerne, N. K. (1974). Towards a network theory of the immune system. *Annual Review of Immunology* 125:373-389.

Jiang, W., Shi, X., Chen, Y., Yongqi, C., Zhao, J., Xu, Z. (2015). Research on tuning parameters for a model predictive controller based on CSA in CSTR process. In: *Chinese Automation Congress (CAC), IEEE*, pp. 247-252.

Jing, C., Hou, J. (2015). SVM and PCA based fault classification approaches for complicated industrial process. *Neurocomputing* 167:636-642.

Khoury, M. (2009). pySTEP. Under the MIT license. <http://sourceforge.net/projects/pystep/>

Koç, D. İ., Koç, M. L. (2015). A genetic programming-based QSPR model for predicting solubility parameters of polymers. *Chemometrics and Intelligent Laboratory Systems* 144:122-127.

- Koza, J. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA.
- Larif, M., Adad, A., Hmammouchi, R., Taghki, A. I., Soulaymani, A., Elmidaoui, A., Lakhlifi, T. (2017). Biological activities of triazine derivatives. Combining DFT and QSAR results. *Arabian Journal of Chemistry* 10: S946-S955.
- Lashkarbolooki, M., Hezave, A. Z., Bayat, M. (2017). Correlating thermal conductivity of pure hydrocarbons and aromatics via perceptron artificial neural network (PANN) method. *Chinese journal of chemical engineering* 25(5):547-554.
- Ławryńczuk, M. (2016). Modelling and predictive control of a neutralisation reactor using sparse support vector machine Wiener models. *Neurocomputing* 205:311-328.
- Leong, C. C., Blakey, S., Wilson, C. W. (2016). Genetic Algorithm optimised Chemical Reactors network: A novel technique for alternative fuels emission prediction. *Swarm and Evolutionary Computation* 27:180-187.
- Lerkkasemsan, N. (2017). Fuzzy logic-based predictive model for biomass pyrolysis. *Applied Energy* 185:1019-1030.
- Lerner, B., Guterman, H., Aladjem, M. (1999). A comparative study of neural network based feature extraction paradigms. *Pattern Recognition Letters* 20(1):7-14.
- Liu, L., Sun, S. Z., Yu, H., Yue, X., Zhang, D. (2016). A modified Fuzzy C-Means (FCM) Clustering algorithm and its application on carbonate fluid identification. *Journal of Applied Geophysics* 129:28-35.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* 1(14), pp. 281-297.
- MathWorks, Inc, Wang, W. C. (1998). *Fuzzy Logic Toolbox: for Use with*

MATLAB: User's Guide. Mathworks, Incorporated.

MATLAB.(2000). The MathWorks Inc. Natick, MA, www.mathworks.com/Products/MATLAB.

Matzinger, P. (2001). Essay 1: the Danger model in its historical context. *Scandinavian journal of immunology* 54:1-2.

McCorduck, P. (2004). *Machines Who Think*. 2nd ed., A. K. Peters Ltd., Natick, Massachusetts, USA.

MendelSolve, A Genetic Algorithm Solver, (2016). Accessed: 17 September 2016. <http://www.bluestretch.com/mendelsolve/index.htm>.

Michalewicz, Z. (1992). *Genetic Algorithms + Data Structures = Evolution programs*. Springer-verlag, New York.

Michalopoulos, J., Papadokonstadakis, S., Arampatzis, G., Lygeros, A. (2001). Modelling of an industrial fluid catalytic cracking unit using neural networks. *Chemical Engineering Research and Design* 79(2):137-142.

Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., Euler, T. (2006). YALE: rapid prototyping for complex data mining tasks. In: *Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining (KDD-06)*.

Mohanty, S. (2009). Artificial neural network based system identification and model predictive control of a flotation column. *Journal of Process Control* 19(6):991-999.

Moore, S. W., Gardner, J. W., Hines, E. L., Göpel, W., Weimar, U. (1993). A modified multilayer perceptron model for gas mixture analysis. *Sensors and Actuators B: Chemical* 16(1-3):344-348.

Moral, H., Aksoy, A., Gokcay, C. F. (2008). Modeling of the activated sludge process by using artificial neural networks with automated architecture screening. *Computers & Chemical Engineering* 32(10):2471-2478.

- Nandi, S., Badhe, Y., Lonari, J., Sridevi, U., Rao, B. S., Tambe, S. S., Kulkarni, B. D. (2004). Hybrid process modeling and optimization strategies integrating neural networks/support vector regression and genetic algorithms: study of benzene isopropylation on Hbeta catalyst. *Chemical Engineering Journal* 97(2):115-129.
- Nayak, J., Naik, B., Behera, H. S., Abraham, A. (2017). Hybrid chemical reaction based metaheuristic with fuzzy c-means algorithm for optimal cluster analysis. *Expert Systems with Applications* 79: 282-295.
- Nongonierma, A. B., FitzGerald, R. J. (2016). Structure activity relationship modelling of milk protein-derived peptides with dipeptidyl peptidase IV (DPP-IV) inhibitory activity. *Peptides* 79:1-7.
- Omidvari, M., Lavasani, S. M. R., Mirza, S. (2014). Presenting of failure probability assessment pattern by FTA in Fuzzy logic (case study: distillation tower unit of oil refinery process). *Journal of Chemical Health and Safety* 21(6):14-22.
- Pandit, Y. P., Badhe, Y. P., Sharma, B. K., Tambe, S. S., Kulkarni, B. D. (2011). Classification of Indian power coals using K-means clustering and Self Organizing Map neural network. *Fuel* 90(1):339-347.
- Pardo, M., Sisk, B. C., Sberveglieri, G., Lewis, N. S. (2006). Comparison of Fisher's linear discriminant to multilayer perceptron networks in the classification of vapors using sensor array data. *Sensors and Actuators B: Chemical* 115(2):647-655.
- Patil-Shinde, V., Mulani, K. B., Donde, K., Chavan, N. N., Ponrathnam, S., Tambe, S. S. (2016). The Removal of arsenite [As (III)] and arsenate [As (V)] ions from wastewater using TFA and TAFA resins: Computational intelligence based reaction modeling and optimization. *Journal of Environmental Chemical Engineering* 4(4):4275-4286.
- Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and*

Journal of Science 2(11):559-572.

- Pérez-Garrido, A., Girón-Rodríguez, F., Bueno-Crespo, A., Soto, J., Pérez-Sánchez, H., Helguera, A. M. (2017). Fuzzy clustering as rational partition method for QSAR. *Chemometrics and Intelligent Laboratory Systems* 166:1-6.
- Pettey, C.B., Leuze, M.R., Grefenstette, J. J. (1987). Genetic algorithms and their applications: proceedings of the second International Conference on Genetic Algorithms: July 28-31, 1987 at the Massachusetts Institute of Technology, Cambridge, MA.
- Podola, B., Melkonian, M. (2012). Genetic Programming as a tool for identification of analyte-specificity from complex response patterns using a non-specific whole-cell biosensor. *Biosensors and Bioelectronics* 33(1):254-259.
- Pooyan, N., Shahbazian, M., Hadian, M. (2015). Simultaneous fault diagnosis using multi class support vector machine in a dew point process. *Journal of Natural Gas Science and Engineering* 23:373-379.
- Rad, M. A. A., Yazdanpanah, M. J. (2015). Designing supervised local neural network classifiers based on EM clustering for fault diagnosis of Tennessee Eastman process. *Chemometrics and Intelligent Laboratory Systems* 146:149-157.
- Rodriguez, R. A., Llanes-Santiago, O., Bernal, D. J. M., Cruz, C. C., Silva, N. A. J., Verdegay, G. J. L. (2017). A novel fault diagnosis scheme applying fuzzy clustering algorithms. *Applied Soft Computing* 58: 605-619.
- Rogina, A., Šiško, I., Mohler, I., Ujević, Ž., Bolf, N. (2011). Soft sensor for continuous product quality estimation (in crude distillation unit). *Chemical Engineering Research and Design* 89(10):2070-2077.
- Rostami, A., Arabloo, M., Ebadi, H. (2017a). Genetic programming (GP) approach for prediction of supercritical CO₂ thermal conductivity. *Chemical Engineering Research and Design* 122:164-175.

- Rostami, A., Ebadi, H., Arabloo, M., Meybodi, M. K., Bahadori, A. (2017b). Toward genetic programming (GP) approach for estimation of hydrocarbon/water interfacial tension. *Journal of Molecular Liquids* 230:175-189.
- Rumelhart, D., Hinton, G., Williams, R. (1986). Learning representations by back-propagating error. *Nature* 323(9):533-536.
- Russell, S. J., Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*. 3rd ed., Prentice Hall, Upper Saddle River, New Jersey, USA.
- Sakthivel, N. R., Nair, B. B., Sugumaran, V. (2012). Soft computing approach to fault diagnosis of centrifugal pump. *Applied Soft Computing* 12(5):1574-1581.
- Scheffczyk, J., Fleitmann, L., Schwarz, A., Bardow, A., Leonhard, K. (2016). Computer-Aided Molecular Design by Combining Genetic Algorithms and COSMO-RS Computer Aided Chemical Engineering 38:115-120.
- Schindler, L., Kerrigan, D., Kelly, J. (2002). Understanding the Immune System. *Science Behind the News - National Cancer Institute*, <http://newscenter.cancer.gov/cancertopics/understandingcancer/immunesystem>.
- Schmidt, M., Lipson, H. (2009). Distilling free-form natural laws from experimental data. *science* 324(5923):81-85.
- Shafiee, A., Arab, M., Lai, Z., Liu, Z., Abbas, A. (2016). Automated process flowsheet synthesis for membrane processes using genetic algorithm: role of crossover operators. In: *26th European Symposium on Computer Aided Process Engineering*. Elsevier BV.
- Sharma, S., Tambe, S. S. (2014). Soft-sensor development for biochemical systems using genetic programming. *Biochemical Engineering Journal* 85:89-100.
- Siddique, N., Adeli, H. (2013). *Computational Intelligence: Synergies of Fuzzy Logic*. *Neural Networks and Evolutionary Computing*, John Wiley &

Sons.

- Soleimani, R., Shoushtari, N. A., Mirza, B., Salahi, A. (2013). Experimental investigation, modeling and optimization of membrane separation using artificial neural network and multi-objective optimization using genetic algorithm. *Chemical Engineering Research and Design* 91(5):883-903.
- Srinivas, M., Patnaik, L. M. (1994). Adaptive probabilities of crossover and mutation in genetic algorithms. *IEEE Transactions on Systems, Man, and Cybernetics* 24(4):656-667.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin* 87(2):245.
- Storn, R., Price, K. (1997). Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization* 11:341-359.
- Sun, K., Huang, S., Jang, S., Wong, D. S. (2016). Development of soft sensor with neural network and nonlinear variable selection for crude distillation unit process. *Computer Aided Chemical Engineering* 38:337-342.
- Tambe, S. S., Kulkarni, B. D., Deshpande, P. B. (1996). Elements of artificial neural networks with selected applications in chemical engineering, and chemical & biological sciences. Simulation & Advanced Controls Inc., Louisville.
- Tan, W. L., Nor, N. M., Bakar, M. A., Ahmad, Z., Sata, S. A. (2012). Optimum parameters for fault detection and diagnosis system of batch reaction using multiple neural networks. *Journal of Loss Prevention in the Process Industries* 25(1):138-141.
- Tanatavikorn, H., Yamashita, Y. (2015). Improving Data Reliability for Process Monitoring with Fuzzy Outlier Detection. *Computer Aided Chemical Engineering* 37:1595-1600.
- Tarca, L. A., Grandjean, B. P., Larachi, F. (2004). Designing supervised

- classifiers for multiphase flow data classification. *Chemical engineering science* 59(16):3303-3313.
- Tatar, A., Naseri, S., Bahadori, M., Hezave, A. Z., Kashiwao, T., Bahadori, A., Darvish, H. (2016). Prediction of carbon dioxide solubility in ionic liquids using MLP and radial basis function (RBF) neural networks. *Journal of the Taiwan Institute of Chemical Engineers* 60:151-164.
- Tóth, A., Hantos, K. M. (2016). A diagnostic method based on clustering qualitative event sequences. *Computers & Chemical Engineering* 95:58-70.
- Vanhatalo, E., Kulahci, M., Bergquist, B. (2017). On the structure of dynamic principal component analysis used in statistical process monitoring. *Chemometrics and Intelligent Laboratory Systems* 167:1-11.
- Vapnik, V. (1995). *The nature of statistical learning theory*. Springer Verlag, New York.
- Vasičkaninová, A., Bakošová, M., Mészáros, A. (2016). Fuzzy Control of a Distillation Column. *Computer Aided Chemical Engineering* 38:1299-1304.
- Vyas, R., Goel, P., Tambe, S. S. (2015). Genetic programming applications in chemical sciences and engineering. In *Handbook of Genetic Programming Applications*; Gandomi, A.H., Amir H., Alavi, Ryan, C. (Eds.), Springer International Publishing, Switzerland pp. 99–140.
- Xue, S., Yan, X. (2017). A new kernel function of support vector regression combined with probability distribution and its application in chemometrics and the QSAR modeling. *Chemometrics and Intelligent Laboratory Systems* 167:96-101.
- Yang, C., Hou, J. (2016). Fed-batch fermentation penicillin process fault diagnosis and detection based on support vector machine. *Neurocomputing* 190:117-123.
- Yang, Z., Shi, X. (2012). Optimization of distillation resources based on

neighborhood-clonal selection learning algorithm. *Journal of Chemical Industry and Engineering* 63(9).

Yu, J. (2012). A Bayesian inference based two-stage support vector regression framework for soft sensor development in batch bioprocesses. *Computers & Chemical Engineering* 41:134-144.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control* 8(3):338–353.

Zhang, M., Liu, X. (2013). A soft sensor based on adaptive fuzzy neural network and support vector regression for industrial melt index prediction. *Chemometrics and Intelligent Laboratory Systems* 126:83-90.

Zhong, Y., Yang, C., Yuchen, C., Xuhua, S. (2015). Process real-time optimization using Clonalg algorithm. In: 27th Chinese Control and Decision Conference (CCDC), IEEE, pp. 743-748.

Zhou, Q., Wu, W., Liu, D., Li, K., Qiao, Q. (2016). Estimation of corrosion failure likelihood of oil and gas pipeline based on fuzzy logic approach. *Engineering Failure Analysis*, 70:48-55.

Zurada, J. M. (1992). *Introduction to Artificial Neural Systems*. West Publishing Company, St. Paul, MN 55101, USA.

CHAPTER 2.

COMPUTATIONAL INTELLIGENCE BASED MODELING OF HIGHER HEATING VALUE OF SOLID BIOMASS FUELS

COMPUTATIONAL INTELLIGENCE BASED MODELING OF HIGHER HEATING VALUE OF SOLID BIOMASS FUELS

Abstract

The renewable nature of biomass fuels has made them an important environment-friendly energy source of the future. The *higher heating value* (HHV) of a biomass fuel is its most important property, as it reflects the energy producing potential of the biomass fuel. The HHV of a fuel is closely correlated with its composition. The composition of solid fuels is expressed in terms of their proximate and ultimate analyses. The existing biomass HHV models based on the proximate or ultimate analysis are mostly based on the conventional empirical linear regression analysis although nonlinear relations have been observed between the biomass HHV and some components of the proximate and ultimate analyses. This observation suggests that nonlinear correlations are expected to fare better than the linear ones for HHV prediction. Accordingly, *computational intelligence* (CI) based methodologies such as *genetic programming based symbolic regression* (GPSR) and *multilayer perceptron artificial neural network* (MLPNN) have been introduced for the development of nonlinear correlations for the prediction of biomass HHV from the components of their proximate or ultimate analyses. The prediction accuracies of the developed CI-based models were compared with the existing high-performance biomass HHV models. This analysis shows that, the GPSR and MLPNN based models are better in terms of possessing prediction accuracies than that of their existing counterparts. Further, range-wise HHV comparisons were also performed between the CI-based and the existing high-performance models. The CI-based biomass HHV models introduced here, due to their excellent prediction and generalization performances possess the potential to replace the conventional linear/nonlinear regression based models.

2.0. INTRODUCTION

Solid biomasses are the second largest energy source after coal. Although coal has a higher energy producing capacity (energy efficiency) as compared to the biomass fuels, its' fast depleting reserves has made biomass as a potential source of energy for the future. The major advantage of biomass fuels is that they are renewable and environment-friendly as compared with the currently dominant energy sources. The sources of biomasses include all once living materials such as plants and organic matter derived from plants. The common usage of solid biomass is either in the form of raw/dried wood/crop or as processed biomass cakes/pellets after removal of moisture. Although the disadvantage of biomasses is their seasonal availability, their huge diversity makes them a potential source for energy generation (Schenk et al., 2008). The major contents of a biomass are the carbohydrates (around 75 wt%) and lignin (around 25 wt%). The former consists of cellulosic or hemicellulosic fibers, imparting strength to the plant structure. Plant-based biofuels are produced mainly from their lignocellulosic components such as stems, fruits, barks, branches, and leaves, which are further processed thermally/biochemically to obtain the biofuels. The diverse biomass resources used in producing the wide variety of bioenergy are derived from (Boundy et al., 2011),

- i. Primary lignocellulosics: These are obtained from the basic resources (plants), such as grasses, trees, and stalks of food crops.
- ii. Secondary lignocellulosics: These are obtained from all the byproducts formed by processing the primary lignocellulosics (pulping black liquor from the paper and pulp industry, sawdust, food waste, manure, etc.).

In order to use the solid biomasses as energy sources (fuel), it is necessary to determine its energy producing capacity in terms of the higher heating value, which forms a necessary exercise before designing combustion/gasification devices.

2.0.1. Higher Heating Value

The *higher heating value* (HHV) of a solid fuel is the amount of heat produced by complete combustion of fuel including the energy leaving with the

water vapors in the flue gases. It is also termed *gross calorific value* or *gross energy* and is commonly expressed in MJ/kg in the SI units. The composition and, thus, the heat content of biomasses vary significantly with the geographic and environmental conditions in which they are grown (Ciolkosz, 2010; Tahir et al., 2011).

The composition of a biofuel is expressed in terms of its *proximate* and *ultimate* analyses. The former a raw measure of the composition of the biomass and is expressed as the weight percentages of *moisture (M)*, *ash (ASH)*, *volatile matter* (weight of matter evaporated when heated to 950°C) (*VM*) and *free carbon* (matter left at 950°C) (*FC*). The ultimate analysis is a finer measure of composition than the proximate analysis and provides its composition (in wt%) in terms the major elemental components, namely, *carbon (C)*, *hydrogen (H)*, and *oxygen (O)*, and minor components *sulfur (S)* and *nitrogen (N)*. The experimental determination of HHV of a biomass is done using an instrument known as *Bomb calorimeter*, and is time-consuming and tedious. The said calorimeter is a high-pressure batch equipment that requires a caution for its safe operation. It has been found that there exist relationships between the heating value of a biomass and its composition. Thus, if this functional relationship is known, then the knowledge of the proximate/ultimate analysis can be used gainfully to compute the HHV of biomass fuels (Parikh et al., 2005). The key advantages of such correlations are as given below.

- i. Efforts in the experimental determination of the HHV of a new biomass sample are saved if its proximate/ultimate analysis is known.
- ii. The correlations provide a fast and easy way of obtaining the biomass HHV.
- iii. These help in the performance modeling and optimal design and operation of combustion and gasification processes involving biomass fuels.

For their wide applicability, the biomass HHV correlations must have excellent prediction accuracy and generalization capability. Accordingly, this chapter presents CI-based correlations with high prediction and generalization performance for predicting the HHV of biomasses.

The next section begins with a brief literature survey of the existing biomass HHV models and explains the need for nonlinear correlations. It is followed by the details of the model development and a comparison of their prediction and generalization performance. Lastly, the main findings of the study are presented.

2.1. SURVEY OF BIOMASS HHV MODELS

2.1.1. Need for Nonlinear Models

Over the years, various attempts have been made to develop biomass HHV models from their proximate or ultimate analysis. Numerous proximate or ultimate analysis based regression models are currently available for the prediction of HHV of a wide range of biofuels (Milne et al., 1990; Jimenez and Gonzales, 1991; Cordero et al., 2001; Channiwala and Parikh, 2002; Parikh et al., 2005; Sheng and Azevedo, 2005; Friedl et al., 2005; Yin, 2011; Daya and Abdul, 2012); a majority of these are linear. The only nonlinear biomass HHV models currently available are those proposed of Friedl et al. (2005) and Daya and Abdul (2012). For developing new CI-based models, a large-sized dataset of 382 proximate and 536 ultimate analyses of biomass samples, and the corresponding HHVs was compiled. This data-set was obtained from the standard research articles and open-source databases (ECN Phyllis, 2012; Hofbauer and BIOBIB, 2012; IEA Bioenergy Task, 2013). The compiled datasets are available on-line at the following address: https://static-content.springer.com/esm/art%3A10.1007%2Fs12155-013-9393-5/MediaObjects/12155_2013_9393_MOESM1_ESM.docx. The prediction accuracies of the existing biomass HHV models for the compiled biomass dataset were evaluated in terms of three statistical metrics, viz *correlation coefficient (CC)*, *root mean squared error (RMSE)*, and *mean absolute percent error (MAPE)* and it was found that the models developed by Parikh et al. (2005) (Eqn. 2.1), Channiwala and Parikh (2002) (Eqn. 2.2), Milne et al. (1990) (Eqn. 2.3) and Freidl et al. (2005) (Eqn. 2.4), possess relatively high biomass HHV prediction accuracies.

$$HHV(MJ/kg) = 0.3536 * FC + 0.1559 * VM - 0.0078 * ASH \quad (2.1)$$

$$HHV(MJ/kg) = 0.3491 * C + 1.1783 * H + 0.1005 * S - 0.1034 * O - 0.0151 * N - 0.0211 * ASH \quad (2.2)$$

$$HHV(MJ/kg) = 0.341 * C + 1.322 * H - 0.12 * (O + H) + 0.0686 * S - 0.0153 * ASH \quad (2.3)$$

$$HHV(KJ/kg) = 3.55 * C^2 - 232 * C - 2230 * H + 51.2 * C * H + 131 * N + 20600 \quad (2.4)$$

where, FC = fixed carbon, VM = volatile matter, ASH = ash, C = carbon, H = hydrogen, O = oxygen, N = nitrogen and S = sulfur, with all variables referring to the weight percentages of respective quantities on dry basis.

The HHVs of the biomass fuel samples from the compiled data-set were plotted against the individual components of the proximate and ultimate analyses; these are shown in Figure (2.1) as scatter-plots. From observation of the scatter-plots, it is found that linear relations exist between the HHV and weight percent composition of the C , FC and ASH , while nonlinear dependence is observed in the scatter-plots of HHV versus weight percentages of VM , H , O , N and S . Thus, linear HHV correlations for biomass fuels from their proximate and ultimate analyses seem to be inappropriate, and nonlinear models may prove to be better in capturing the underlying nonlinearities, resulting in better prediction accuracies.

The main difficulty in regression based modeling of the HHV from the components of the proximate and ultimate analyses is that the exact forms of the correlations are unknown. This limitation can be overcome by using the CI-based exclusively data-driven nonlinear modeling formalisms such as *genetic programming* (GP) and *artificial neural network* (ANN). An ANN such as *multilayer perceptron artificial neural network* (MLPNN), through the use of a pre-specified nonlinear transfer function in its hidden layer, can efficiently capture input-output nonlinear relationships.

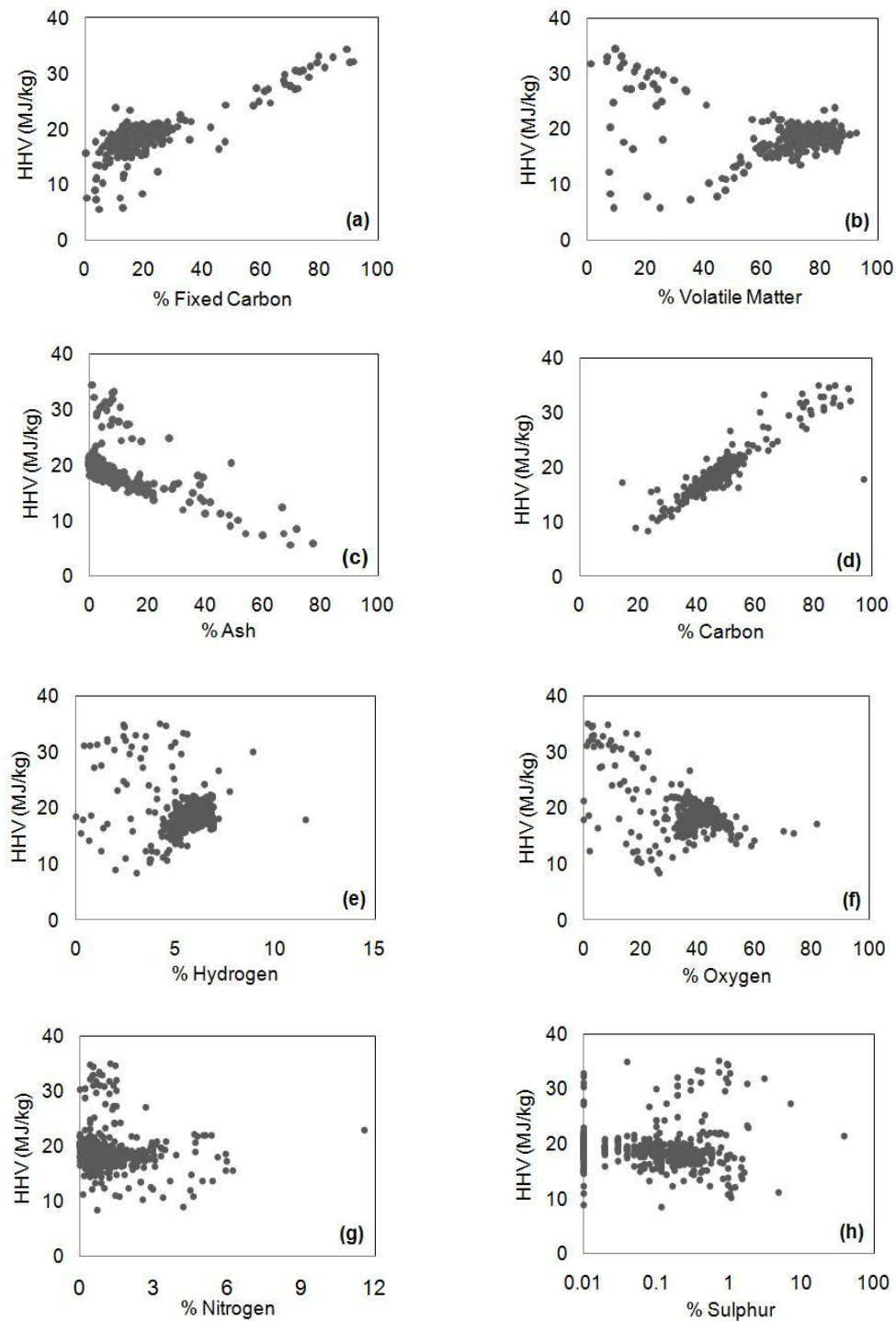


Figure 2.1: Scatter-plots of biomass HHV versus weight percentages of various components of the proximate and ultimate analysis on dry basis.

The relatively newer CI-based symbolic regression performing methodology, namely *genetic programming* (GP) (Koza, 1992; Poli et al., 2008) possesses an added advantage over ANNs (as also the conventional regression methods), that they do not require to pre-specify the form and the parameters of

the fitting function to model the input-output relations (see Chapter 1, Section (1.2.2)). This attractive feature of GPSR (GP-based symbolic regression) has been exploited to develop the nonlinear biomass HHV predicting models possessing high prediction and generalization performance. Additionally, *multi-layer perceptron neural network* (MLPNN) (Freeman and Skapura, 1991; Tambe et al., 1996) has been used to develop the corresponding nonlinear biomass HHV predicting models.

Following are the notable features of this study:

- CI-based methodologies introduced for modeling of *HHV* of solid biomass fuels.
- A wide range of biomasses (refer Section (2.0) and the dataset available online) has been considered for the CI-based model development.
- Large-sized datasets covering a variety of biomasses have been used in the CI-based model development.
- The ultimate analysis dataset used in the CI-based biomass HHV model development has been the largest employed so far when compared with similar studies conducted previously.

The prediction accuracies and generalization performances of the developed CI-based biomass HHV predicting models were rigorously compared with the existing best models and it was found that the CI-based HHV models performed much better than their existing counterparts.

2.2. CI-BASED BIOMASS HHV MODEL DEVELOPMENT

The proximate analysis dataset contains three biomass property attributes namely, *fixed carbon (FC)*, *volatile matter (VM)* and *ash (ASH)*, while the ultimate analysis dataset consists of five attributes namely, *carbon (C)*, *hydrogen (H)*, *oxygen (O)*, *nitrogen (N)* and *sulfur (S)*. For model development the datasets were partitioned into training and test sets in the 85/15 ratio.

2.2.1. GPSR-Based Model Development

The GPSR-based biomass HHV prediction models were developed using *Eureqa Formulize* software package (Schmidt and Lipson, 2009). The set of

operators used in generating candidate solutions are: {+, -, /, *, const, exp, sqrt, power}. Several GPSR-based modeling runs were performed to study the effect of variations in the GPSR parameters on the converged models and the best performing GPSR models were retained for further scrutiny.

To choose the best GPSR-based biomass HHV models, the *CC*, *RMSE* and *MAPE* magnitudes of the predicted biomass HHV values in respect to both the training and test sets were compared. The generalization capabilities of the GPSR-based models were verified by evaluating the referred statistical metrics of the test sets. Finally, the GPSR-based models that gave the best prediction accuracies (in terms of high *CC* values and low *RMSE* and *MAPE* values) for both the training and test sets were retained as the final biomass HHV models. Alongside it was also ensured that the model complexities of the chosen GPSR-based models are low. The retained best GPSR-based biomass HHV (MJ/kg) predicting models are given below:

- Proximate analysis based GPSR model (GPSR-Model 1):

$$HHV = 0.365 * FC + 0.131 * VM + \frac{1.397}{FC} + \frac{328.568 * VM}{10283.138 + 0.531 * (FC)^3 * ASH - 6.863 * (FC)^2 * ASH} \quad (2.5)$$

where, *FC*, *VM* and *ASH* denote weight percentages of fixed carbon, volatile matter and ash respectively on dry basis.

- Ultimate analysis based GPSR model (GPSR-Model 2):

$$HHV = 0.367 * C + \frac{53.883 * O}{2.131 * (C)^2 - 93.299} + \frac{C * H - 115.971}{10.472 * H + 0.129 * C * O} - \frac{91.531}{35.299 + N} + \frac{232.698}{77.545 + S} \quad (2.6)$$

where, *C*, *H*, *O*, *N* and *S* denote the weight percentages of carbon, hydrogen, oxygen, nitrogen and sulfur respectively on dry basis.

2.2.2. MLPNN-Based Model Development

The same biomass dataset as used in constructing the GPSR-based biomass HHV model as was used in the MLPNN-based model development. The two best MLPNN-based biomass HHV models based on the proximate and ultimate analyses were designed and developed using *RapidMiner* software package (Mierswa et al., 2006). Various MLPNN architectures and the EBP training specific parameters were analyzed heuristically and the best performing models were chosen, based on the highest *CCs* and lowest *RMSEs* corresponding to the model prediction of HHV magnitudes. The design of both the MLPNN models consisted of two hidden layers in their architecture and these models were trained using the EBP training algorithm (Rumelhart et al., 1986) available in the package. The dataset was normalized in the [-1, +1] range, prior to MLPNN training, while the predicted output HHV was in the de-normalized form. The architectural design details and the EBP-specific parameters chosen for the development of an optimal MLPNN-based biomass HHV models are given in Table (2.1).

Table 2.1: Architectural details of developed MLPNN-based biomass HHV models

Model No.	Model-inputs	Data-set	Training-set	Test-set	<i>N</i>	<i>L</i>	<i>M</i>
1	<i>FC, VM and ASH</i>	382	322	60	3	5	3
2	<i>C, H, O, N and S</i>	536	456	80	5	6	4

N = number of input layer nodes, *L* = number of nodes in first hidden layer, *M* = number of nodes in second hidden layer.

Both the MLPNN models consisted of log-sigmoid transfer function in the hidden layer nodes and linear transfer function in the output layer nodes.

2.3. RESULTS AND DISCUSSION

A comparison of the HHV prediction accuracy and generalization performance possessed by the developed CI-based models with that of the existing best performing models was carried out on the basis of three statistical metrics viz. *CC*, *RMSE* and *MAPE*.

2.3.1. Biomass HHV Models Based on the Proximate Analysis

The magnitudes of CC , $RMSE$ and $MAPE$ in respect of the CI-based biomass HHV models as also the best performing biomass HHV model of Parikh et al. (2005) (Eqn. 2.1) are listed in Table (2.2), wherein it is observed that both the CI-based models (GPSR-Model 1 and MLPNN-Model 1) have performed better as compared to the biomass HHV model of Parikh et al. (2005). Especially, it is noticed that the CC values for the HHV predictions for the training (>0.96) and test (>0.95) sets by the CI-based models are higher (0.955, 0.954) than that for the predictions by the model by Parikh et al. (2005). Also, the $RMSE$ and $MAPE$ magnitudes in respect of the training (test) set HHV predictions by the CI-based models are significantly lower than the corresponding $RMSE$ and $MAPE$ magnitudes of the model by Parikh et al. (2005).

Table 2.2: Comparison of biomass HHV prediction performances of GPSR-Model 1, MLPNN-Model 1 and Parikh et al's model based on the proximate analysis as input

Model	Training-set			Test-set		
	CC_{Trn}	$RMSE_{Trn}$	$MAPE_{Trn}$	CC_{Tst}	$RMSE_{Tst}$	$MAPE_{Tst}$
GPSR-Model 1	0.97	0.94	4.01	0.96	0.97	4.37
MLPNN-Model 1	0.97	0.93	3.80	0.96	0.99	4.43
Parikh et al. (2005)	0.95	1.25	5.20	0.95	1.12	4.45

Similar observations can be inferred from the parity plots in Figure (2.2), wherein it can be observed that the plots for the HHV predictions of the CI-based models show a much lower scatter (indicating higher prediction accuracies) as compared to the corresponding plot of predictions made by the model of Parikh et al. (2005).

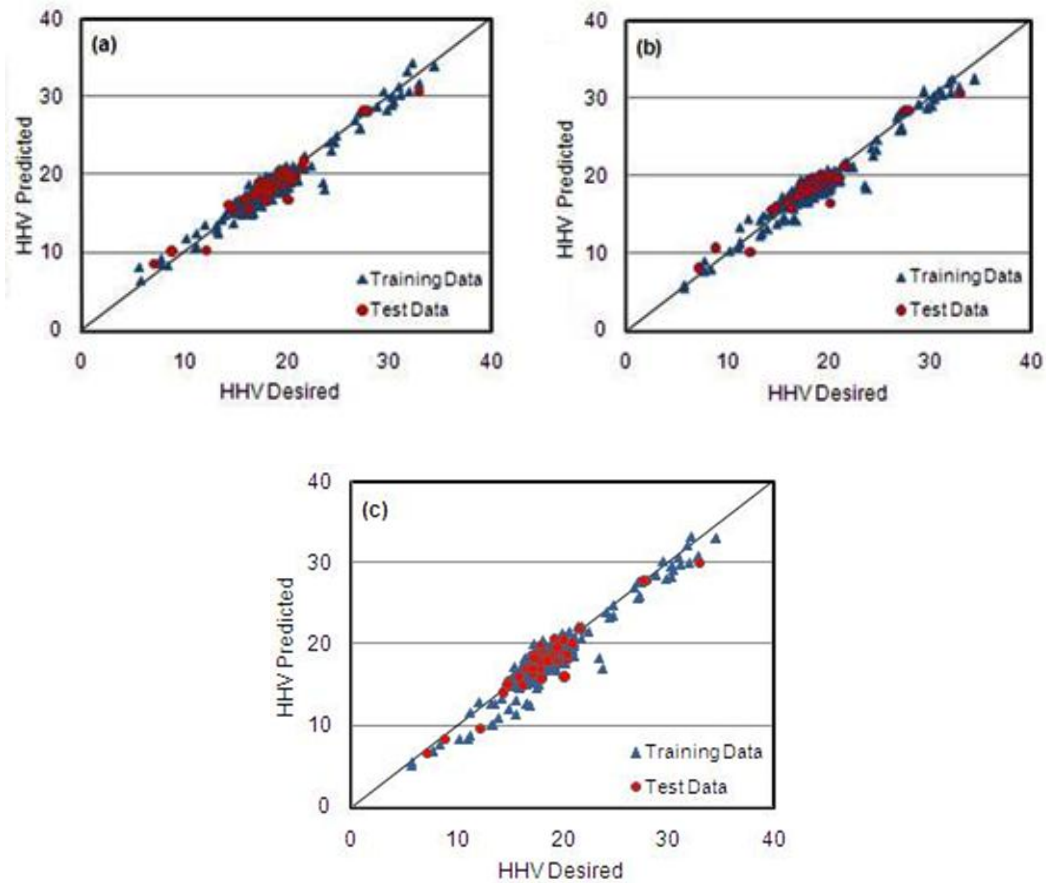


Figure 2.2: Parity plots of the experimental HHVs versus those predicted by the proximate analysis based following models: (a) GPSR-Model 1, (b) MLPNN-Model 1 and (c) Parikh et al. (2005).

A comparison of the HHV prediction accuracies and generalization performances of the CI-based models with the competing linear models was also performed to verify their prediction and generalization performance in different (low, medium and high) HHV ranges. The results of this comparison are given in Table (2.3). The three experimental HHV (MJ/kg) ranges considered are (were the model predicted HHVs were compared with the experimental HHV values) are: 0-16 (low), 16-25 (medium), and 25-35 (high).

Table 2.3: Range-wise comparison of biomass HHV prediction accuracies and generalization performances of proximate analysis based models

Model	HHV (MJ/kg) range								
	0-16			16-25			25-35		
	<i>CC</i>	<i>RMSE</i>	<i>MAPE</i>	<i>CC</i>	<i>RMSE</i>	<i>MAPE</i>	<i>CC</i>	<i>RMSE</i>	<i>MAPE</i>
GPSR-Model 1	0.95	1.03	7.87	0.76	0.92	3.66	0.89	1.05	2.98
MLPNN-Model 1	0.95	1.11	7.50	0.76	0.90	3.50	0.89	1.10	3.04
Parikh et al. (2005)	0.92	1.54	8.98	0.69	1.18	4.74	0.89	1.19	3.11

The results of the mentioned range-wise comparison are as follows:

- In the low HHV range, the prediction and generalization performance of the CI-based models is better than that of the competing linear model (Eqn. 2.1).
- In the medium HHV range, the CI-based HHV models as well as the competing linear model possess lower values of *CCs* as compared to those observed for the low HHV range. However, as inferred from the *CC*, *RMSE* and *MAPE* magnitudes, the HHV prediction accuracy and generalization performance of both the CI-based models is superior than the competing linear models.
- In the high HHV range, all biomass HHV prediction models depicted comparable performance, with the CI-based models performing marginally better with lower values of *RMSEs* and *MAPEs* as compared to the linear competing model.

2.3.2. Biomass HHV Models Based on the Ultimate Analysis

Here, the CI-based biomass HHV models were compared using the same statistical metrics (*CC*, *RMSE* and *MAPE*) with the three competing models proposed by Channiwala and Parikh (2002) (Eqn. 2.2), Milne et al. (1990) (Eqn. 2.3) and Friedl et al. (2005) (Eqn. 2.4). The results of these comparisons are listed in Table (2.4); it is observed in this table that the *CC* values relative to the experimental HHVs in training and test sets and the corresponding magnitudes predicted by the CI-based models are higher than those computed by the three

competing models. Also, *RMSE* and *MAPE* values pertaining to the HHV predictions by the CI-based models are lower when compared with the corresponding HHV predictions by the three non-CI based empirical models. The results are clearly indicative of the fact that the CI-based models using elements of the ultimate analysis as inputs for the prediction of biomass HHV exhibit better prediction accuracies and generalization performance than their competing non-CI based models. Figure (2.3) also supports these conclusions by revealing a lower scatter in the parity plots corresponding to the HHV predictions by the CI-based models than that existing in the plots related to the predictions of the non-CI models.

Table 2.4: Comparison of biomass HHV prediction accuracies and generalization performances of the GPSR-Model 2 and MLPNN-Model 2 with the models of Channiwala and Parikh (2002), Milne et al. (1990) and Friedl et al. (2005) based on ultimate analysis

Model	Training-set			Test-set		
	CC_{Trn}	$RMSE_{Trn}$	$MAPE_{Trn}$	CC_{Tst}	$RMSE_{Tst}$	$MAPE_{Tst}$
GPSR-Model 2	0.95	1.08	3.62	0.95	0.94	3.63
MLPNN-Model 2	0.97	0.86	3.33	0.95	0.98	3.67
Channiwala and Parikh (2002)	0.89	1.88	4.99	0.93	1.32	4.86
Milne et al. (1990)	0.88	1.97	5.35	0.92	1.45	5.25
Friedl et al. (2005)	0.93	1.38	4.50	0.95	1.03	4.01

As conducted for the proximate analysis based biomass HHV prediction models, the performance of the ultimate analysis based models was also compared in three ranges of HHVs. Here, the prediction and generalization potential of the CI-based models was compared with their three competing models, viz. Channiwala and Parikh (2002) (Eqn. 2.2), Milne et al. (1990) (Eqn. 2.3) and Friedl et al. (2005)

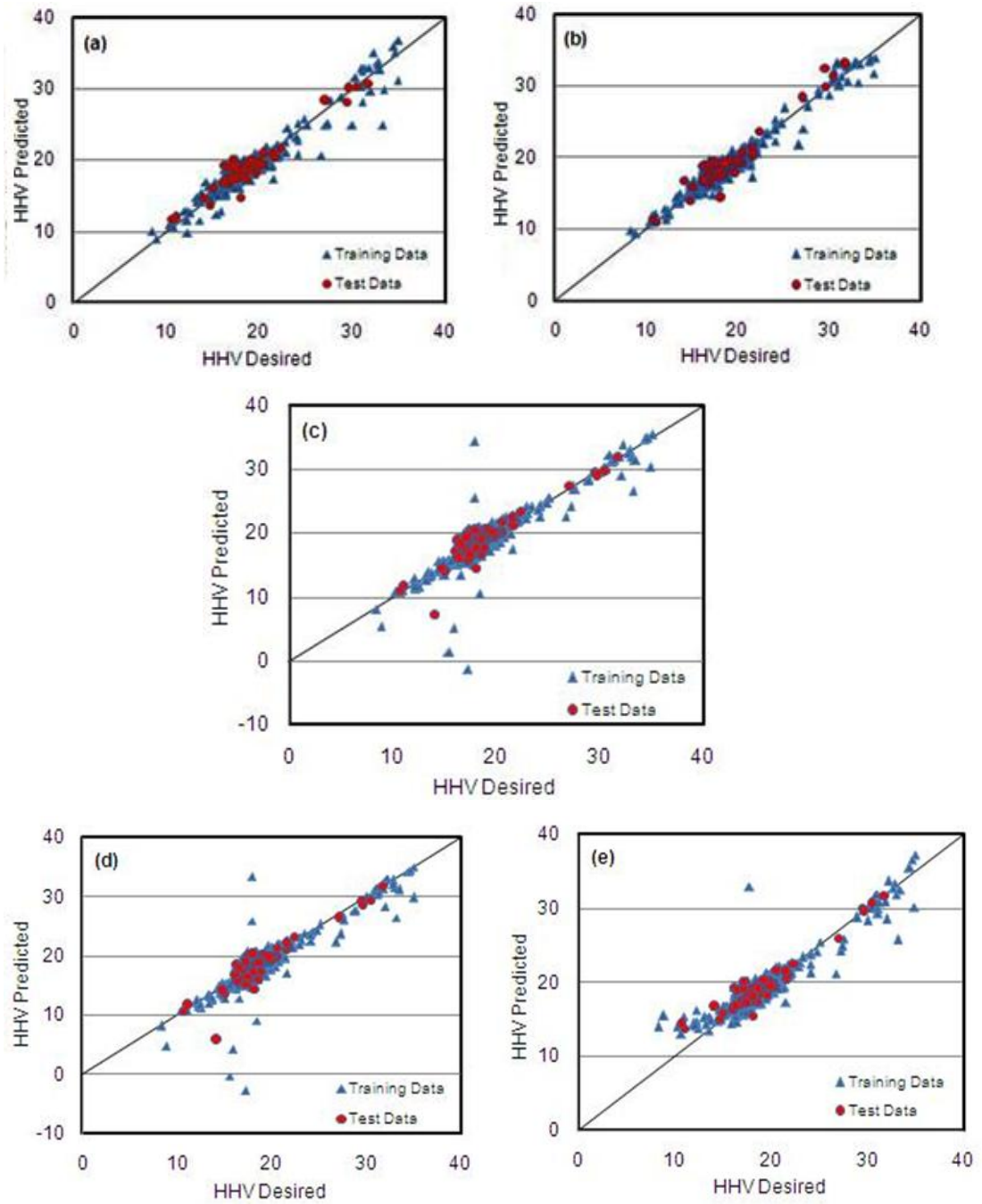


Figure 2.3: Parity plots of the experimental HHVs versus those predicted by the ultimate analysis based following models: (a) GPSR-Model 2, (b) MLPNN-Model 2, (c) Channiwala and Parikh (2002), (d) Milne et al. (1990) and (e) Friedl et al. (2005).

(Eqn. 2.4). Table (2.5) lists the results of this comparison in terms of the three statistical metrics, viz. *CC*, *RMSE* and *MAPE*, and the key findings thereof are given below.

- The CI-based models show significantly higher prediction accuracy and lower generalization error (also termed “out-of-sample” error) in the low HHV range than the competing models that are not CI-based. The largest improvement is observed while comparing MLPNN-Model 2 and the model of Milne et al. (1990), with the former outperforming the latter model’s *CC* by 100%.
- In the medium HHV range, both the CI-based models exhibit better prediction and generalization performance when compared with that of the non-CI models, with both CI models exhibiting closely matching HHV prediction performance.
- In the high HHV range (25–36 MJ/kg), both the MLPNN-Model 2, and the competing presently high performing models show comparable HHV prediction and generalization capabilities with the GPSR-Model 2 performing sub-optimally when compared with the rest of the models.

Table 2.5: Range-wise statistical analysis of HHV prediction accuracy and generalization performance of ultimate analysis based models

Model	HHV (MJ/kg) range								
	0-16			16-25			25-36		
	<i>CC</i>	<i>RMSE</i>	<i>MAPE</i>	<i>CC</i>	<i>RMSE</i>	<i>MAPE</i>	<i>CC</i>	<i>RMSE</i>	<i>MAPE</i>
GPSR-Model 2	0.84	1.16	6.38	0.82	0.84	3.15	0.75	2.55	5.77
MLPNN-Model 2	0.90	0.90	5.21	0.84	0.81	3.13	0.83	1.55	3.89
Channiwala and Parikh (2002)	0.50	2.78	7.94	0.63	1.66	4.76	0.82	1.83	3.38
Milne et al. (1990)	0.45	3.10	9.21	0.63	1.71	5.00	0.83	1.99	4.04
Friedl et al. (2005)	0.64	2.13	13.17	0.73	1.10	3.41	0.81	2.22	4.75

The MLPNN models in this study employed the sigmoid transfer function to compute the outputs of the nodes in their hidden layers. Thus, these models are inherently nonlinear. Also, the GPSR-based biomass HHV models (Eqn. (2.5) and (2.6)) are nonlinear. The fact that all CI-based nonlinear models possess an improved biomass HHV prediction performance as compared to the existing best

competing linear models, clearly suggests that for the prediction of HHV of biomass fuels using the constituents of proximate or ultimate analyses, nonlinear models are indeed more suitable than the linear models.

2.4. CONCLUSION

An extensive literature survey of biomass HHV predicting models revealed that a majority of the existing models are based on the conventional empirical regression approaches. Also, most of the high-performing biomass HHV models based on the components of the proximate and/or ultimate analyses are pre-dominantly linear. A thorough analysis of the proximate and ultimate analyses and the corresponding HHV values reveals that the HHVs are nonlinearly related to a few of the proximate/ultimate analysis components. Accordingly, two computational intelligence (CI) methodologies namely genetic programming assisted symbolic regression (GPSR) and multi-layer perceptron neural network (MLPNN) were utilized to design and develop models predicting HHV of solid biomass fuels. A total of four CI-based models, with two GPSR-based and two MLPNN-based biomass HHV models were developed, which respectively used the components of the proximate and ultimate analyses, as their inputs. The GPSR methodology has a major attractive feature that given the example dataset, it is capable of generating optimized linear/nonlinear fitting function and associated parameters that best fit the data. In spite of this attractive feature, GP based symbolic regression has rarely been explored for modeling in fuel science/engineering. The prediction and generalization performance of both the CI-based models was compared with the competing high-performance models. The said comparison revealed that all CI-based models have an excellent HHV prediction performance as compared to the competing conventional models. In the conducted exercise of comparison of the models' prediction ability in three specific ranges of HHV (i.e., low, medium and high), the GPSR-Model 2 with ultimate analysis constituents as inputs, has yielded excellent performance in the low and medium HHV ranges; it, however, fared sub-optimally in the high HHV range (25-36 MJ/kg). The CI (GPSR and MLPNN) based nonlinear biomass HHV models developed in this study since possess an excellent prediction and

generalization performance have the potential to replace the existing biomass HHV correlations.

NOMENCLATURE

- L Number of first hidden layer nodes in MLPNN
- M Number of second hidden layer nodes in MLPNN
- N Number of input layer nodes in MLPNN

REFERENCES

- Boundy, B., Diegel, S. W., Wright, L., Davis, S. C. (2011). Biomass resources overview in biomass energy data book. 4th edn., U.S. Department of Energy, <http://cta.ornl.gov/bedb>. Accessed: 06 April 2013
- Channiwala, S. A., Parikh, P. P. (2002). A unified correlation for estimating HHV of solid, liquid and gaseous fuels. *Fuel* 81:1051–1063.
- Ciolkosz, D. (2010). Characteristics of biomass as a heating fuel. Document Code #UB043, Pennsylvania State University, pp 1–4. <http://extension.psu.edu/pubs/ub043>. Accessed: 10 April 2013
- Cordero, T., Marquez, F., Rodriguez-Mirasol, J., Rodriguez, J. (2001). Predicting heating values of lignocellulosics and carbonaceous materials from proximate analysis. *Fuel* 80:1567–1571.
- Daya, R. N., Abdul, S. P. (2012). Estimation of higher heating value of biomass from proximate analysis: a new approach. *Fuel* 99:55–63.
- ECN Phyllis, (2012). The composition of biomass and waste. <http://ecn.nl/phyllis/single.html>. Accessed 09 April 2013
- Freeman, J. A., Skapura, D. M. (1991). *Neural Networks: Algorithms, Applications, and Programming Techniques*. Addison-Wesley Reading.

- Friedl, A., Padouvas, E., Rotter, H., Varmuza, K. (2005). Prediction of heating values of biomass fuel from elemental composition. *Analytica Chimica Acta* 544:191–198.
- Hofbauer, H., BIOBIB, (2012). A database for biofuels. Institute of Chemical Engineering, Vienna University of Technology, Austria, Vienna. <http://vt.tuwien.ac.at/biobib/info.html>. Accessed: 09 April 2013
- IEA Bioenergy Task, 32. (2013). Biomass combustion and cofiring. <http://ieabcc.nl/database/biomass.php>. Accessed: 09 April 2013
- Jimenez, L., Gonzales, F. (1991). Study of the physical and chemical properties of lignocellulosic residues with a view to the production of fuels. *Fuel* 70:947–950.
- Koza, J. R. (1992). Genetic programming: on the programming of computers by means of natural selection. MIT, Cambridge.
- Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., Euler, T. (2006). YALE: rapid prototyping for complex data mining tasks. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-06).
- Milne, T. A., Brennan, A. H., Glenn, B. H. (1990). Source book of methods of analysis for biomass and biomass conversion processes. Elsevier, London
- Parikh, J., Channiwala, S. A., Ghosal, G. K. (2005). A correlation for calculating HHV from proximate analysis of solid fuels. *Fuel* 84: 487–494.
- Poli, R., Langdon, W. B., McPhee, N. F. (2008). A field guide to genetic programming. Published via <http://lulu.com/> and freely available at <http://gp-field-guide.org.uk/>; (With contributions by J. R. Koza). Accessed: 22 January 2013
- Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature* 323:533–536.
- Schenk, P. M., Thomas-Hall, S. R., Stephens, E., Marx, U. C., Mussnug, J. H.,

- Posten, C., Kruse, O., Hankamer, B. (2008). Second generation biofuels: high-efficiency microalgae for biodiesel production. *Bioenergy Research* 1:20–43.
- Schmidt, M., Lipson, H. (2009). Distilling free-form natural laws from experimental data. *Science* 324(5923):81–85.
- Sheng, C., Azevedo, J. L. T. (2005). Estimating the higher heating value of biomass fuels from basic analysis data. *Biomass and Bioenergy* 28(5): 499–507.
- Tahir, M. H. N., Casler, M. D., Moore, K. J., Brummer, E. C. (2011). Biomass yield and quality of reed canary grass under five harvest management systems for bioenergy production. *Bioenergy Research* 4:111–119.
- Tambe, S. S., Kulkarni, B. D., Deshpande, P. B. (1996). Elements of artificial neural networks with selected applications in chemical engineering, and chemical & biological sciences. *Simulation & Advanced Controls*, Louisville
- Yin, C. Y. (2011). Prediction of higher heating values of biomass from proximate and ultimate analysis. *Fuel* 90:1128–1132.

CHAPTER 3.

PREDICTION OF ELEMENTAL COMPOSITION OF SOLID BIOMASS FUELS FROM PROXIMATE ANALYSIS USING COMPUTATIONAL INTELLIGENCE BASED MODELS

PREDICTION OF ELEMENTAL COMPOSITION OF SOLID BIOMASS FUELS FROM PROXIMATE ANALYSIS USING COMPUTATIONAL INTELLIGENCE BASED MODELS

Abstract

The elemental composition of a biomass fuel is an important measure of its energy content and its information is valuable in designing of combustion processes for efficient and clean utilization of the fuel. Obtaining the elemental composition (termed “*ultimate analysis*”) via standard laboratory procedures is expensive, time-consuming, and requires a skilled analyst. The proximate analysis of a fuel represents its *ash*, *fixed carbon* and *volatile matter* content and relatively easy to conduct. Currently, two linear models are available for the prediction of elemental composition of solid biomass fuels from their proximate analysis. These show applicability over a limited range of percentage compositions of the elements and the constituents of proximate analysis of the fuel. Accordingly, this study utilizes three computational intelligence (CI) methodologies, namely *genetic programming* (GP), *artificial neural network* (ANN), and *support vector regression* (SVR) for modeling the elemental composition of a large variety of biomass fuels using the constituents of their proximate analysis as inputs. The models developed are applicable over wider ranges of the constituents of the proximate analysis, and provide much better prediction and generalization performance as compared with the existing linear models.

3.0. INTRODUCTION

Biomass fuels, due to their renewable nature are seen as a promising “Green” source of energy for the future. Considering the tremendous importance being attached to the sustainable energy, the usage of biomass in energy production is expected to rise exponentially. Biomass fuels are obtained from the remains of plants and animals such as dried leaves, stems, dried plant-fruits and also various forms of fossil fuels originating from the long-term natural processing of buried remains of animals and plants. The *ultimate analysis* of a biomass fuel expresses its composition on an elemental basis representing the weight percentages of *ash*, *carbon*, *hydrogen*, *nitrogen*, *sulfur*, and (by difference) *oxygen*. The major components of the ultimate analysis include carbon (*C*), hydrogen (*H*) and oxygen (*O*), while nitrogen and sulfur are the minor elements. The knowledge of the ultimate analysis of a biomass fuel is helpful in the efficient design and operation of the biomass fuel utilizing facilities (Parikh et al., 2007). It is directly related to the energy content of the fuel and plays an important role in:

- (a) The mass and energy balance calculations of the biomass-utilizing processes such as gasification, combustion and pyrolysis.
- (b) The calculation of the stoichiometric requirement of air for the complete combustion of the biomass fuel for achieving enhanced combustion efficiency.
- (c) The optimal design of combustion equipment for maintaining appropriate biomass fuel-to-air ratio.

A relatively cruder characterization of a biomass fuel is performed in terms of its *proximate analysis*, which constitutes weight percentages of *fixed carbon (FC)*, *volatile matter (VM)*, *ash (ASH)*, and *moisture*. The standard analytical ASTM procedures are available for the sample preparation and measurement of components of the proximate and ultimate analyses of solid fuels (ASTM D5142-04, 2004; ASTM D3176-09, 2009).

The experimental determination of the elemental composition of a fuel requires very expensive equipment and skilled analysts (Cordero et al., 2001), while the proximate analysis can be easily done using standard laboratory

equipments by any engineer/researcher (Parikh et al., 2007). Thus, correlations predicting the elemental composition of a biomass fuel from the constituents of its proximate analysis would be a convenient and cost-effective alternative to ultimate analysis. From an extensive literature survey, it is observed that there are only two such models by namely Parikh et al. (2007) and Shen et al. (2010). Both the models are linear and the average absolute error of their predictions of *C*, *H* and *O* (in weight percentages) are greater than 3%, 4%, and 3%, respectively. Also, the applicability of these conventional models is limited to lower ranges of the inputs (weight percentages of the components of the proximate analysis). In order to develop models with improved performance in predicting the elemental composition of biomass fuels and applicability over wider ranges of the inputs, a large and diverse biomass property dataset was compiled. This dataset consists of the proximate analysis and the corresponding elemental composition records of 830 biomass samples of a variety of biomasses. Using the dataset, nonlinear models were developed using CI-based *genetic programming* (GP) (Koza, 1992; Poli et al., 2008), *multilayer perceptron neural network* (MLPNN) (Freeman and Skapura, 1991; Bishop, 1995) and *support vector regression* (SVR) (Vapnik, 1995; Ivanciuc, 2007) methodologies. Notably, the usage of the relatively new and infrequently used GP based symbolic regression (GPSR) method resulted in simpler correlations that are easier to use in practice (see Section (1.2.2) for more details of GPSR methodology). In this study, a total of nine CI-based models (three each of GPSR, MLPNN and SVR) have been developed for the prediction of weight percentages of carbon, hydrogen and oxygen (all on dry basis), respectively of biomass fuels using the components of the corresponding proximate analysis as inputs.

Thus, the notable features of this study are:

- CI-based modeling has been introduced for the prediction of elemental composition of solid biomass fuels.
- The dataset compiled for the development of the said CI-based models has been the largest so far utilized in the development of similar type of models.

- The compiled dataset of the biomass samples includes a wide variety of biomass types such as waste sludge, paper and pulp waste, processing wastes, various grades of coal, chars, plants, crops and forest trees.
- The CI-based models developed have wider applicability in terms of their ranges of the model inputs as compared to the competing linear non-CI-based models (see Table (3.1)).

Table 3.1: Details of the proximate and ultimate analyses dataset of biomass fuels

Component	Training-set (620)*		Test-set (125)		Validation-set (85)	
	Range	Mean	Range	Mean	Range	Mean
<i>Proximate analysis</i>						
Fixed Carbon	5.04-89.6	20.73	6.43-87.17	20.70	10.8-82.2	22.47
Volatile Matter	7.4-94.73	70.43	9.93-87.9	73.81	11.6-85.65	73.98
Ash	0.1-56.1	8.81	0.18-47.5	5.48	0.29-25.5	3.53
<i>Ultimate analysis</i>						
Carbon	22.35-92.86	47.98	24.4-91.85	49.49	36.81-81.7	50.77
Hydrogen	0.52-11.42	5.60	0.73-6.37	5.47	1.33-6.75	5.60
Oxygen	2.32-52.06	36.17	4.25-44.99	38.61	5.55-45.13	39.41

*indicates number of data-points in the set; all values are in weight percentage on dry basis

The prediction performances of the CI-based models has been compared with the existing, linear models of Parikh et al. (2007), and Shen et al. (2010). The rest of this chapter is organized as follows. In the forthcoming sections, the methodologies used to develop the respective CI-based models are discussed. In the “Results and Discussion” section, a comparison of the prediction performance of the CI-based models and competing linear models is done. For identifying the best models for the prediction of the elemental composition of solid biomass fuels, standard statistical tests have been performed, the results of which are also discussed in the same section. Finally, the “Conclusion” section provides the concluding remarks.

3.0.1. Need for Non-Linear Elemental Composition Models

Presently, there are only two linear regression-based models available for the prediction of the elemental composition of solid biomass fuels from the

components of their proximate analysis. These correlations proposed by Parikh et al. (2007) (Eqns. 3.1–3.3) and Shen et al. (2010) (Eqns. 3.4–3.6), are given below:

$$C = 0.637FC + 0.455VM \quad (3.1)$$

$$H = 0.052FC + 0.062VM \quad (3.2)$$

$$O = 0.304FC + 0.476VM \quad (3.3)$$

$$C = 0.635FC + 0.460VM - 0.095ASH \quad (3.4)$$

$$H = 0.059FC + 0.060VM + 0.010ASH \quad (3.5)$$

$$O = 0.340FC + 0.469VM - 0.023ASH \quad (3.6)$$

where, FC = fixed carbon, VM = volatile matter, ASH = ash, C = carbon, H = hydrogen and O = oxygen are the weight percentages (on dry basis) of the respective components of solid biomass fuels. In order to examine the nature of the relationships (whether linear or nonlinear) existing between the individual biomass composition elements and the corresponding components of the proximate analysis, scatter-plots were generated by plotting the individual components of the proximate analysis against each component of the elemental composition as shown in Figure (3.1).

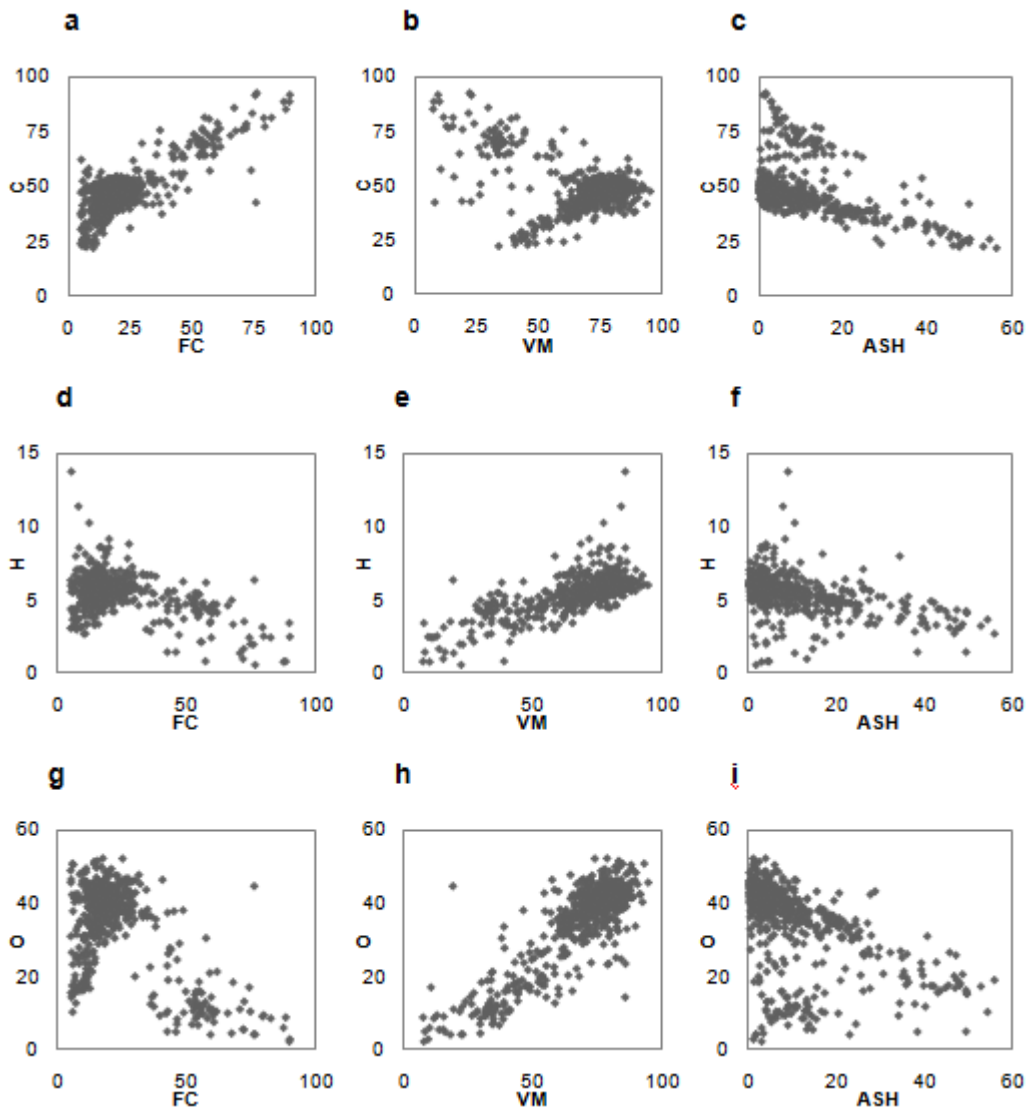


Figure 3.1: Scatter-plots of components of the elemental composition versus components of the proximate analysis of biomass fuels; C: wt% carbon, H: wt% hydrogen, O: wt% oxygen, FC: wt% fixed carbon, VM: wt% volatile matter and ASH: wt% ash in biomass fuels.

From the scatter-plots, it is observed that the relations between some of the components of the proximate analysis and the corresponding elemental composition of biomass fuels are nonlinear. For instance, a strong nonlinear dependence is witnessed (see Figure 3.1(a)) between the C and FC values in the first half of their ranges while the variables, C and ASH show a weak nonlinear interdependency. The hydrogen weight percentage also shows a strong nonlinear relation with FC , and weak nonlinear relationships with VM and ASH (see Figures 3.1(d), (e) and (f)). Oxygen in the biomass too shows a mild nonlinear correlation with fixed carbon (see Figure 3.1(g)). From these observations, it can be unambiguously inferred that the models considering the above-stated nonlinear

dependencies are most likely to display a better prediction and generalization performance in predicting the weight percentages of the individual constituents of the elemental composition (C , H and O) of biomass fuels from the components (FC , VM and ASH) of their proximate analysis.

3.1. CI-BASED MODELS FOR THE PREDICTION OF ELEMENTAL COMPOSITION OF BIOMASS FUELS

The CI-based (GPSR, MLPNN and SVR) elemental composition predicting models for solid biomass fuels were developed using proximate analysis data and the corresponding elemental composition values of a wide variety of biomass fuels. A majority of the dataset was compiled from the open access database, i.e., ECN's Phyllis database (ECN Phyllis, 2012) and a few data were obtained from the published research papers, totaling 830 data-points. The compiled dataset is available on-line at the following address: https://static-content.springer.com/esm/art%3A10.1007%2Fs13198-014-0324-4/MediaObjects/13198_2014_324_MOESM1_ESM.doc.

For the development of CI-based models, the dataset was partitioned into *training*, *test* and *validation* sets in percentages 75%, 15% and 10%, respectively; the statistical analysis of the data are provided in Table (3.1). The dataset was pre-processed by normalizing (except for the development of the SVR based models) the inputs (proximate analysis constituents) in the range $[-1, +1]$. The outputs (targets) were supplied to the GPSR and MLPNN based learners in normalized form in the same range as the inputs. The normalized outputs obtained from these models were de-normalized to generate the rescaled outputs. The performances of each of the developed CI-based models were evaluated in terms of three statistical metrics, namely, *coefficient of correlation (CC)*, *root mean square error (RMSE)*, and *mean absolute percent error (MAPE)* relative to the experimental and model predicted weight percentage values of C , H and O . These predicted outputs from the constructed CI-based models were compared with those computed using the competing linear models of Parikh et al. (2007) and Shen et al. (2010). Finally, to identify the best performing CI-based model from among the three GPSR, MLPNN and SVR based models for predicting the weight percentages of C , H and

O individually, a statistical test known as the Steiger's z-test (Steiger, 1980) was conducted. This test using coefficient of correlation (CC) values of competing models unambiguously identifies the best model (see Section (1.8) for a detailed description of the test).

3.1.1. Development of GPSR-Based Biomass Elemental Composition Predicting Models

The *Eureka Formulize* (Schmidt and Lipson, 2009) software package was used to develop the GPSR-based models. This package has several attractive features (see Section (1.2.2)) for developing models possessing an excellent generalization capability. Several runs were conducted, by systematically varying GPSR parameters thus obtaining multiple expressions with different structures. The identification of the best GPSR-based biomass elemental composition models was done based on the lower complexity of the expression and its higher prediction and generalization performance. The best GPSR-based models satisfying the stated criteria for the prediction of carbon, hydrogen and oxygen weight percentages are given below:

(a) Carbon content (wt%) predicting optimal model (GPSR-Model 1):

$$C = 35.255(0.1834 + 1.271\hat{x}_1 + 0.3891\hat{x}_2 + 0.2564\hat{x}_1^2 - 0.2873\hat{x}_2^3\hat{x}_3) + 57.605 \quad (3.7)$$

(b) Hydrogen content (wt%) predicting optimal model (GPSR-Model 2):

$$H = \frac{5.45(-\hat{x}_1 - \hat{x}_3)}{(2.409 + \hat{x}_2)} + 3.062425 \quad (3.8)$$

(c) Oxygen content (wt%) predicting optimal model (GPSR-Model 3):

$$O = 24.87(0.2451 + 1.495\hat{x}_2 + 0.3608\hat{x}_3 + 0.8407\hat{x}_1^3) + 27.19 \quad (3.9)$$

where, \hat{x}_1 , \hat{x}_2 and \hat{x}_3 are the normalized variables of percentages of fixed carbon (FC), volatile matter (VM) and ash (ASH) in the biomass on dry-basis respectively, expressed as,

$$\hat{x}_1 = \frac{(2FC - 94.64)}{84.56} \quad (3.10)$$

$$\hat{x}_2 = \frac{(2VM - 102.13)}{87.33} \quad (3.11)$$

$$\hat{x}_3 = \frac{(2ASH - 56.2)}{56} \quad (3.12)$$

3.1.2. Development of MLPNN-Based Biomass Elemental Composition Predicting Models

Three MLPNN models for the prediction of weight percentages of carbon, hydrogen and oxygen, respectively were developed using *Rapid Miner* software package (Mierswa et al., 2006). Each of the optimal MLPNN models have two hidden layers in its architecture, the details of which are given in the Table (3.2). The table also lists the parameter values of the Error-Back-Propagation (EBP) algorithm used in training the three optimally performing MLPNN models.

Table 3.2: Dataset partitioning and architectural details of the developed MLPNN-based biomass elemental composition models

Model No.	Model inputs	Data-set			N	L	M	η	μ
		Training -set	Test -set	Validation -set					
1	<i>FC, VM and ASH</i>	620	125	85	3	5	3	0.2	0.35
2	<i>FC, VM and ASH</i>	620	125	85	3	5	3	0.4	0.15
3	<i>FC, VM and ASH</i>	620	125	85	3	5	3	0.2	0.3

N = number of nodes in the input layer, L = number of nodes in the 1st hidden layer, M = number of nodes in the 2nd hidden layer, TF = Transfer function, η = Learning rate, μ = Momentum coefficient.

3.1.3. Development of SVR-Based Biomass Elemental Composition Predicting Models

The optimal SVR-based models for the prediction of weight percentages of C , H and O , respectively were also developed using *RapidMiner* software package

(Mierswa et al., 2006). Table (3.3) shows the details of these SVR-based models along with the parameter values pertaining to the “ANOVA” kernel function used in the SVR formulation. The epsilon-insensitive loss function was used in the SVR model formulation. The SVR parameters as listed in the table were chosen heuristically to obtain the optimal SVR models.

Table 3.3: Dataset partitioning and parameter details of the developed SVR-based biomass elemental composition models

Model No.	Model inputs	Data-set			γ	ε	SVs
		Training -set	Test-set	Validation-set			
1	<i>FC, VM and ASH</i>	620	125	85	0.1	5	104
2	<i>FC, VM and ASH</i>	620	125	85	0.05	0.7	116
3	<i>FC, VM and ASH</i>	620	125	85	0.1	5	126

γ = Gamma (Kernel parameter), ε = Epsilon (tube-width), SVs = Number of support vectors; Kernel used: Radial Basis Function (RBF)

3.2. RESULTS AND DISCUSSION

The prediction and generalization performances of the CI-based models in predicting the weight (%) magnitudes (on dry basis) of *C*, *H* and *O*, was rigorously compared with that exhibited by the linear models of Parikh et al. (2007) and Shen et al. (2010). The latter two models are valid over the following narrow ranges of the proximate and ultimate analyses constituents (on dry basis) of the biomass samples.

- *FC*: 4.7-38.4 wt%, *VM*: 57.2-90.6 wt%, *ASH*: 0.1-24.6wt%, *C*: 36.2-53.1 wt%, *H*: 4.36-8.3 wt% and *O*: 31.37-49.5 wt%.

In contrast, the present modeling study utilizes a large number of biomass samples that are valid over the following wider ranges of the components of the proximate and ultimate analyses.

- *FC*: 5.04-89.6 wt%, *VM*: 7.4-94.73 wt%, *ASH*: 0.1-56.1 wt%, *C*: 22.35-92.86 wt%, *H*: 0.52-11.42 wt% and *O*: 2.32-52.06 wt%.

Specifically, in a dataset of 830 biomass samples considered in this study, a close to 200 samples lie outside the ranges considered by the models proposed by

Parikh et al. (2007) and Shen et al. (2010). Accordingly, we have evaluated performance of each CI-based model in two ranges that is, “within” and “outside” the ranges considered by the models of Parikh et al. (2007) and Shen et al. (2010). This performance assessment is conducted using 614 (“within range”) and 216 (“outside range”) number of sample data. Finally, the CI-based models were compared with each other using Steiger’s test (Steiger, 1980) for testing the hypothesis that two correlation coefficient magnitudes relative to the predictions of a pair of models are statistically equal. The null hypotheses ($H_0: CC_{AB} = CC_{AC}$) is rejected if ‘ p ’ value is lesser than 0.05, else it is accepted. Here subscript, ‘A’ denotes the experimental values, while ‘B’ and ‘C’ indicate output values predicted by models ‘B’ and ‘C’, respectively.

3.2.1. Comparison of Models Predicting Carbon (wt%)

Table (3.4) shows a comparison of the prediction and generalization performance of the three CI-based and the two existing linear models for carbon (wt%). From the table listing CC , $RMSE$ and $MAPE$ values, it is evident that, the prediction and generalization performance of the CI-based models is clearly much better than the linear models. This can be seen in terms of higher CC s (>0.9) corresponding to the predictions of almost all CI-based models when compared with the CC magnitudes (<0.85) returned by the linear models’ predictions of carbon (wt%). The major improvement by the CI-based models over the existing linear models is seen through the much lower $RMSE$ s (nearly two times reduction, as compared to the existing linear models), and $MAPE$ s. A range-wise comparison of the models as seen in Table (3.5), reveals that the CI-based models developed here show marginally better performance than the linear models for *within* range data and much better performance for “out-of- range” data. The result of the Steiger’s statistical test (see Table (3.6)) for testing of the null hypothesis of equivalence of correlation coefficients indicates that the MLPNN and SVR models fair comparably well, while when compared with the GPSR-based model they fair better. Figure (3.2) shows the parity plots of comparison of the prediction of the carbon (wt%) values by the CI-based and the existing best linear model of Parikh et al. (2007). A larger scatter in the parity-plot for the carbon (wt%) predictions pertaining to the model of Parikh et al. (2007) is observed as compared to the plots of predictions done by the CI-based models.

Table 3.4: Comparison of model performances of CI-based and existing linear-regression models for the prediction of carbon (wt%)

Model	Training-set			Test-set			Validation-set		
	CC_{Trn}	$RMSE_{Trn}$	$MAPE_{Trn}$	CC_{Tst}	$RMSE_{Tst}$	$MAPE_{Tst}$	CC_{Vln}	$RMSE_{Vln}$	$MAPE_{Vln}$
GPSR-Model 1	0.89	4.24	6.38	0.98	1.37	1.82	0.98	1.91	2.42
MLPNN- Model 1	0.91	4.20	6.59	0.98	2.10	3.30	0.98	1.72	2.50
SVR-Model 1	0.91	3.96	6.17	0.98	1.64	2.56	0.99	1.28	1.81
Parikh et al. (2007)	0.75	6.98	7.96	0.81	6.76	4.99	0.84	6.84	4.71
Shen et al. (2010)	0.72	7.38	8.89	0.74	7.02	5.14	0.75	7.09	4.66

Table 3.5: Range-wise comparison of the performance of CI-based and existing linear-regression models for prediction of carbon (wt%)

Model	Range	Training-set			Test -set			Validation-set		
		CC_{Trn}	$RMSE_{Trn}$	$MAPE_{Trn}$	CC_{Tst}	$RMSE_{Tst}$	$MAPE_{Tst}$	CC_{Vln}	$RMSE_{Vln}$	$MAPE_{Vln}$
GPSR-Model 1	WR	0.66	2.89	4.76	0.94	1.00	1.57	0.90	1.15	2.01
	OR	0.92	6.28	9.95	0.99	3.07	4.04	0.96	4.61	5.51
MLPNN-Model 1	WR	0.70	2.82	4.95	0.95	1.69	3.11	0.93	1.29	2.29
	OR	0.93	6.26	10.24	0.98	4.20	4.95	0.97	12.62	4.04
SVR-Model 1	WR	0.68	2.68	4.60	0.94	1.31	2.35	0.92	0.92	1.64
	OR	0.93	5.87	9.63	0.99	3.33	4.40	0.97	2.77	3.04
Parikh et al. (2007)	WR	0.68	2.75	4.74	0.93	1.52	2.77	0.90	1.11	1.94
	OR	0.86	11.83	15.10	0.97	20.48	24.11	0.96	19.71	25.49
Shen et al. (2010)	WR	0.69	2.78	4.83	0.94	1.40	2.51	0.91	1.00	1.69
	OR	0.84	12.56	17.87	0.96	21.39	27.82	0.95	20.49	26.94

WR: Within range data, OR: Outside range data.

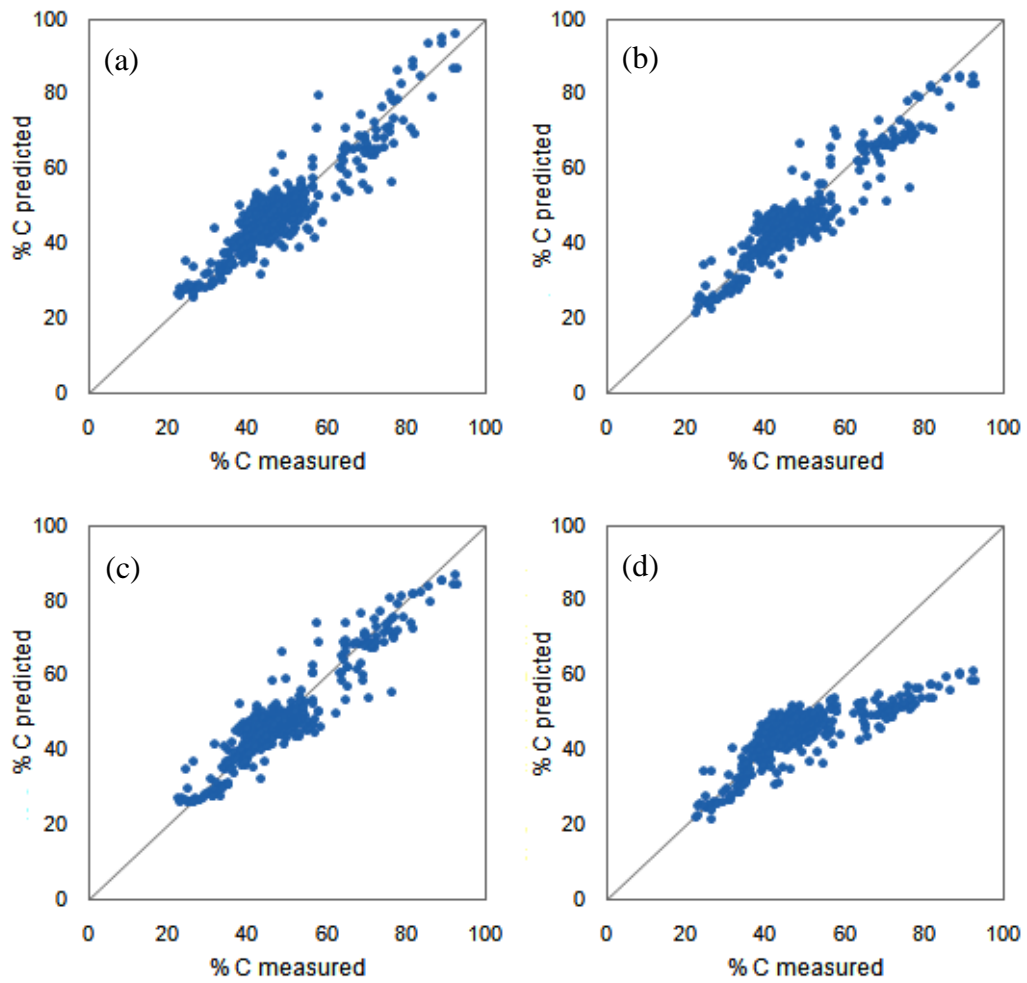


Figure 3.2: Parity-plots of model-predicted and the corresponding experimental values pertaining to the carbon (wt%) predicting following models: (a) GPSR-Model 1, (b) MLPNN-Model 1, (c) SVR-Model 1, and (d) Parikh et al. (Parikh et al., 2007).

Table 3.6: Results of the statistical comparison of the correlation coefficients pertaining to CI-based models predicting the weight percent of carbon

Model pair (B-C)	CC_{AB}	CC_{AC}	CC_{BC}	df	p	H_0
GPSR-MLPNN	0.91	0.93	0.98	829	2.56e-9	Reject
GPSR-SVR	0.91	0.93	0.98	829	8.85e-8	Reject
MLPNN-SVR	0.93	0.93	0.99	829	0.428	Accept

df : degrees of freedom, p : probability, H_0 : null hypothesis proposed

3.2.2. Comparison of Models Predicting Hydrogen (wt%)

The prediction and generalization performances of the hydrogen (wt%) predicting models are indicated in Table (3.7), wherein the CI-based models consistently exhibit better performances ($CCs > 0.74$ for all the sets) as compared to the existing linear models ($CCs < 0.67$). The prediction performance of the MLPNN-Model 2 is slightly better among the developed CI models ($CCs \sim 0.76$, $RMSEs < 0.87$ and $MAPEs < 11.36$) and is also the best overall. Same is indicated in Figure (3.3), which shows the parity plots of comparison of the prediction of the hydrogen (wt%) values by the CI-based and the existing best linear model of Parikh et al. (2007). Similar to the case of carbon (wt%) predictions for “within range” and “out of range” data, the CI-based models predicting Hydrogen (wt%) magnitudes exhibit much better prediction performance (see Table (3.8)) for *out-of-range* data while they produce comparable results for *within-the-range* data. The Steiger’s statistical test (Table (3.9)) for significant differences between the correlation coefficients reveal that among the CI-based models, the MLPNN and SVR-based models predicting the hydrogen weight percentages fair marginally better than the GPSR-based Model 2.

Table 3.7: Comparison of model performances of CI-based and existing linear-regression models for prediction of hydrogen (wt%)

Model	Training-set			Test-set			Validation-set		
	CC_{Trn}	$RMSE_{Trn}$	$MAPE_{Trn}$	CC_{Tst}	$RMSE_{Tst}$	$MAPE_{Tst}$	CC_{Vln}	$RMSE_{Vln}$	$MAPE_{Vln}$
GPSR-Model 2	0.76	0.64	9.61	0.75	0.83	12.89	0.75	0.84	9.73
MLPNN-Model 2	0.76	0.68	9.99	0.77	0.82	11.35	0.76	0.86	9.14
SVR-Model 2	0.77	0.64	10.04	0.76	0.82	12.90	0.74	0.82	10.99
Parikh et al. (2007)	0.66	0.75	11.84	0.50	1.08	19.73	0.56	1.06	14.63
Shen et al. (2010)	0.59	0.80	13.11	0.35	1.16	22.27	0.36	1.15	16.46

Table 3.8: Range-wise comparison of the performance of CI-based and existing linear-regression models for prediction of hydrogen (wt%)

Model	Range	Training-set			Test- set			Validation-set		
		CC_{Trn}	$RMSE_{Trn}$	$MAPE_{Trn}$	CC_{Tst}	$RMSE_{Tst}$	$MAPE_{Tst}$	CC_{Vln}	$RMSE_{Vln}$	$MAPE_{Vln}$
GPSR-Model 2	WR	0.45	0.47	5.67	0.20	0.61	5.28	0.10	0.51	4.08
	OR	0.64	1.12	27.32	0.80	1.65	61.27	0.88	1.78	41.05
MLPNN-Model 2	WR	0.44	0.54	6.38	0.21	0.64	5.73	0.14	0.57	5.08
	OR	0.66	1.13	26.21	0.83	1.52	47.03	0.89	1.74	31.67
SVR-Model 2	WR	0.46	0.49	6.76	0.21	0.66	7.23	0.13	0.55	6.31
	OR	0.68	1.07	24.74	0.82	1.47	48.88	0.85	1.65	36.90
Parikh et al. (2007)	WR	0.45	0.52	6.05	0.17	0.66	5.10	0.10	0.55	4.17
	OR	0.42	1.38	37.84	0.35	2.40	112.64	0.89	2.37	72.52
Shen et al. (2010)	WR	0.44	0.50	5.91	0.16	0.64	5.26	0.10	0.53	4.03
	OR	0.33	1.56	45.41	0.16	2.71	130.32	0.72	2.66	85.32

WR: Within range data, OR: Outside range data.

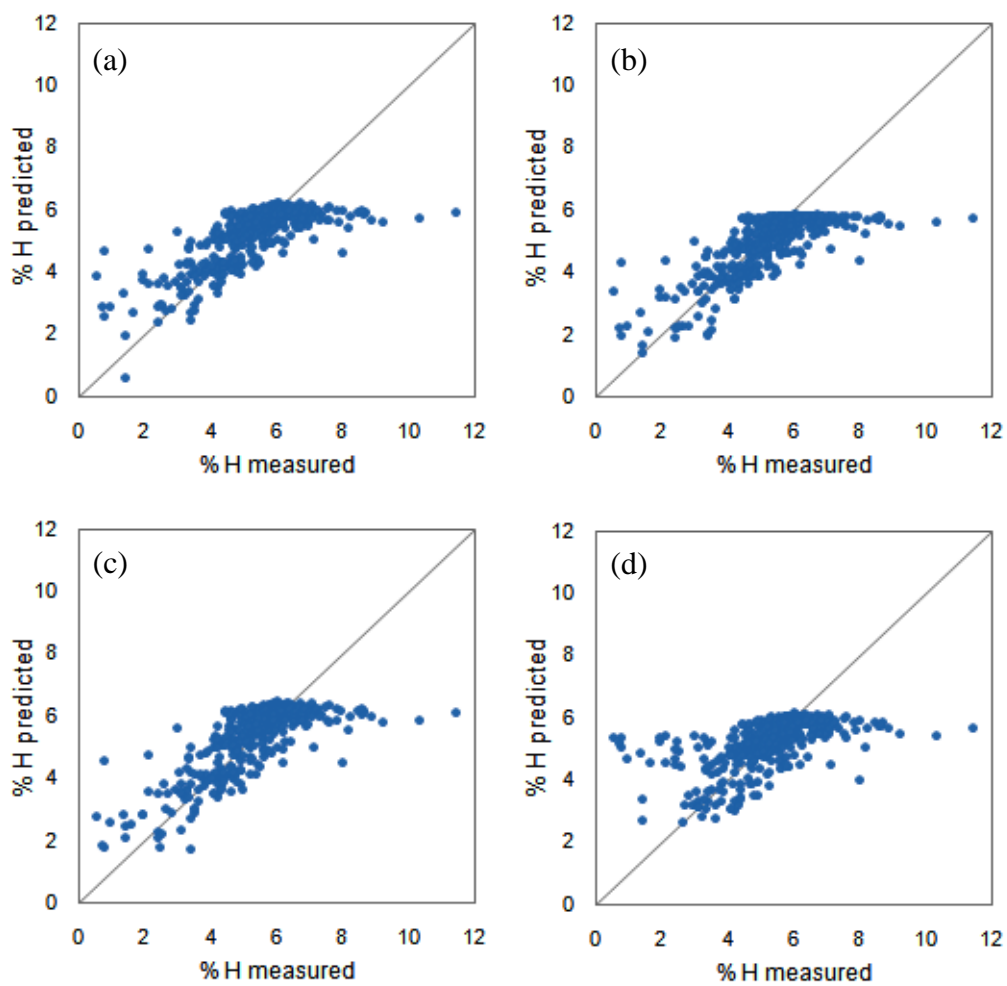


Figure 3.3: Parity-plots of model-predicted and the corresponding experimental values pertaining to the hydrogen (wt%) predicting following models: (a) GPSR-Model 2, (b) MLPNN-Model 2, (c) SVR-Model 2, and (d) Parikh et al. (Parikh et al., 2007).

Table 3.9: Results of the statistical comparison of the correlation coefficients pertaining to CI-based models predicting the weight percentage of hydrogen

Model pair (B-C)	CC_{AB}	CC_{AC}	CC_{BC}	df	p	H_0
GPSR-MLPNN	0.75	0.76	0.99	829	0.013	Reject
GPSR-SVR	0.75	0.76	0.98	829	0.012	Reject
MLPNN-SVR	0.76	0.76	0.98	829	0.382	Accept

df : degrees of freedom, p : probability, H_0 : null hypothesis proposed

3.2.3. Comparison of Models Predicting Oxygen (wt%)

The performance of models predicting oxygen (wt%) values in biomass fuels (see Table (3.10)) shows that the CI-based models outperform (as indicated by the corresponding higher CC s (>0.89) of the CI-based models) the existing linear models. The most marked reduction is seen in the $RMSE$ s relative to the predictions by the CI-based models (~ 4.0 for the training set and ~ 2.0 for the other sets) as compared to higher $RMSE$ s relative to the predictions by the existing linear models (>6). The range-wise comparison of the predictions of the developed CI-based models for the prediction of oxygen (wt%) (see Table (3.11)) shows same trends as witnessed in the similar comparison of models predicting carbon and hydrogen weight percentages. The statistical significance (Steiger's test) test for the evaluation of differences in CC magnitudes in respect of predictions by pairs of CI-based models reveals that the MLPNN-based model performs better than the remaining two models. This is seen in Figure (3.4), wherein a marginally smaller scatter is observed in the parity-plot of the MLPNN model predicted values as compared to the parity-plots of other models.

Table 3.10: Comparison of model performances of CI-based and existing linear-regression models for prediction of oxygen (wt%)

Model	Training-set			Test-set			Validation-set		
	CC_{Trn}	$RMSE_{Trn}$	$MAPE_{Trn}$	CC_{Tst}	$RMSE_{Tst}$	$MAPE_{Tst}$	CC_{Vln}	$RMSE_{Vln}$	$MAPE_{Vln}$
GPSR-Model 3	0.89	4.56	7.11	0.99	1.36	1.80	0.98	2.10	2.33
MLPNN-Model 3	0.90	4.84	15.14	0.99	2.13	8.45	0.98	2.14	5.71
SVR-Model 3	0.89	4.45	12.18	0.98	1.53	5.24	0.98	1.56	4.20
Parikh et al. 2007	0.79	7.69	30.93	0.89	6.30	26.37	0.94	6.50	21.09
Shen et al. 2010	0.75	7.99	32.14	0.82	6.83	28.30	0.91	7.12	23.01

Table 3.11: Range-wise comparison of the performance of CI-based and existing linear-regression models for prediction of oxygen (wt%)

Model	Range	Training-set			Test-set			Validation-set		
		CC_{Trn}	$RMSE_{Trn}$	$MAPE_{Trn}$	CC_{Tst}	$RMSE_{Tst}$	$MAPE_{Tst}$	CC_{Vln}	$RMSE_{Vln}$	$MAPE_{Vln}$
GPSR-Model 3	WR	0.56	3.10	5.82	0.93	0.91	1.66	0.91	1.02	1.87
	OR	0.84	6.74	27.20	0.92	3.29	29.18	0.92	5.47	40.58
MLPNN-Model 3	WR	0.61	3.26	6.71	0.92	1.81	4.03	0.90	1.30	2.69
	OR	0.85	7.19	33.80	0.92	3.92	46.60	0.86	5.14	28.35
SVR-Model 3	WR	0.62	2.95	5.75	0.93	1.16	2.11	0.91	1.00	1.76
	OR	0.84	6.65	26.39	0.90	3.29	32.22	0.85	3.64	22.49
Parikh et al. 2007	WR	0.60	3.10	6.27	0.93	1.36	2.95	0.89	0.94	1.76
	OR	0.65	12.98	85.49	0.21	19.12	228.13	0.87	18.79	166.10
Shen et al. 2010	WR	0.62	3.06	6.19	0.93	1.34	2.91	0.91	0.91	1.72
	OR	0.58	13.58	89.56	0.01	20.83	247.06	0.72	20.62	182.70

WR: Within range data, OR: Outside range data.

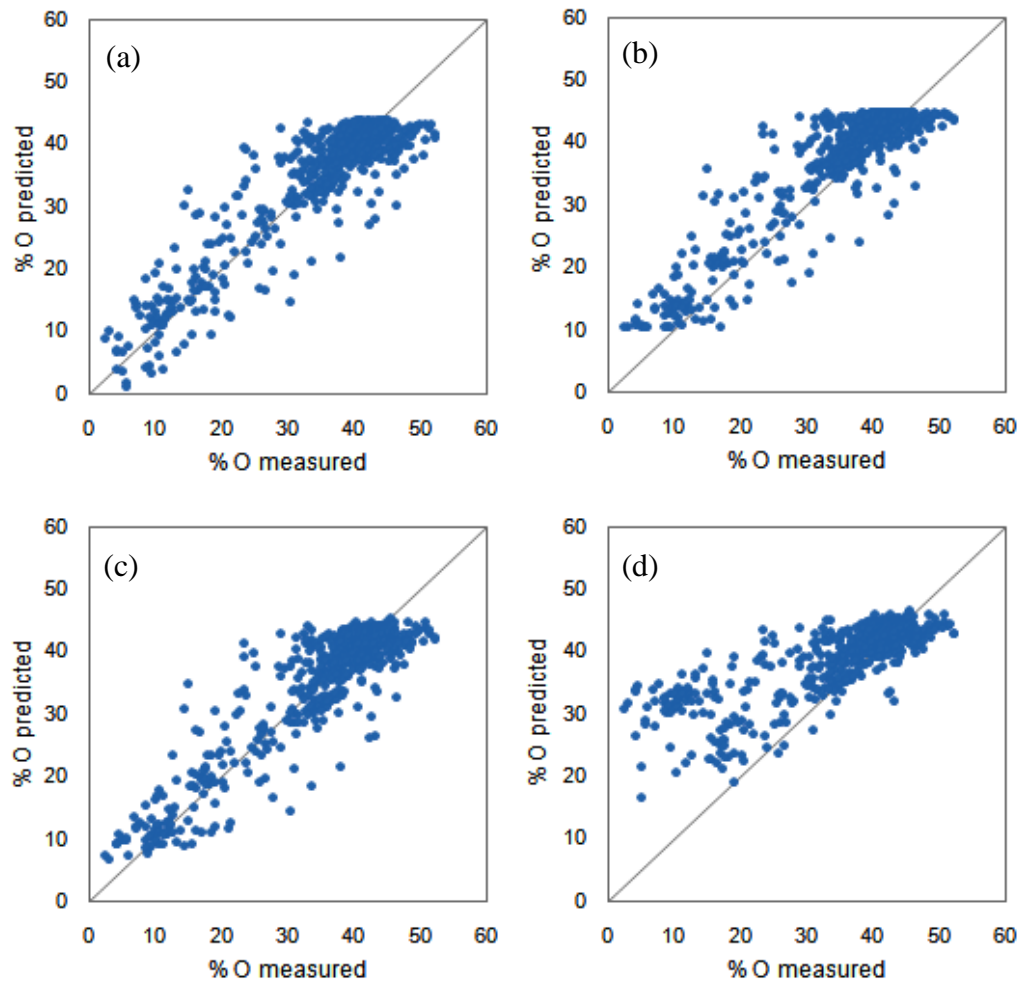


Figure 3.4: Parity-plots of model-predicted and the corresponding experimental values pertaining to the oxygen (wt%) predicting following models: (a) GPSR-Model 3, (b) MLPNN-Model 3, (c) SVR-Model 3, and (d) Parikh et al. (Parikh et al., 2007).

Table 3.12: Results of the statistical comparison of the correlation coefficients pertaining to CI-based models predicting the weight percent of oxygen

Model pair (B-C)	CC_{AB}	CC_{AC}	CC_{BC}	df	p	H_0
GPSR-MLPNN	0.91	0.92	0.99	829	4.12e-7	Reject
GPSR-SVR	0.91	0.92	0.99	829	0.037	Reject
MLPNN-SVR	0.92	0.92	0.99	829	5.59e-4	Reject

df : degrees of freedom, p : probability, H_0 : null hypothesis proposed

Overall it is seen that all the developed CI-based models give much better prediction and generalization accuracies as compared to the existing linear-regression-based conventional models.

3.3. CONCLUSION

There exist only two models for quantitatively estimating the elemental composition, (mainly carbon, hydrogen and oxygen) of solid biomass fuels from the components of their proximate analysis. These models are linear in character and are applicable over limited weight percentage (wt%) ranges of the components involved. It is observed from the scatter plots that some of the individual components of the elemental composition of biomass fuels exhibit low/high nonlinear dependencies with respect to the components of their proximate analysis. Accordingly, nine CI-based nonlinear models (that is three GPSR-based, three MLPNN-based and three SVR-based) were developed in this study for the estimation of elemental composition of biomass fuels from proximate analysis constituents. The CI-based model development was done using a large dataset comprising of 830 samples encompassing a much wider range of percent elemental and proximate analysis composition. The results of the elemental composition estimation clearly suggest that all the developed CI-based (GPSR, MLPNN and SVR) models possess excellent prediction accuracy and generalization performance with the MLPNN and SVR models yielding the best results for the prediction of carbon and hydrogen weight percentages, and the

GPSR and MLPNN models faring better in predicting oxygen weight percentage. When the developed CI models were compared with the existing linear models using “within the range” data— that was considered in the previous studies—they were found to provide marginally improved performance, while for “out-of-range” data (that is those falling outside the ranges considered in earlier studies), the CI models outperform the existing models. This result clearly points towards a wider applicability of CI-based models. All the CI-based models are found to estimate the elemental composition magnitudes with better accuracies when compared to the prevailing linear models. The Steiger’s statistical significance test for correlation coefficients (*CCs*) reveals that for the prediction of carbon (wt%) and hydrogen (wt%) the MLPNN and SVR-based models, and for oxygen (wt%) the MLPNN-based model are more appropriate. More importantly, all CI-based models perform equally well compared to the existing linear models. This study clearly shows that CI-based formalisms provide an attractive strategy for the estimation of elemental composition of solid biomass fuels from the components of their proximate analysis. The proposed CI-based modeling approach can further be extended gainfully for the prediction of other useful properties of wide variety of solid, liquid and gaseous fuels.

NOMENCLATURE

H_0	Null hypothesis in Steiger’s test
L	Number of first hidden layer nodes in MLPNN
M	Number of second hidden layer nodes in MLPNN
N	Number of input layer nodes in MLPNN
p	Probability
\hat{x}	Normalized variable

Greek letters

μ	Momentum coefficient in the EBP algorithm
ε	Epsilon (tube width); a precision parameter in SVR

- η Learning rate in the EBP algorithm
- γ Kernel gamma of radial basis function in SVR

REFERENCES

- ASTM D3176-09.(2009). Standard practice for ultimate analysis of coal and coke. ASTM International, Pennsylvania, USA.
- ASTM D5142-04.(2004). Standard Test Method for Proximate Analysis of the Analysis Sample of Coal and Coke by Instrumental Procedure.ASTM International, West Conshohocken, PA, USA.
- Bishop, C. M. (1995). Neural networks for pattern recognition. Oxford University Press, Oxford.
- Cordero, T., Marquez, F., Rodriquez, M. J., Rodriguez, J. J. (2001).Predicting heating values of lignocellulosic and carbonaceous materials from proximate analysis. Fuel 80:1567–1571.
- ECN Phyllis. (2012). The composition of biomass and waste. Webpage: <http://www.ecn.nl/phyllis/single.html>. Accessed: 15April 2014
- Freeman, J. A., Skapura, D. M. (1991). Neural Networks: Algorithms, Applications, and Programming Techniques. Addison-Wesley Reading.
- Ivanciuc, O. (2007). Applications of Support Vector Machines in Chemistry, In: Lipkowitz, K.B., Cundari, T.R., (Eds.), Reviews in Computational Chemistry. Wiley-VCH., Weinheim, pp. 291-400.
- Koza, J. R. (1992). Genetic programming: on the programming of computers by means of natural selection. MIT, Cambridge.
- Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., Euler, T. (2006). YALE: rapid prototyping for complex data mining tasks, in: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-06).

- Parikh, J., Channiwala, S. A., Ghosal, G. K. (2007). A correlation for calculating elemental composition from proximate analysis of biomass materials. *Fuel* 86:1710–1719.
- Poli, R., Langdon, W. B., McPhee, N. F. (2008). A field guide to genetic programming. Published via <http://lulu.com/> and freely available at <http://gp-field-guide.org.uk/>; (With contributions by J. R. Koza). Accessed: 22 January 2013
- Schmidt, M., Lipson, H. (2009). Distilling free-form natural laws from experimental data. *Science* 324:81–85.
- Shen, J., Zhu, S., Liu, X., Zhang, H., Tan, J. (2010). The prediction of elemental composition of biomass based on proximate analysis. *Energy Conversion and Management* 51(5):983–987.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin* 87:245–251.
- Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer Verlag, New York, USA.

CHAPTER 4.

DEVELOPMENT OF HIGH PERFORMING MODELS FOR THE PREDICTION OF HIGHER HEATING VALUE OF COALS OF DIFFERENT RANKS AND FROM DIVERSE GEOGRAPHIES USING COMPUTATIONAL INTELLIGENCE BASED METHODOLOGY

DEVELOPMENT OF HIGH PERFORMING MODELS FOR THE PREDICTION OF HIGHER HEATING VALUE OF COALS OF DIFFERENT RANKS AND FROM DIVERSE GEOGRAPHIES USING COMPUTATIONAL INTELLIGENCE BASED METHODOLOGY

Abstract

The *higher heating value* (HHV) directly reflects the energy producing potential of coal and, therefore, it is an important property of coal. Its magnitude can be estimated using coal's proximate and/or ultimate analyses. Speedy and accurate estimation of HHV is important for energy calculations involved in the design and operation of coal's combustion and gasification processes. As the analytical determination of coal's HHV is tedious, and costly, numerous correlations for estimating the HHV of coal from the constituents of its proximate and/or ultimate analyses have been proposed over the years; even now efforts to propose correlations predicting HHV of coals more accurately continue. Accordingly, this study proposes five proximate/ultimate analyses based nonlinear models for the prediction of HHV of coals. The principal attributes of these models are: (i) a relatively newer computational intelligence based methodology, namely, *genetic programming* based *symbolic regression* (GPSR) possessing a number of attractive features has been introduced for developing the models, (ii) a huge number (7682) of data pertaining to a variety of coals from several countries have been utilized in the model development, (iii) the prediction accuracy and generalization abilities of the GPSR-based models is excellent, and (iv) the models possess much lower complexity when compared with the other computational intelligence (CI) based HHV prediction models. A comparison of the prediction and generalization performance of the individual GPSR-based models developed in this study with that of the currently available coal HHV models indicate that the models in the former category have consistently outperformed the existing models and therefore possess a high potential to replace the existing models.

4.0. INTRODUCTION

On a quantitative basis, coal is the world's largest consumed solid fuel for energy production. Its high energy content and abundant availability makes coal a vital energy source for the present and future. The total heat content of a unit mass of coal is determined in terms of the *higher heating value* (HHV), also known as *gross calorific value*. It is defined as the amount of heat evolved when a unit weight of the fuel is burnt completely and the combustion products cooled to a standard temperature of 298 K.

The HHV of a coal sample is experimentally determined using a bomb calorimeter as per ASTM standard procedures (ASTM, 2011). Similar to solid biofuels (see Chapter 2), the HHV of a coal sample is highly correlated with its proximate and/or ultimate analyses. The *proximate analysis* determines the content of coal's four important individual components, namely *moisture*, *volatile matter*, *ash*, and *fixed carbon* while the *ultimate analysis* measures contents of the following basic elements: *carbon*, *hydrogen*, *nitrogen*, *sulphur*, and, *oxygen*. These analyses are initially performed on the "as received" basis and, subsequently, can be converted to different bases such as the "dry" and "dry ash-free (DAF)". The HHV of coal, being its energy content indicator, is used widely in the optimal design and operation of coal-based combustion and gasification processes, their pollution compliance assessment, and, determining the coal's rank (Singh and Kakati, 1994).

4.1. SURVEY OF COAL HHV MODELS

Attempts to develop coal's HHV prediction models started in the early nineteenth century, wherein the elemental composition of a fuel was directly related to its HHV. Later, empirical coal's HHV correlations (mostly linear regression based) proved to exhibit improved results in terms of higher accuracies of prediction (Grummel and Davis, 1933; Selvig and Gibson, 1945; Mason and Gandhi, 1983; Neavel et al., 1986; Urkan and Arikol, 1989; Mazumdar and Konovalov, 2002; Sheng and Azevedo, 2005; Majumder et al., 2008; Mesroghli et al., 2009; Yin, 2011; Kavšek et al., 2013). These attempts were predominantly directed towards the development of models for coals of specific rank(s) and/or

from specific geographic locations (Selvig and Gibson, 1945; Neavel et al., 1986; Urkan and Arikol, 1989; Küçükbayrak et al., 1991; Cordero et al., 2001; Mazumdar and Konovalov, 2002; Parikh et al., 2005; Patel et al., 2007; Majumder et al., 2008; Akkaya, 2009; Kavšek et al., 2013). Thus, these models, although showed high prediction and generalization abilities for the coal samples under study, however, exhibited poor performance at predicting HHV of coals of different ranks and/or from different regions. Recently, CI-based models have been developed to predict the HHV of coal and other solid biomasses. Specifically, these nonlinear models possessing high prediction and generalization performance are based on *artificial neural networks* (ANN) (Patel et al., 2007; Mesroghli et al., 2009; Verma et al., 2010; Kavšek, 2013; Akkaya, 2013; Feng et al., 2015), *co-active neuro-fuzzy inference system* (CANFIS) (Verma et al., 2010), *alternating conditional expectation* (ACE) (Feng et al. 2015), and *support vector regression* (SVR) (Feng et al., 2015; Tan et al., 2015) methodologies. Simultaneously, the recent emphasis is also on the development of generalized HHV correlations comprising several ranks of coal as also unified HHV models for a wide range of biomass fuels including coal (Channiwala and Parikh, 2002). A good review of the coal HHV correlations developed over-the-years is given by Mathew et al. (2014); an elaborated list of the HHV correlations for solid biomass fuels (including coals) is given by Garcia et al. (2014a, 2014b). A list of HHV predicting correlations is provided in Table (4.1), with their bases of development.

Table 4.1: A comprehensive list of the correlations for predicting HHV of coal

Eqn. No.	Model Reference	HHV Correlation	Material	HHV Basis	N
4.1	Majumder et al. (2008)	$HHV \text{ (MJ/kg)} = -0.03 \text{ }ASH - 0.11 \text{ }M + 0.33 \text{ }VM + 0.35 \text{ }FC$	Coal	Wet, Proximate	250
4.2	Mesroghli et al. (2009)	$HHV \text{ (MJ/kg)} = 37.777 - 0.647 \text{ }M - 0.387 \text{ }A - 0.089 \text{ }VM$	Coal	Wet, Proximate	4540
4.3	Kavšek et al. (2013)	$HHV \text{ (MJ/kg)} = -3.57 + 0.31 \text{ }VM + 0.34 \text{ }FC$	Coal	Wet, Proximate	64
4.4	Dulong (Selvig and Gibson, 1945)	$HHV \text{ (Kcal/kg)} = 81 \text{ }C + 342.5 \text{ } (H - (O/8)) + 22.5 \text{ }S - 6 \text{ } (9 \text{ }H - M)$	Coal	Wet, Ultimate	
4.5	Mesroghli et al. (2009)	$HHV \text{ (MJ/kg)} = -26.29 + 0.275 \text{ }ASH + 0.605 \text{ }C + 1.352 \text{ }H + 0.840 \text{ }N + 0.321 \text{ }S$	Coal	Wet, Ultimate	4540
4.6	Mesroghli et al. (2009)	$HHV \text{ (MJ/kg)} = 6.971 + 0.269 \text{ }C + 0.195 \text{ }N - 0.061 \text{ }ASH - 0.251 \text{ }O_{\text{ex}} + 1.08 \text{ }H_{\text{ex}} - 0.21 \text{ }M$	Coal	Wet, Ultimate	4540
4.7	Cordero (2001)	$HHV \text{ (MJ/kg)} = 0.3543 \text{ }FC + 0.1708 \text{ }VM$	Biomass	Dry, Proximate	
4.8	Parikh et al. (2005)	$HHV \text{ (MJ/kg)} = 0.3536 \text{ }FC + 0.1559 \text{ }VM - 0.0078 \text{ }ASH$	Biomass	Dry, Proximate	550
4.9	Ghugare et al. (2014)	$HHV \text{ (MJ/kg)} = 0.365 \text{ }FC + 0.131 \text{ }VM + (1.397 / FC) + (328.568 \text{ }VM / (10283.138 + 0.531 \text{ }FC^3 \text{ }ASH - 6.863 \text{ }FC^2 \text{ }ASH))$	Biomass	Dry, Proximate,	
4.10	Neavel et al. (1986)	$HHV \text{ (BTU/lb)} = 145.9 \text{ }C + 569.6 \text{ }H - 53.89 \text{ }O + 43.08 \text{ }S - 6.3 \text{ }ASH$	Coal	Dry, Ultimate	120
4.11	Mason and Gandhi (1983)	$HHV \text{ (BTU/lb)} = 198.11 \text{ }C + 620.31 \text{ }H + 80.93 \text{ }S + 44.95 \text{ }ASH - 5153$	Coal	Dry, Ultimate	775
4.12	IGT (1978)	$HHV \text{ (BTU/lb)} = 146.58 \text{ }C + 568.78 \text{ }H + 29.4 \text{ }S - 6.58 \text{ }A - 51.53 \text{ } (O + N)$	Biomass	Dry, Ultimate	700
4.13	Boie (1953)	$HHV \text{ (MJ/kg)} = 0.3517 \text{ }C + 1.1626 \text{ }H + 0.1047 \text{ }S - 0.111 \text{ }O$	Organics	DAF, Ultimate	
4.14	Dulong (Selvig and Gibson, 1945)	$HHV \text{ (MJ/kg)} = 0.3383 \text{ }C + 1.443 \text{ } (H - (O/8)) + 0.0942 \text{ }S$	Coal	DAF, Ultimate	
4.15	Grummel and Davis (1933)	$HHV \text{ (MJ/kg)} = (0.0152 \text{ }H + 0.9875) \text{ } ((C/3) + H - ((O - S)/8))$	Coal	DAF, Ultimate	

N: Number of coal samples in the respective study, *M*: Moisture (%), *FC*: Fixed carbon (%), *VM*: Volatile matter (%), *ASH*: Ash (%), *C*: Carbon (%), *H*: Hydrogen (%), *H_{ex}*: Hydrogen (%) (external), *O*: Oxygen (%), *O_{ex}*: Oxygen (%) (external), *N*: Nitrogen (%), *S*: Sulphur (%), DAF: Dry ash-free basis (proximate/ultimate analysis constituents are based on the weight (%)) and the respective basis of computation is indicated in the “HHV Basis” column)

4.1.1. Need for Nonlinear Coal HHV Models

As observed in Table (4.1), most of the earlier coal HHV prediction correlations based on the proximate/ultimate analysis are linear. Recently, a few studies, such as by Patel et al. (2007) and Ghugare et al. (2014) have examined the nonlinearities existing between the HHV of coal/biomass and the constituents of the corresponding proximate and ultimate analyses. For capturing the stated nonlinear dependencies, Patel et al. (2007) developed ANN based nonlinear HHV prediction models for coals used in Indian thermal power stations; these models indeed possess better HHV prediction and generalization abilities than their linear counterparts. Using a large data set, Ghugare et al. (2014) also developed computational intelligence based nonlinear models for the prediction of HHV of solid biomass fuels; these models have been found to possess much better prediction accuracy and generalization capability when compared with the linear models. A small portion of the large dataset utilized by Ghugare et al. (2014) in the development of the CI-based models pertained to coals. It can thus be seen that there exists a need to develop exclusively coal-specific nonlinear models possessing high HHV prediction accuracy and generalization performance. Accordingly, *genetic programming based symbolic regression* (GPSR) (Koza, 1990; Poli et al., 2008) methodology has been introduced in this study to develop high performing models for predicting the HHV of coals of different ranks and from different geographical locations. Compared to ANNs and SVR, the GPSR is an infrequently used *computational intelligence* (CI) based data-driven modeling method, and being used for the first time for developing coal-specific HHV models. The GPSR-based models have several advantages, which have been detailed in Chapter 1, Section (1.2.2). For the development of the GPSR-based coal HHV models a huge dataset of coal samples from several geographical regions and of various ranks has been utilized. This data set consists of the constituents of coals' proximate and ultimate analyses (model inputs) and the corresponding HHVs.

The next section in this chapter discusses the development of the GPSR-based coal HHV models. The developed GPSR-based models are further

compared with their existing high-performance counterparts. Finally, in the “conclusion” section the outcomes of the study are summarized.

4.2. GPSR-BASED COAL HHV MODEL DEVELOPMENT

The extensive data set used in the development of GPSR based models was compiled from the open coal databases available on-line from, (i) U.S. Geological Survey Coal Quality (U.S.G.S. COALQUAL) database (Brag et al., 2009), (ii) World coal database from the U.S.G.S. (Tewalt et al., 2010), and (iii) Kentucky coal database (Kentucky Geological Survey, 2015). It consists of property data of 7682 coal samples encompassing all ranks of coal. A majority of the compiled coal database contained data of coal samples from various sub-regions of the U.S. (~85%) and the remaining (~15%) from the major coal mining countries of the world from various continents. For modeling, the dataset was partitioned into *training*, *test* and *validation* sets in the ratio 7.5:1.5:1, ensuring uniform distribution of the data in all the sets. The training set was used for developing the models and the test and validation sets were respectively used to evaluate and validate the generalization capabilities of the developed models. The prediction accuracies of each model were estimated in terms of three statistical measures, namely, *coefficient of correlation (CC)*, *root mean square error (RMSE)* and *mean absolute percent error (MAPE)* computed using the measured and model predicted coal HHV values.

Five different GPSR-based models predicting the HHV of coals were developed using *Eureqa Formulize* (Schmidt and Lipson, 2009) software package (see Section (1.2.2) for details). Here, several GPSR runs were conducted by extensively varying the seed function and other GPSR algorithmic options to generate parsimonious models with high prediction accuracies. In this study, five different GPSR-based models have been developed for predicting the HHV of coal using different analysis bases as given below.

- (i) Proximate analysis based model using data on “as-received” basis
- (ii) Ultimate analysis based model using data on “as-received” basis
- (iii) Proximate analysis based model using data on “dry” basis
- (iv) Ultimate analysis based model using data on “dry” basis

- (v) Ultimate analysis based model using data on “dry ash-free” basis

The corresponding five overall best performing GPSR-based models for the prediction of coal HHV are as given below.

- (a) *GPSR-Model 1* (proximate analysis based model using “as-received basis” data)

$$HHV = 0.816 VM + 0.7008 FC + 0.2786 ASH + 0.004835 M^2 + 0.002808 M * ASH - 34.08 - 0.003026 VM^2 \quad (4.16)$$

- (b) *GPSR-Model 2* (ultimate analysis based model using “as-received basis” data)

$$HHV = 1.178H + 0.3608 C + 0.2505 N + 0.1048 S + 0.0002277 ASH^2 - 1.548 - 0.1217 O \quad (4.17)$$

- (c) *GPSR-Model 3* (proximate analysis based model using “dry basis” data)

$$HHV = 0.397 FC_{Dry} + 0.006441 FC_{Dry} \times VM_{Dry} + 0.004573 VM_{Dry} \times ASH_{Dry} - 5.359 \quad (4.18)$$

- (d) *GPSR-Model 4* (ultimate analysis based model using “dry basis” data)

$$HHV = 1.215 H_{Dry} + 0.3597 C_{Dry} + 0.1501 N_{Dry} + 0.108 S_{Dry} + 0.0001834 ASH_{Dry}^2 - 1.628 - 0.09964 O_{Dry} \quad (4.19)$$

- (e) *GPSR-Model 5* (ultimate analysis based model using “dry ash-free” data)

$$HHV = 1.882 H_{DAF} + 0.4011 C_{DAF} + 0.002221 C_{DAF} \times N_{DAF} + 0.001805 C_{DAF} \times S_{DAF} - 7.282 - 0.002338 O_{DAF}^2 - 0.07857 H_{DAF}^2 \quad (4.20)$$

where, M = moisture, FC = fixed carbon, VM = volatile matter, ASH = ash, C = carbon, H = hydrogen, O = oxygen, N = nitrogen and S = sulphur are the input (predictor) variables expressed in weight percentages (on “as received” basis). The corresponding quantities on “dry” basis are denoted as, FC_{Dry} , VM_{Dry} , ASH_{Dry} , C_{Dry} , H_{Dry} , O_{Dry} , N_{Dry} and S_{Dry} while the quantities on “dry ash-free” basis are referred to as, C_{DAF} , H_{DAF} , O_{DAF} , N_{DAF} and S_{DAF} .

4.3. RESULTS AND DISCUSSION

In this section, the results of a comparison of the HHV prediction and generalization performance of five GPSR-based models (Eqns. (4.16)–(4.20)) and the existing high-performance models are presented. For this comparison, CC ,

RMSE and *MAPE* values pertaining to the HHV predictions by all the stated models were evaluated using the same training, test and validation set data as used in the development of five GPSR models.

4.3.1. Comparison of Proximate Analysis Based Models (“as received” basis)

It may be noted that currently, only a few models are available for the prediction of coal HHV from their proximate analysis conducted on “as-received” basis, although usage of such models would be more convenient and cost-effective than the ultimate analysis based models. The performances of GPSR-Model 1 was compared with that of the three prevailing models by Majumder et al.(2008), Mesroghli et al. (2009), and Kavšek et al. (2013), and the results are shown in Table (4.2). It is observed from the table that the *CC* magnitudes pertaining to the HHVs predicted by the GPSR-Model 1 are quite high (~0.99). Also, the Model 1 exhibits much lower *RMSE* and *MAPE* values. These results clearly indicate the superiority of the GPSR-Model 1 in providing an excellent HHV prediction accuracy and generalization performance. Same conclusion can be also drawn from the lower scatter (when compared with the predictions of the best existing model of Majumder et al. (2008)) seen in the parity plots (see Figure (4.1)) portraying the GPSR-Model 1 predicted coal HHV values and the experimental HHV values.

Table 4.2: Comparison of prediction performances of proximate analysis based coal HHV models using “as-received” basis data

Model	Training-set			Test-set			Validation-set		
	<i>CC</i>	<i>RMSE</i>	<i>MAPE</i>	<i>CC</i>	<i>RMSE</i>	<i>MAPE</i>	<i>CC</i>	<i>RMSE</i>	<i>MAPE</i>
GPSR-Model 1	0.99	0.79	2.52	0.99	0.70	2.20	0.99	0.69	1.99
Majumder et al. (2008)	0.98	1.27	4.38	0.98	1.19	3.99	0.98	1.11	3.51
Mesroghli et al. (2009)	0.97	2.63	9.46	0.97	2.68	9.22	0.97	2.38	7.88
Kavšek et al. (2013)	0.97	4.39	14.67	0.98	4.42	14.87	0.98	4.51	15.28

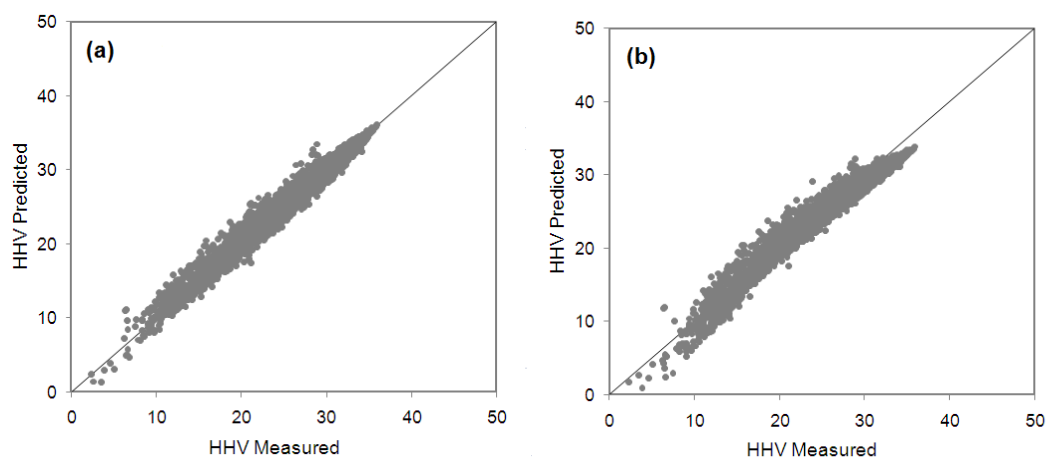


Figure 4.1: Comparative parity plots of model-predicted and the corresponding experimental values pertaining to coal HHVs based on proximate analysis on an as-received basis (a) GPSR-Model 1, (b) Majumder et al. (2008).

4.3.2. Comparison of Ultimate Analysis Based Models (“as received” basis)

The performance monitoring statistical measures pertaining to the HHV predictions made by GPSR-Model 2 and three other existing models by Dulong (Selvig and Gibson, 1945) and Mesroghli et al. (2009) are given in Table (4.3). From the table, minor differences are observed among the *CC* values of competing models, and major differences are noticed in the corresponding *RMSE* and *MAPE* values; here, among all competing models, marginally lower *RMSE* and *MAPE* magnitudes pertaining to the predictions of the GPSR based model indicates its superiority (albeit borderline) over other models. This inference is also corroborated by the slightly lower scatter (see Figure (4.2)) exhibited by the HHV predictions made by GPSR-Model 2.

Table 4.3: Comparison of prediction performances of ultimate analysis based coal HHV models using “as-received” basis data

Model	Training-set			Test-set			Validation-set		
	CC	RMSE	MAPE	CC	RMSE	MAPE	CC	RMSE	MAPE
GPSR-Model 2	0.99	0.27	0.85	0.99	0.27	0.81	0.99	0.26	0.79
Dulong (Selvig and Gibson, 1945)	0.99	0.82	3.04	0.99	0.81	2.99	0.99	0.80	2.87
Mesroghli et al. (2009)	0.98	2.48	9.26	0.98	2.45	8.89	0.98	2.21	7.70
Mesroghli et al. (2009)	0.96	5.11	18.54	0.96	5.06	17.62	0.96	4.47	14.62

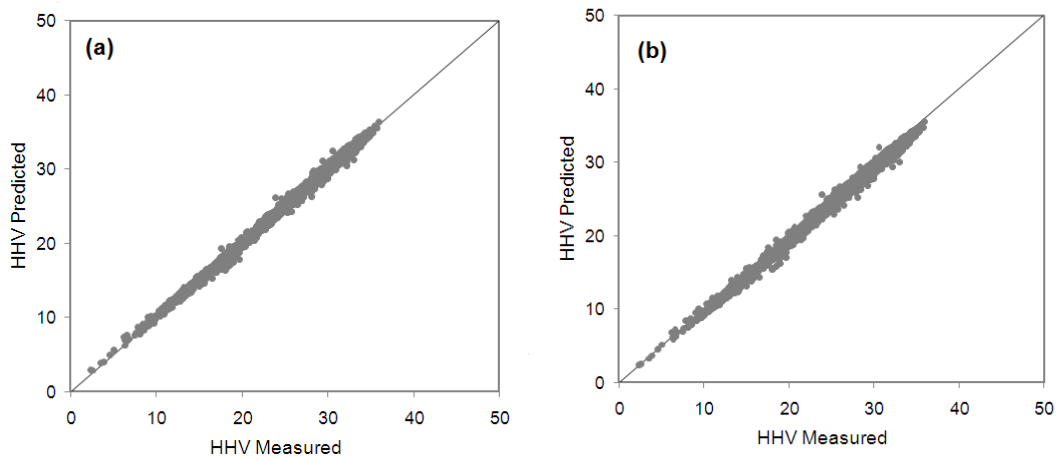


Figure 4.2: Comparative parity plots of model-predicted and the corresponding experimental values pertaining to coal HHVs based on ultimate analysis on an as-received basis (a) GPSR-Model 2, (b) Dulong (Selvig and Gibson, 1945).

4.3.3. Comparison of Proximate Analysis Based Models (“dry” basis)

A comparison of GPSR-Model 3 and three existing models, namely, by Cordero et al. (2001), Parikh et al. (2005), and Ghugare et al. (2014) in predicting accurately the HHV of coals and generalizing the learned knowledge to “out-of-sample” data, is provided in Table (4.4). It may be noted that the model by Ghugare et al. (2014) is essentially for computing the HHV of solid biomasses. Its inclusion here as a competitive model stems from the fact that the large dataset used for its development contained a small subset of coal data. From the tabulated values, it is observed that the prediction and generalization performance of GPSR-based model 3 is much better than the existing three models with the *CC* magnitudes greater than 0.9 for all the three datasets. This is also reflected in the substantially lower values of *RMSEs* and *MAPEs* pertaining to the HHV predictions made by the GPSR-Model 3. In Figure (4.3), predictions of the GPSR-based and the existing model with best prediction accuracy by Cordero et al. (2001) are compared, in which the GPSR-Model 3 shows a lower scatter indicating better performance.

Table 4.4: Comparison of prediction performances of proximate analysis based coal HHV models using “dry” basis data

Model	Training-set			Test-set			Validation-set		
	<i>CC</i>	<i>RMSE</i>	<i>MAPE</i>	<i>CC</i>	<i>RMSE</i>	<i>MAPE</i>	<i>CC</i>	<i>RMSE</i>	<i>MAPE</i>
GPSR-Model 3	0.91	1.74	5.01	0.90	1.62	4.62	0.91	1.58	4.58
Cordero et al. (2001)	0.87	4.76	14.80	0.88	4.89	15.51	0.88	5.04	16.07
Parikh et al. (2005)	0.87	5.36	17.00	0.87	5.51	17.80	0.88	5.67	18.42
Ghugare et al. (2014)	0.85	5.49	17.49	0.86	5.66	18.3	0.86	5.83	18.89

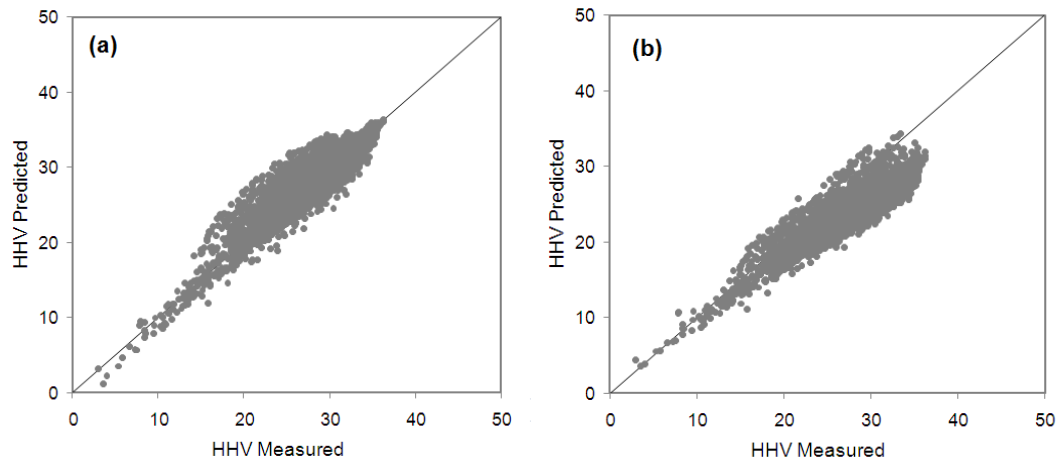


Figure 4.3: Comparative parity plots of model-predicted and the corresponding experimental values pertaining to coal HHVs based on proximate analysis on dry basis (a) GPSR-Model 3, (b) Cordero et. al. (2001).

4.3.4. Comparison of Ultimate Analysis Based Models (“dry” basis)

A number of correlations are available for the prediction of HHV of coal from its ultimate analysis on the dry basis (Mathew et. al., 2014; Garcia et al., 2014a; 2014b). In Table (4.5), a comparison of the GPSR-Model 4 has been made with the stated prevailing three high performing models. From the *CC*, *RMSE* and *MAPE* magnitudes listed in this table, it is noticed that all four models possess comparable HHV prediction and generalization capabilities with GPSR-Model 4 performing marginally better than its three competing models. The parity plots in Figure (4.4) also depict the closely comparable performance of the GPSR-based model and its closest competing model by Neavel et al. (1986).

Table 4.5: Comparison of prediction performances of ultimate analysis based coal HHV models using “dry” basis data

Model	Training-set			Test-set			Validation-set		
	CC	RMSE	MAPE	CC	RMSE	MAPE	CC	RMSE	MAPE
GPSR-Model 4	0.99	0.29	0.78	0.99	0.29	0.76	0.99	0.27	0.75
Neavel et al. (1986)	0.99	0.31	0.85	0.99	0.31	0.82	0.99	0.29	0.78
Mason and Gandhi (1983)	0.99	0.32	0.87	0.99	0.32	0.86	0.99	0.31	0.84
IGT (1978)	0.99	0.32	0.87	0.99	0.32	0.86	0.99	0.31	0.84

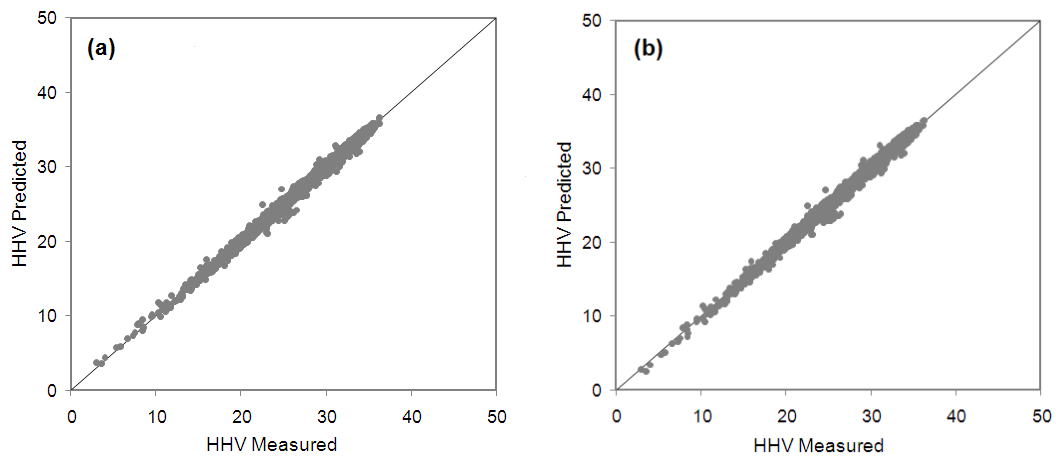


Figure 4.4: Comparative parity plots of model-predicted and the corresponding experimental values pertaining to coal HHVs based on ultimate analysis on dry basis (a) GPSR-Model 4, (b) Neavel et al. (1986).

4.3.5. Comparison of Ultimate Analysis Based Models (“dry ash-free” basis)

Similar to the ultimate analysis (“dry” basis) based models (Mathew et al., 2014; Garcia et al., 2014a; 2014b) a number of models are available for the prediction of HHV of coal/biomass from the constituents of their ultimate analyses on dry ash-free basis. In this study, three such models endowed with superior HHV prediction performance by, namely, Grummel and Davis (1933), Boie (1953), and Dulong’s equation (Selvig and Gibson 1945) were used for comparison with the GPSR-Model 5. The results of this comparison are presented in Table (4.6), which shows that the GPSR-based model possesses marginally improved prediction accuracy and generalization ability compared with that possessed by each of the three competing models. The said observation is also

corroborated by the lower scatter seen in the parity plot of the HHV predictions made by GPSR-Model 5 in Figure (4.5(a)) when compared with that corresponding to the predictions by the model of Boie (1953).

Table 4.6: Comparison of prediction performances of ultimate analysis based coal HHV models using “dry ash-free” basis data

Model	Training-set			Test-set			Validation-set		
	CC	RMSE	MAPE	CC	RMSE	MAPE	CC	RMSE	MAPE
GPSR-Model 5	0.99	0.35	0.77	0.99	0.35	0.77	0.98	0.35	0.77
Boie (1953)	0.99	0.68	1.80	0.98	0.63	1.69	0.98	0.64	1.68
Dulong (Selvig and Gibson 1945)	0.98	0.57	1.39	0.98	0.57	1.40	0.98	0.56	1.32
Grummel and Davis (1933)	0.98	0.53	1.22	0.98	0.53	1.26	0.98	0.54	1.25

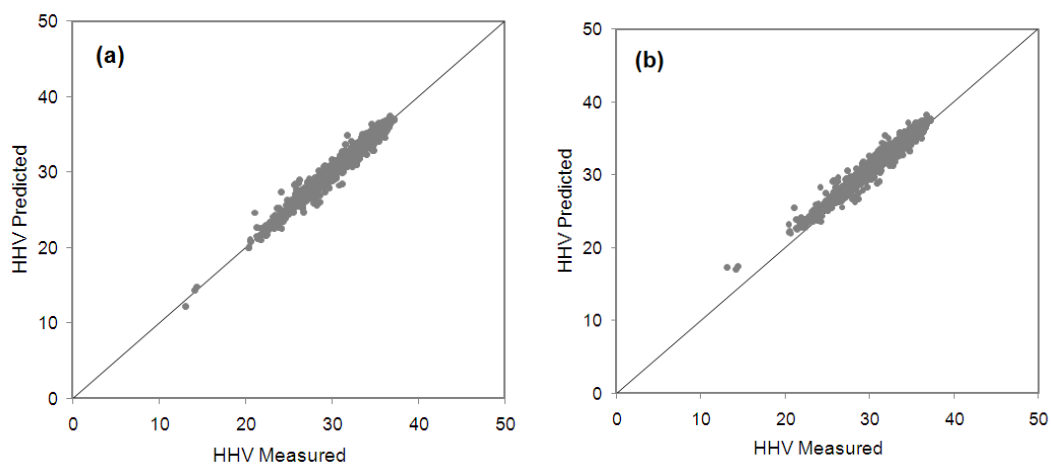


Figure 4.5: Comparative parity plots of model-predicted and the corresponding experimental values pertaining to coal HHVs based on ultimate analysis on dry ash-free basis (a) GPSR-Model 5, (b) Boie (1953).

4.3.6. Statistical Comparison of Coal HHV Models

From the above-described results of comparisons, it is observed that the GPSR-based models have outperformed the existing linear models. However, in some cases the difference between the performance of the GPSR-based model and the corresponding competing model is small. Thus, the Steiger’s statistical test (Steiger, 1980) was performed for unambiguously identifying the best model among the competing ones. This test compares the *CC* values pertaining to the predictions of two competing models to assess whether there exists a statistically

significant difference between them (see Section (1.8) for details). The results of this test are presented in Table (4.7), indicating that the p -values of the test for all the five GPSR-based models, when compared with their counterparts are zero. Thus, there exists a statistically significant difference between the HHV prediction performances of the GPSR-based models and their prevailing counterparts. Since all the GPSR-based models possess higher prediction and generalization performance than their counterparts (as indicated by their higher CC values), it can be inferred that GPSR-based models predicting the HHV of coals have outperformed the corresponding existing models.

Table 4.7: Results of the statistical comparison of the prediction performances pertaining to GPSR-based coal HHV models

Model pair (B-C)	CC_{AB}	CC_{AC}	CC_{BC}	df	p	H_0^*
GPSR-Model 1–Majumder et al. (2008)	0.990	0.983	0.992	7681	0	Reject
GPSR-Model 2–Dulong (Selvig and Gibson 1945)	0.998	0.998	0.999	7681	0	Reject
GPSR-Model 3–Cordero et al. (2001)	0.911	0.878	0.972	7681	0	Reject
GPSR-Model 4–Neveal et al. (1986)	0.997	0.997	0.999	7681	0	Reject
GPSR-Model 5–Boie (1953)	0.991	0.990	0.998	7681	0	Reject

df : degrees of freedom, p : probability

4.4. CONCLUSION

In the past, several correlations have been proposed for predicting the HHV of coals from their proximate and/or ultimate analyses. These are mostly linear, rank and geography-specific, and were developed using a limited amount of data. It is well-proven that HHV of a coal sample and some constituents of its proximate and ultimate analyses exhibit nonlinear interdependencies. Thus, nonlinear models are expected to exhibit better HHV prediction accuracies than the linear ones. Accordingly, this study employed the relatively newer computational intelligence methodology, namely genetic programming based symbolic regression (GPSR) for developing a number of models for the prediction of HHV of coals. In particular, five GPSR-based HHV predicting models were developed by employing a huge dataset of 7682 coal samples belonging to multiple ranks and from major coal-mining geographic regions from the world. The data set contains proximate/ultimate analyses conducted using “as received”, “dry” and “dry-ash-free” basis and these were used as inputs to the GPSR-based

models. All five models were found to possess excellent prediction accuracies and generalization capabilities. Additionally, the HHV prediction and generalization performance of each GPSR-based model was compared with that of the existing high performance yielding models. The results obtained thereof using the statistical Steiger's test clearly indicate that the GPSR-based models outperform the existing models. In summary, the high performing GPSR-based models presented here owing to their lower complexity, and superior HHV prediction and generalization abilities exhibit the potential to replace the existing models for HHV prediction of coals.

NOMENCLATURE

p	Probability
H_0	Null hypothesis for Steiger's test

REFERENCES

- Akkaya, A. V. (2009). Proximate analysis based multiple regression models for higher heating value estimation of low rank coals. *Fuel Process Technology* 90:165–170.
- Akkaya, A. V. (2013). Predicting Coal Heating Values Using Proximate Analysis via a Neural Network Approach, *Energy Sources. Part A* 35:253–260.
- ASTM, International, ASTM D5865–10.(2011). Standard Test Method for Gross Calorific Value of Coal and Coke.
- Boie, W. (1953). *Energietechnik* 3:309.
- Bragg, L. J., Oman, J. K., Tewalt, S. J., Oman, C. J., Rega, N. H., Washington, P. M., Finkelman, R. B. (2009). U.S. Geological Survey Coal Quality (COALQUAL) database: Version 2.0. U.S. GEOLOGICAL SURVEY OPEN-FILE REPORT 97-134. <http://energy.er.usgs.gov/products/databases/CoalQual/index.htm>. Accessed: 25 November 2014

- Channiwala, S. A., Parikh, P. P. (2002). A unified correlation for estimating HHV of solid, liquid and gaseous fuels. *Fuel* 81:1051–1063.
- Cordero, T., Marquez, F., Rodriguez, M. J., Rodriguez, J. J. (2001). Predicting heating values of lignocellulosic and carbonaceous materials from proximate analysis. *Fuel* 80:1567–1571.
- Feng, Q., Zhang, J., Zhang, X., Wen, S. (2015). Proximate analysis based prediction of gross calorific value of coals: A comparison of support vector machine, alternating conditional expectation and artificial neural network. *Fuel Processing Technology* 129:120–129.
- García, R., Pizarro, C., Lavín, A. G., Bueno, J. L. (2014a). Spanish biofuels heating value estimation. Part I: Ultimate analysis data. *Fuel* 117:1130–1138.
- García, R., Pizarro, C., Lavín, A. G., Bueno, J. L. (2014b). Spanish biofuels heating value estimation. Part II: Proximate analysis data. *Fuel* 117:1139–1147.
- Ghugare, S. B., Tiwary, S., Elangovan, V., Tambe, S. S. (2014). Prediction of Higher Heating Value of Solid Biomass Fuels using Artificial Intelligence Formalisms. *Bioenergy Research* 7:681–692.
- Grumel, E. S., Davis, I. A. (1933). A New Method of Calculating the Calorific Value of a Fuel From Its Ultimate Analysis. *Fuel* 1:199–203.
- IGT, Institute of Gas Technology. (1978). Coal conversion systems technical data book. NTIS, Springfield, VA.
- Kavšek, D., Bednárová, A., Biro, M., Kranvogel, R., Vončina, D.B., Beinrohr, E. (2013). Characterization of Slovenian coal and estimation of coal heating value based on proximate analysis using regression and artificial neural networks. *Central European Journal of Chemistry* 11:1481-1491.
- Kentucky Geological Survey.(2015). Coal database by the Kentucky Geological Survey, University of Kentucky.

<http://kgs.uky.edu/kgsweb/DataSearching/Coal/Quality/QualitySearch.asp>.

Accessed: 25 November 2014

- Koza, J. R. (1990). Genetically breeding populations of computer programs to solve problems in artificial intelligence, In: Proceedings of the 2nd International IEEE Conference on Tools for Artificial Intelligence, pp. 819–827.
- Küçükbayrak, S., Dürüs, B., Meriçboyu, A. E., Kadioğlu, E. (1991). Estimation of calorific values of Turkish lignites. *Fuel* 70(8):979-981.
- Majumder, A. K., Jain, R., Banerjee, P., Barnwal, J. P. (2008). Development of a new proximate analysis based correlation to predict calorific value of coal. *Fuel* 87:3077–3081.
- Mason, D. M., Gandhi, K. N. (1983). Formulas for calculating the calorific value of coal and coal chars: development, tests, and uses. *Fuel Processing Technology* 7:11–22.
- Mathews, J. P., Krishnamoorthy, V., Louw, E., Tchaptala, A. H. N., Marcano, F. C., Karri, V., Alexis, D. A., Mitchell, G. D. (2014). A review of the correlations of coal properties with elemental composition. *Fuel Processing Technology* 121:104–113.
- Mazumdar, B. K., Konovalov. (2002). Correlation for molar heats of combustion of organic compounds: its adoption for coal. *Fuel* 81:119–124.
- Mesroghli, S., Jorjani, E., Chelgani, S. C. (2009). Estimation of gross calorific value based on coal analysis using regression and artificial neural networks. *International Journal of Coal Geology* 79:49–54.
- Neavel, R. C., Smith, S. E., Hippo, E. J., Miller, R. N. (1986). Interrelationships between coal compositional parameters. *Fuel* 65(3):312–320.
- Parikh, J., Channiwala, S. A., Ghosal, G. K. (2005). A correlation for calculating HHV from proximate analysis of solid fuels. *Fuel* 84:487–494.

- Patel, S. U., Kumar, B. J., Badhe, Y. P., Sharma, B. K., Saha, S., Biswas, S., Chaudhury, A., Tambe, S. S., Kulkarni, B. D. (2007). Estimation of gross calorific value of coals using artificial neural networks. *Fuel* 86:334–344.
- Poli, R., Langdon, W. B., McPhee, N. F., (2008). A field guide to genetic programming. Published via <http://lulu.com/> and freely available at <http://gp-field-guide.org.uk/>; (With contributions by J. R. Koza). Accessed: 25 November 2014.
- Schmidt, M., Lipson, H. (2009). Distilling free-form natural laws from experimental data, *Science* 324:81–85.
- Selvig, W. A., Gibson, F. H. (1945). Calorific value of coal, In: H.H. Lowry (Ed.), *Chemistry of Coal Utilization*. vol. 1, J. Wiley & Sons, New York, pp. 132–144.
- Sheng, C., Azevedo, J. L. T. (2005). Estimating the higher heating value of biomass fuels from basic analysis data. *Biomass and Bioenergy* 28:499–507.
- Singh, P. H., Kakati, M. C. (1994). New models for prediction of specific energy of coal. *Fuel* 73:301–303.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin* 87:245–251.
- Tan, P., Zhang, C., Xia, J., Fang, Q. Y., Chen, G. (2015). Estimation of higher heating value of coal based on proximate analysis using support vector regression. *Fuel Processing Technology* 138:298-304.
- Tewalt, S. J., Belkin, H. E., SanFilipo, J. R., Merrill, M. D., Palmer, C. A., Warwick, P. D., Karlsen, A. W., Finkelman, R. B., Park, A. J. (2010). Chemical analyses in the World Coal Quality Inventory, version 1: U.S. Geological Survey Open-File Report comp. Accessed: 25 November 2014
- Urkan, M. K., Arikol, M. (1989). Correlations for the heating value of Turkish coals. *Fuel* 68:527–530.

Verma, A. K., Singh, T. N., Monjezi, M. (2010). Intelligent prediction of heating value of coal. *Iranian Journal of Earth Sciences* 2:32–38.

Yin, C. Y. (2011). Prediction of higher heating values of biomass from proximate and ultimate analysis. *Fuel* 90:1128–1132.

CHAPTER 5.

DEVELOPMENT OF COMPUTATIONAL INTELLIGENCE BASED SOFT-SENSOR MODEL FOR STYRENE POLYMERIZATION PROCESS AND ITS APPLICATION IN A MODEL PREDICTIVE CONTROL SCHEME

DEVELOPMENT OF COMPUTATIONAL INTELLIGENCE BASED SOFT-SENSOR MODEL FOR STYRENE POLYMERIZATION PROCESS AND ITS APPLICATION IN A MODEL PREDICTIVE CONTROL SCHEME

Abstract

A mathematical model relating the key process variable, which is tedious and difficult/impossible to measure online, to the process variable that can be measured online easily, is termed as a *soft-sensor*. Soft-sensors are valuable alternatives to the conventional hardware sensors in chemical process industries. They majorly assist in the tasks of process monitoring and control. The quality of the final polymer product in a typical polymerization process is characterized by its average molecular weight, whose online measurement is difficult, and time-consuming. Also, the modern polymerization processes involve multiple reactions occurring in series and parallel along with complex heat and mass transfer phenomena. These features create difficulties for process modeling by conventional approaches such as phenomenological modeling. *Genetic programming* based *symbolic regression* (GPSR) is a computational intelligence-based exclusively data-driven modeling methodology that has rarely been used for soft-sensor development and identification of polymer processes. Thus, in this work a GPSR-based soft-sensor modeling approach is proposed for the process identification of an *industrial scale continuous styrene polymerization reactor process*. The developed GPSR-based soft-sensor models relate the average molecular weight of the effluent polystyrene product—which is difficult to measure online—to the input parameters of the reactor. Specifically, two soft sensor models were developed possessing high prediction and generalization performance. The best GPSR-based model was advantageously used in the design of a *model predictive control* (MPC) scheme for the styrene polymerization reactor process.

5.0. INTRODUCTION

The modern chemical processes often operating at high production rates need sophisticated control systems to tackle various issues related to the safety and environmental regulation in addition to the basic needs of optimizing operating costs and maintaining highest standards of product quality. This, in turn, motivated new developments in automatic process control technology. The universal drive for consistent high product quality, efficient use of energy, and an increasing environmental awareness impose far stricter demands on chemical process control systems. This has paved the way for modern research, resulting in significant advancements in control technologies. In the last few decades, the industry has seen successful applications of innovative control technologies based on *model predictive control* (MPC) and more recently intelligent control (Ogunnaike and Ray, 1994).

Soft-sensor is a recent and novel approach that has greatly assisted in process identification and control. It is a model that mathematically captures the relation between the key process variables, which are tedious and difficult/impossible to measure online, to the process variables that can be measured easily. *Computational intelligence* (CI) based methodologies such as artificial neural networks (ANNs) (Jos de Assis and Maciel Filho, 2000; Meleiro and Finho, 2000; Devogelaere et al., 2002; Li et al., 2005; Desai et al., 2006) and support vector regression (SVR) (Feng et al., 2003; Yan et al., 2004; Desai et al., 2006) have been used in the development of soft-sensors for process monitoring and control applications. The relatively newer CI-based exclusively data-driven modeling methodology, namely, *genetic programming* based *symbolic regression* (GPSR) (Koza, 1994) though possesses some attractive characteristics (see Section (1.2.2)), has been rarely used in soft-sensor development. Accordingly, in this study we demonstrate the capabilities of GPSR for developing dynamic soft-sensor models of a *continuous styrene polymerization reactor*; next, the model has been effectively utilized for implementing model based control (MPC) for the stated polymerization process.

5.1. PROCESS IDENTIFICATION OF POLYMERIZATION PROCESSES

Process identification is essentially the development of mathematical models of a process/system to accurately capture its dynamic (unsteady-state) behavior. It relates the process outputs to its inputs. Such a process model is mainly required for different process tasks such as process optimization, monitoring, and simulation to appropriately design the relevant control system. Often, empirical modeling approaches are employed for process identification, as these can model any linear/nonlinear process behavior from process input-output/stimulus-response data. In general, polymerization processes are difficult to model due to following reasons (Yoon et al., 2004):

- They involve multiple sequential/parallel reactions with associated mechanisms and kinetics (including parameters) that are hard to determine accurately.
- The reactions are multi-phase and the underlying thermodynamics, and heat and mass transfer phenomena are complex.
- Due to the multiphase reactions and/or non-ideal mixing, non-homogeneities may occur in the reaction mass, thus creating additional modeling difficulties.
- The quality of the polymer product is characterized by structural properties of the final polymer such as molecular weight, copolymer composition, and sequence length distribution, which are difficult to measure accurately online.

Due to all the difficulties/limitations mentioned above the empirical modeling approaches have been increasingly used for the modeling and identification of polymerization processes, a good review of which can be found in Yoon et al. (2004) and Richards and Congalidis (2006). In addition to the conventional empirical modeling approaches such as the step and impulse response modeling, the newer modeling strategies such as Volterra series modeling, bilinear modeling, Hammerstein modeling, Wiener modeling, and autoregressive moving-average (ARMA) modeling have gained popularity for modeling of polymerization

reactions (Na and Rhee, 2000). However, the models developed using such modeling strategies are mostly linear and computationally expensive to execute due to their complexities. In recent years, computational intelligence (CI) based exclusively data-driven modeling formalisms such as ANNs, SVR, and fuzzy logic (FL) are becoming popular in various engineering applications due to their ability of representing the behavior of complex and nonlinear processes with an excellent accuracy. However, the transfer functions in an ANN model, the kernel function in an SVR model, and the membership functions in an FL model is required to be appropriately pre-specified to obtain a suitably accurate process model. In comparison, the GPSR-based models are easier to develop and it can automatically model linear/nonlinear relations between process input-output variables. The details of GPSR method and its advantages are given in Section (1.2.2).

In this study, the GPSR methodology was utilized for the process identification of an industrial scale continuous styrene polymerization process. Specifically, two GPSR-based soft sensor models were first developed from the input-output data of the polymerization reactor for capturing the dynamics of a “difficult-to-measure” controlled variable. Next, the best GPSR-based soft-sensor model was used for the implementation of an MPC strategy for controlling the polystyrene quality produced by the styrene polymerization reactor.

5.2. MODEL PREDICTIVE CONTROL OF POLYMERIZATION PROCESSES

The traditional proportional-integral-derivative (PID) feedback control schemes are often found to perform un-satisfactorily for complex chemical processes due to the presence of high nonlinearities with strong interactions among the process variables. Another common problem with the use of PID control systems is that they require frequent tuning to optimally set their parameters for achieving best control performance. This need stems from the variations in the dynamics of chemical processes which make the previously tuned parameters invalid, thus deteriorating the control performance over time.

Polymerization processes commonly exhibit such characteristics and thus numerous advanced control schemes have been implemented to satisfy the various desirable controller properties such as accuracy, consistency, robustness and adaptive nature. With the advancement in computing hardware, software and networking technologies, achieving these goals have become relatively easier. The *model predictive control* (MPC) strategy is one such approach built on modern day digital technologies that utilizes a software-based model of the process to predict the process behavior and thereby implementing most appropriate control action. This is especially useful in adapting to the changing process conditions.

Polymerization processes—due to their previously mentioned difficulties—offer numerous challenges for process measurements, and accurate and robust process control. Such processes hardly run smoothly and minor disturbances in any input(s) or state variables can easily upset the process behavior. For smooth operation of such processes and thereby obtaining a consistent product quality at the desired rate, a set of influential process variables (termed “controlled variables”) must be maintained/changed appropriately over the duration of the polymerization reaction. The MPC scheme is one of the most efficient and appropriate strategies for controlling a polymerization process (Hur et al., 2003; Richards and Congalidis, 2006; Bindlish, 2015).

For styrene polymerization reaction, linear and polynomial ARMA models were developed and utilized in the MPC framework for the reactor control (Na and Rhee, 2000, Na and Rhee, 2002). The main disadvantage of the ARMA models is their higher complexity, which makes them computationally expensive to execute in software. Recently, ANN-based models were used for the predictive control of batch polystyrene reactor (Hosen et al., 2011); however, due to their complex structures, even the ANN-based models create high computational loads during their evaluation. In comparison, the GPSR-based models are less complex and thus easier to execute. MPC has been applied satisfactorily to several polymerization processes of varying complexities utilizing models ranging from the linear regression based on the recently developed CI-based models. A detailed review of such research work is given by Richards and Congalidis (2006).

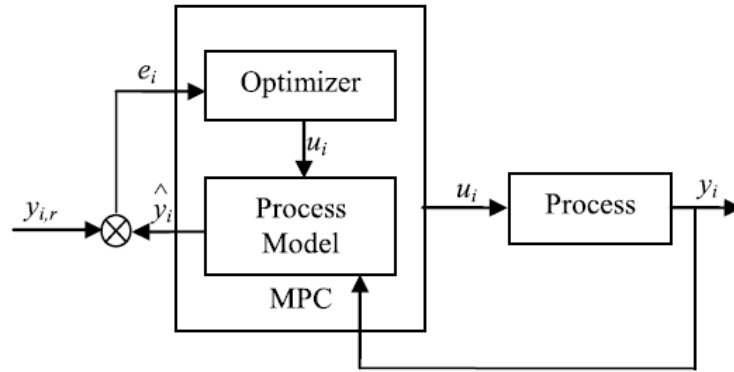


Figure 5.1: Conceptual block-diagram of an MPC scheme.

The block diagram of the MPC scheme applied here for controlling the styrene polymerization process is shown in Figure (5.1). The process model is used to predict the current value of the controlled (output) variable using the current and past input and state information. The process model is often implemented in the form of discrete equations. The variation in set-point of the controlled variable is calculated from the objectives of economic optimization of the process. In general, the economical optimization objectives include minimization of manufacturing cost by reducing the utilization of costly raw-materials or maximization of production rates/profits. Using the current and past values of the manipulated and controlled variables in the identified process model the MPC block computes a sequence of the future (multi-step-ahead) controlled variable values and the optimizer calculates optimized values of the manipulated variable so as to closely follow the set-point trajectory of the controlled variable. Essentially, the MPC computes the current move of the manipulated input variable (Δu_i) to achieve the control goal.

The error to be minimized by the optimizer for achieving the control goal is the difference between the reference set point and the value of the control variable predicted by the model at each time step. For following the reference set-point trajectory the MPC minimizes an objective function, ' J ' of the form:

$$J(\Delta u_k) = \sum_{k=1}^{H_p} [\gamma(y_{k,r} - \hat{y}_k)^2 + \lambda \Delta u_k^2] \quad (5.1)$$

where, $y_{k,r}$ = reference trajectory values of the control variable at the k^{th} time instant, \hat{y}_k = model predicted value of the control variable at the k^{th} time instant, and H_p = prediction horizon. The control moves, $\Delta u_k, \Delta u_{k+1}, \dots, \Delta u_{k+H_c-1}$, are the optimum future (at discrete time intervals) changes in the input (manipulated) variable 'u' so that the controlled variable 'y' closely follow the reference trajectory, where, H_c = control horizon. The equation is applicable for a single input-single output (SISO) process, wherein, ' γ ' and ' λ ' are the penalizing constants for excess error and control moves respectively, which take into account the physical limitations of control system hardware. More details of the MPC and its implementations are given by Ogunnaike and Ray (1994) and Seborg et al. (2004).

5.3. STYRENE POLYMERIZATION PROCESS

The polymerization reactor is usually the heart of the polymer production process and very often is tedious to model due to reasons explained earlier. The continuous styrene polymerization reactor is a typical example of such a difficult-to-model chemical process. In order to develop GPSR-based models for the reactor, process input-output data were generated by employing the "first-principles" process model of an industrial scale styrene polymerization reactor (Sotomayor and Odloak, 2005). Figure (5.2) shows a schematic of the styrene polymerization reactor, wherein an exothermic, free-radical solution polymerization reaction is carried out in a jacketed continuous stirred tank reactor (CSTR). The CSTR is fed with the monomer (pure styrene), an initiator (azobisisobutyronitrile) and a solvent (benzene), at their constant steady-state rates.

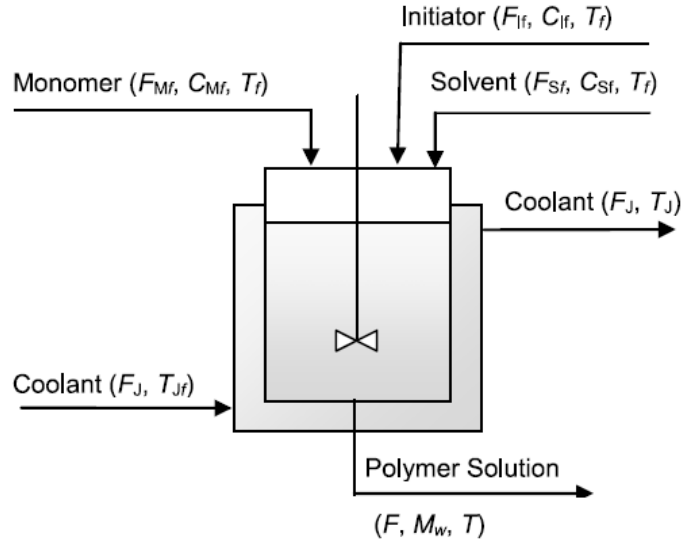


Figure 5.2: Schematic of the continuous polystyrene reactor process.

The initiator initiates a sequence of reactions to produce the polystyrene molecules, which are characterized by their average molecular weight. The equations describing the dynamics of the styrene polymerization reactor are given below.

$$\frac{dC_I}{dt} = \frac{F_I C_{If} - FC_I}{V} - k_I C_I \quad (5.2)$$

$$\frac{dC_M}{dt} = \frac{F_M C_{Mf} - FC_M}{V} - k_P C_M C_P \quad (5.3)$$

$$\frac{dT}{dt} = \frac{F(T_f - T)}{V} + \frac{(-\Delta H_R)k_P C_M C_P}{\rho C_p} - \frac{UA(T - T_J)}{\rho C_p} \quad (5.4)$$

$$\frac{dT_J}{dt} = \frac{F_J(T_{Jf} - T_J)}{V_J} + \frac{UA(T - T_J)}{\rho_J C_{pJ} V_J} \quad (5.5)$$

$$\frac{dD_0}{dt} = 0.5k_t C_P^2 - \frac{FD_0}{V} \quad (5.6)$$

$$\frac{dD_1}{dt} = M_M k_P C_M C_P - \frac{FD_1}{V} \quad (5.7)$$

$$M_W = \frac{D_1}{D_0} \quad (5.8)$$

$$C_P = \sqrt{\frac{2f_i k_i C_I}{k_T}} \quad (5.9)$$

$$k_i = k_{0i} e^{(-E_i/T)}, i = I, P, T \quad (5.10-5.12)$$

$$F = F_I + F_S + F_M \quad (5.13)$$

where, t = time and all the parameters/ steady-state variables and their values used in the simulation are listed in Table (5.1).

Table 5.1: Process parameters/variables of the phenomenological styrene polymerization reactor model

Sr. No.	Variable	Description	Parameter/Steady-state variable value	Unit
1	C_I	Initiator concentration in reactor	6.6832×10^{-2}	mol/L
2	C_{If}	Initiator feed concentration	0.5888	mol/L
3	C_M	Monomer concentration in reactor	3.3245	mol/L
4	C_{Mf}	Monomer feed concentration	8.6981	mol/L
5	D_0	Concentration of the dead polymer chains (moles)	2.7547×10^{-4}	mol/L
6	D_1	Concentration of the dead polymer chains (mass)	16.110	g/L
7	E_I	Activation energy for initiation reaction	14897	K
8	E_P	Activation energy for propagation reaction	3557	K
9	E_T	Activation energy for termination reaction	843	K
10	f_i	Initiator efficiency	0.6	
11	F_I	Initiator feed flowrate	108	L/hr
12	F_J	Jacket fluid feed flowrate	471.6	L/hr
13	F_M	Monomer feed flowrate	378	L/hr
14	F_S	Solvent feed flowrate	459	L/hr
15	k_{0I}	Arrhenius equation frequency factor for initiation reaction	2.142×10^{17}	h^{-1}
16	k_{0P}	Arrhenius equation frequency factor for propagation reaction	3.816×10^{10}	L/mol hr
17	k_{0T}	Arrhenius equation frequency factor for termination reaction	4.50×10^{12}	L/mol hr
18	MW	Monomer molecular weight	104.14	g/mol
19	T	Reactor mass temperature	323.56	K
20	T_f	Feed temperature	330	K
21	T_J	Jacket fluid temperature	295	K
22	UA	Overall heat transfer coefficient \times heat transfer area of reactor	2.52×10^5	cal/K hr
23	$-\Delta H_R$	Heat of polymerization reaction	16700	cal/mol
24	η	Intrinsic viscosity of polymer solution	2.9091	L/g

Sr. No.	Variable	Description	Parameter/Steady-state variable value	Unit
25	ρC_p	Mean density of reactor fluid \times mean heat capacity of reactor fluid	360	cal/K L
26	$\rho_j C_{pJ}$	Density of cooling jacket fluid \times heat capacity of cooling jacket fluid	966.3	cal/K L

The effluent polymer molecular weight (M_w) is the most important indicator of the final polymer product quality and, thus, need to be controlled tightly in any industrial polymerization process (Yoon et al., 2004). However, due to the multiple sequential and simultaneous reactions occurring in polymerization processes the final polymer product is composed of chains of polymer molecules of varying molecular weights. Thus, the quality of the final polymer product is estimated from its average molecular weight. The traditional analytical methods of estimating the molecular weight distribution and, thus, the average molecular weight of a polymer product include gel-permeation or size-exclusion chromatography, which require costly instruments and time-consuming procedures; thus, their readings are available off-line (Richards and Congalidis, 2006). The average molecular weight of the resultant polymer and, the corresponding polymer yield and quality are highly sensitive to the polymerization operating conditions. The most important operating parameter that significantly affects polymerization is the reaction temperature (Richards and Congalidis, 2006). Accordingly, we have developed GPSR-based soft sensor models relating the molecular weight of the polystyrene polymer produced M_w , to the coolant feed flowrate, F_j .

5.3.1. Process Identification of Styrene Polymerization Process Using GPSR Methodology

For capturing the dynamics of the styrene polymerization process using the GPSR methodology, a discrete time single input-single output (SISO) model of the following form was used.

$$M_{W,k+1} = f(M_{W,k}, M_{W,k-1}, M_{W,k-2}, F_{J,k}, F_{J,k-1}, F_{J,k-2}) \quad (5.14)$$

where, $M_{W,k+1}$ is a one-step ahead predicted value of the effluent polymer molecular weight (controlled/output) variable, subscript 'k' is the sampling instant, $F_{J,k}$ is the current/past values of the coolant flowrate (manipulated/input variable) to the jacket, f denotes the functional relationship between the current and past (time lagged) values of the input and the future predicted output variable value.

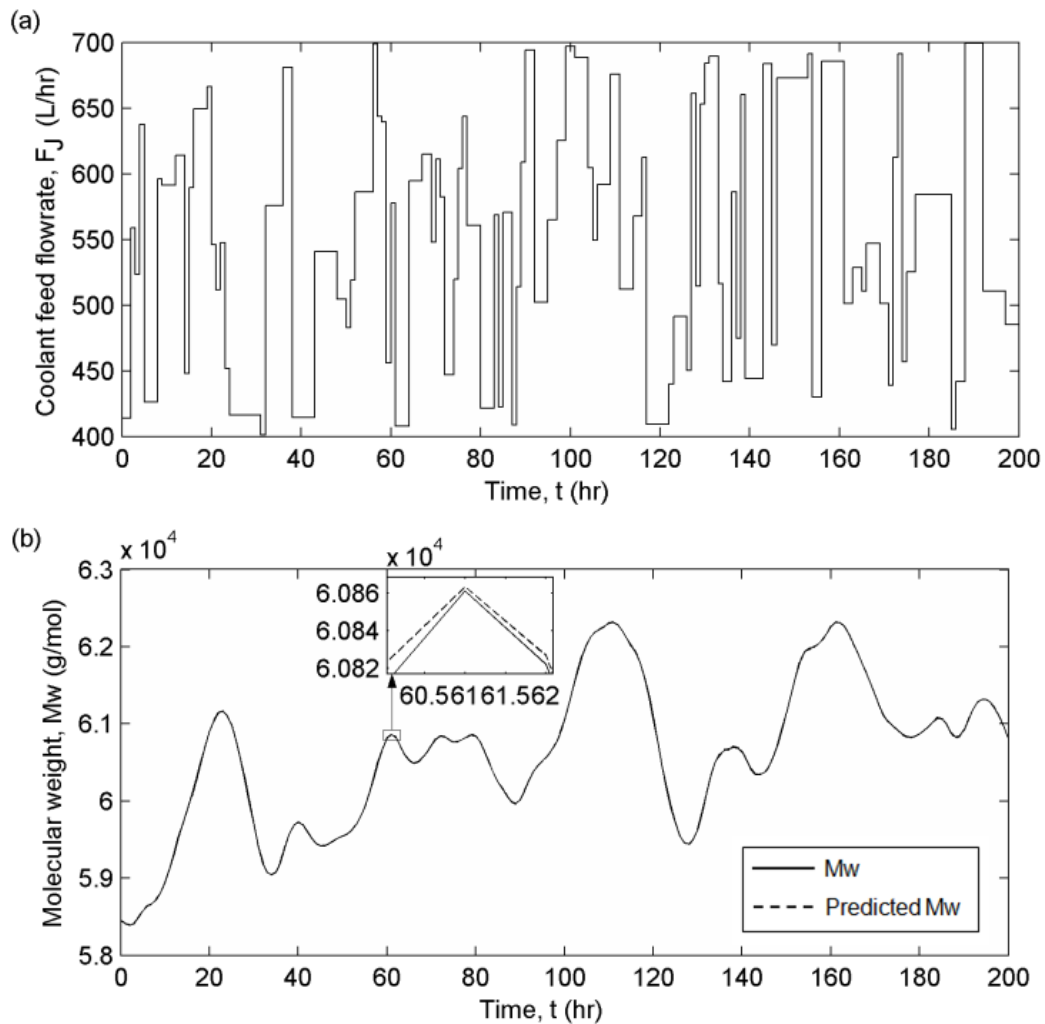


Figure 5.3: Stimulus-response profiles of the polystyrene reactor process, (a) random sequence in input (coolant feed-flowrate), (b) response of effluent polymer molecular weight and the GPSR-Model 1 predicted polymer molecular weight (inset: enlarged view).

A large-sized dataset, composed of 10000 data-points were generated by imparting a pseudo-random sequence in F_J with a sampling time of 1 hr and switching probability of $P_s = 0.5$ to the process model. It may be noted that in real practice, the process dataset should be collected by physically disturbing the

process in open loop mode. In the absence of such data, in the present study, the phenomenological model of styrene polymerization has been used for process data generation and demonstrating the GPSR based soft-sensor's utility in implementing an MPC. A small portion of the input and the resulting output sequences are shown in Figure (5.3). The total dataset was partitioned into training and test sets in the ratio 3:1 for the development of the GPSR-based models. The GPSR-based model was built using *Eureqa Formulize* software (Schmidt and Lipson, 2009). A number of variants of GPSR-based soft-sensor models were generated by systematically varying the GPSR algorithmic parameters. The performance evaluation of these models was done using standard statistical parameters namely, *coefficient of correlation (CC)* and *root mean squared error (RMSE)* between the model-predicted molecular weight values and their desired magnitudes. Among the several GPSR models generated, the two best models (i.e., possessing highest *CC* and lowest *RMSE* magnitudes) are as follows:

GPSR-Model 1:

$$M_{Wn,k+1} = 2.263M_{Wn,k} + 0.4226M_{Wn,k-2} + 0.01471F_{Jn,k} + 0.002204F_{Jn,k-1} - 0.0005894 - 1.698M_{Wn,k-1} - 0.001905F_{Jn,k}^2 \quad (5.15)$$

GPSR-Model 2:

$$M_{Wn,k+1} = 1.985M_{Wn,k} + \frac{F_{Jn,k}}{(58.53 + F_{Jn,k-2})} - M_{Wn,k-1} \quad (5.16)$$

where, M_{Wn} and F_{Jn} are the normalized forms of M_W and F_J respectively, expressed as:

$$M_{Wn} = \frac{M_W - M_{W,\min}}{M_{W,\max} - M_{W,\min}} \quad (5.17)$$

$$F_{Jn} = \frac{F_J - F_{J,\min}}{F_{J,\max} - F_{J,\min}} \quad (5.18)$$

where, $M_{W,\min} = 57362$, $M_{W,\max} = 63201$, $F_{J,\min} = 400.05$, and $F_{J,\max} = 699.95$ are the scaling parameters. Figure (5.3) shows the random input step sequence of the manipulated variable, F_J , and the resulting trajectory of the controlled variable,

M_w (represented by a continuous line), and its GPSR-Model 1 (Eqn. 5.15) predicted values (represented by a dashed line).

5.4. RESULTS AND DISCUSSION

5.4.1. Performance Analysis of the GPSR Based Soft Sensor Models

The GPSR-based soft sensors (Eqns. 5.15 and 5.16) relate the current and lagged values of average molecular weight of the effluent polymer (that is difficult-to-measure online), to the current and lagged values of the coolant feed flowrate. Both the GPSR models exhibit good prediction and generalization performances, thus accurately modeling the dynamic behavior of the styrene polymerization process. Especially, the GPSR-Model 1 shows the best performance with a CC of more than 0.999 and $RMSE$ of around 9×10^{-4} (see Table (5.2)) for both the training and test sets. Same can also be inferred from Figure (5.3(b)) in which it is observed that the predicted (by the GPSR-based soft sensor model) ' M_w ' trajectory (indicated by dashed line) closely follows the actual molecular weight time-trajectory (indicated by continuous line). This clearly demonstrates that the GPSR based model has been able to accurately and effectively identify the strong nonlinear dynamics possessed by the continuous styrene polymerization process. The notable feature of GPSR-based process models is that they represent mathematical relations in a simpler form as compared to other CI-based models such as ANNs, SVR and FL. The use of such simple models considerably reduces the computational load on the MPC during its implementation.

Table 5.2: Prediction performance of the GPSR-based soft-sensor models for styrene polymerization process

Model	Training-set		Test-set	
	CC_{Trn}	$RMSE_{Trn}$	CC_{Tst}	$RMSE_{Tst}$
GPSR-Model 1	0.999	9.2×10^{-4}	0.999	9×10^{-4}
GPSR-Model 2	0.999	3.6×10^{-3}	0.999	3.5×10^{-3}

5.4.2. Model Predictive Control of Styrene Polymerization Process Using GPSR-Based Soft-Sensor Model

The MPC scheme designed for the industrial-scale styrene polymerization process (see Figure (5.2)) was implemented for controlling the molecular weight, M_w , of the effluent polystyrene (which is difficult to measure online) by adjusting the coolant feed-flowrate F_J to the jacket. The jacket coolant feed-flowrate (F_J) is found to directly affect the reactor temperature and, consequently, the effluent polystyrene molecular weight. The molecular weight of the effluent polystyrene has a strong relationship with the intrinsic viscosity of the effluent polymer solution, according to the equation (Hur, 2003),

$$\eta = 0.0012M_w^{0.71} \quad (5.19)$$

The intrinsic viscosity, η (L/g), of the effluent polystyrene solution can be easily estimated online using a continuous viscosity analysis instrument and, thus, it is used as a direct measure of M_w and, consequently, as the controlled variable. The MPC scheme uses GPSR-Model 1 (Eqn. 5.15) for predicting the future values of M_w (η) over the prediction horizon of $H_p = 10$; this is done using values of η and F_J at the current and past time instants. The MPC then determines an optimized sequence of the future moves of F_J , over the control horizon $H_c = 4$. The prediction and the control horizons are the number of time steps in the near future over which the respective predictive calculations are done. The weight factors of the MPC objective function were heuristically determined to be $\gamma = 1$ and $\lambda = 0.015$. From all the future calculations, only the first control move was implemented following which the future model predictions and control moves were re-computed based on the new conditions. The MPC scheme utilizing the GPSR-based process model was simulated for two case studies, namely set-point tracking, and disturbance rejection. The control performance of the GPSR-based model utilizing MPC scheme was compared with that of an appropriately tuned PID controller. The PID controller was fine-tuned using the Cohen-Coon settings (Cohen and Coon, 1953) for stable operation.

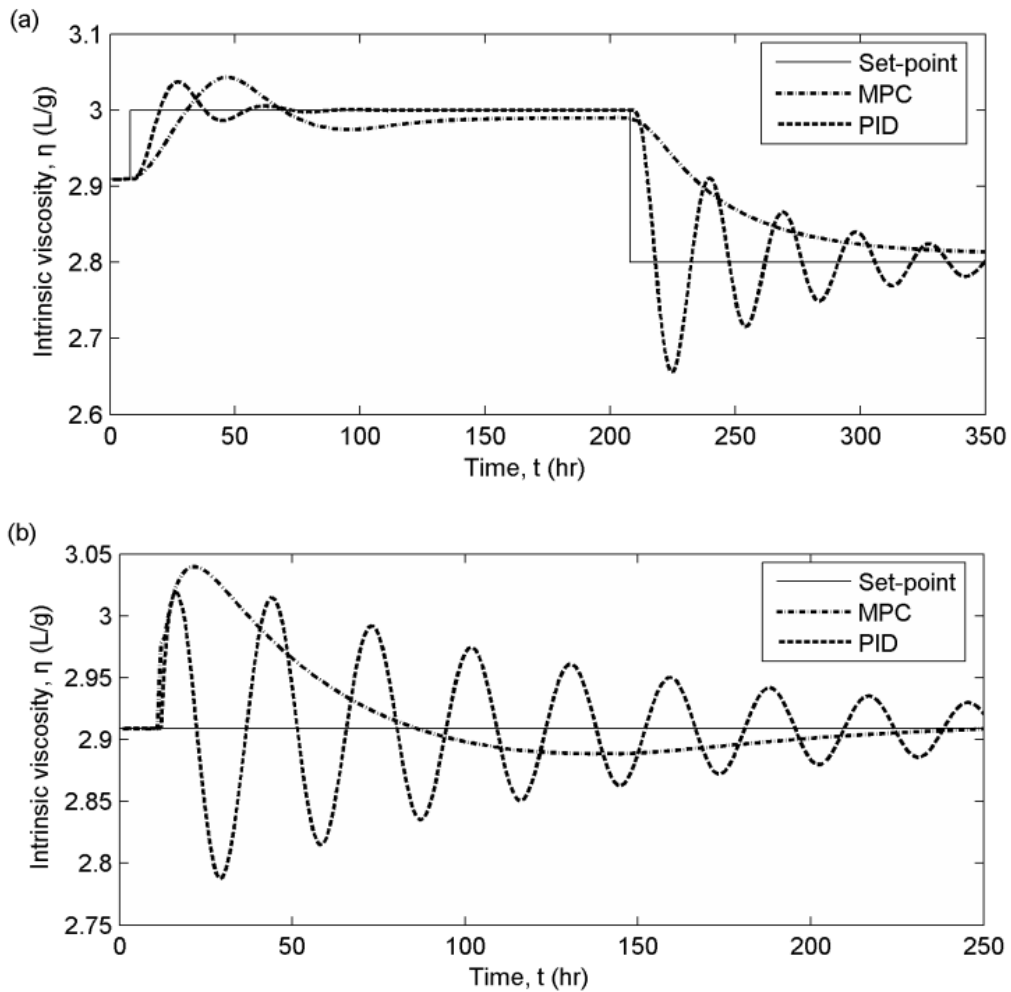


Figure 5.4: Controlled variable profiles of the simulated MPC and PID control schemes for the polystyrene reactor process, (a) Set-point tracking, (b) Disturbance rejection.

(a) Set-point Tracking

Polymer grade-transitions during the operation of continuous polymerization reactors are often required to produce polymers of different grades. In the transitory period, the nonlinear process behavior becomes more pronounced (Srinivasan et al., 2005). Such grade transition phases can be followed quite well by an MPC scheme. An MPC-based grade transition scheme has been implemented using the GPSR-based soft-sensor model. This was done by simulating sudden changes in the process set-point over time as shown in Figure (5.4(a)). The set-point value of the intrinsic viscosity (controlled variable) of the effluent polymer was increased from its steady state value of

2.9091 L/g to 3 L/g and then decreased to 2.8 L/g over time. From the actual intrinsic viscosity profile, it is seen that the MPC has taken appropriate action to accommodate the set-point changes by adjusting F_J , thereby following the imparted set-point. The PID controller takes a shorter time to reach the higher set-point magnitude; however, it exhibits greater fluctuations in doing so. In the case of a lowered set-point also, the PID controller shows high fluctuations in the values of the controlled variable while reaching the desired set-point. In comparison, the MPC controller follows a smooth path towards the set-point in both cases.

(b) Disturbance Rejection

In the disturbance rejection case study, the feed monomer concentration was abruptly increased (step input) from its steady state value of 3.3245 mol/L to 3.7 mol/L at 20 hr as shown in Figure (5.4(b)). This step input caused a disturbance in the process behavior in the form of a sudden rise in the intrinsic viscosity of the effluent polymer at that time instant (see Figure 5.4(b)). In response, the MPC takes immediate action to suppress the disturbance by appropriately manipulating F_J over time so as to bring back the process to its set-point (steady-state) allowing only minor oscillations. In comparison, the PID controller drives the controlled variable to its set value in a highly oscillatory manner spread over a longer period of time.

5.5. CONCLUSION

A novel approach of using the CI-based genetic programming based symbolic regression (GPSR) methodology for process identification of a complex, nonlinear dynamic chemical process system is presented in this work. Styrene polymerization is such a complex, nonlinear process that is difficult to model phenomenologically. Accordingly, GPSR-based soft sensor dynamic models for an industrial scale styrene polymerization reactor were developed. The developed soft-sensor models dynamically relate the “difficult to measure” output variable (average molecular weight of the effluent polystyrene) to the important input variable (jacket coolant flowrate) of the process. The GPSR-based soft sensor models possess excellent prediction and generalization performances as

indicated by the high CCs of 0.999 and low $RMSEs$ ($<9.3 \times 10^{-4}$) for both the training and test sets. The best GPSR-based model (possessing highest prediction and generalization performance) was used in a simulated model predictive control (MPC) scheme for the continuous styrene polymerization reactor. The MPC scheme was found to perform better in both set-point tracking and disturbance rejection tasks than an appropriately tuned PID controller. The proposed approach of GPSR-based modeling can be gainfully applied to model any linear/nonlinear phenomena for utilization in model based control of nonlinear chemical processes.

NOMENCLATURE

A	Heat transfer area of reactor
C_I	Initiator concentration in reactor
C_{If}	Initiator feed concentration
C_M	Monomer concentration in reactor
C_{Mf}	Monomer feed concentration
C_p	Mean heat capacity of reactor fluid
C_{pJ}	Heat capacity of cooling jacket fluid
D_0	Concentration of the dead polymer chains (moles)
D_I	Concentration of the dead polymer chains (mass)
E_I	Activation energy for initiation reaction
E_P	Activation energy for propagation reaction
E_T	Activation energy for termination reaction
f_i	Initiator efficiency
F_I	Initiator feed flowrate
F_J	Jacket fluid feed flowrate
F_M	Monomer feed flowrate
F_S	Solvent feed flowrate

J	Least squared objective function
k_{0I}	Arrhenius equation frequency factor for initiation reaction
k_{0P}	Arrhenius equation frequency factor for propagation reaction
k_{0T}	Arrhenius equation frequency factor for termination reaction
M_w	Monomer molecular weight
t	Time
T	Reactor mass temperature
T_f	Feed temperature
T_j	Jacket fluid temperature
u	Input (manipulated) variable
U	Overall heat transfer coefficient
$-\Delta H_R$	Heat of polymerization reaction
y	Output (controlled) variable

Greek letters

ρ	Mean density of reactor fluid
ρ_j	Density of cooling jacket fluid
η	Intrinsic viscosity of polymer solution
γ	Penalizing constants for excess error
λ	Penalizing constants for excess control moves

REFERENCES

Bindlish, R. (2015). Nonlinear model predictive control of an industrial polymerization process. *Computers & Chemical Engineering* 73:43-48.

Cohen, G., Coon, G. A. (1953). Theoretical consideration of retarded

control. *Trans. Asme*, 75:827-834.

de Assis, A. J., Maciel Filho, R. (2000). Soft sensors development for on-line bioreactor state estimation. *Computers & Chemical Engineering* 24(2-7):1099-1103.

Desai, K., Badhe, Y., Tambe, S. S., Kulkarni, B. D. (2006). Soft-sensor development for fed-batch bioreactors using support vector regression. *Biochemical Engineering Journal* 27(3):225-239.

Devogelaere, D., Rijckaert, M., Leon, O. G., Lemus, G. C. (2002). Application of feedforward neural networks for soft sensors in the sugar industry. In: *Proceedings of Neural Networks, IEEE SBRN VII Brazilian Symposium*, pp. 2-6.

Feng, R., Shen, W., Shao, H. (2003). A soft sensor modeling approach using support vector machines. In: *Proceedings of the American Control Conference, IEEE*, vol. 5, pp. 3702-3707.

Ghugare, S. B., Tiwary, S., Elangovan, V., Tambe, S. S. (2014). Prediction of higher heating value of solid biomass fuels using artificial intelligence formalisms. *BioEnergy Research* 7(2):681-692.

Hosen, M. A., Hussain, M. A., Mjalli, F. S. (2011). Control of polystyrene batch reactors using neural network based model predictive control (NNMPC): An experimental investigation. *Control Engineering Practice* 19(5):454-467.

Hur, S. M., Park, M. J., Rhee, H. K. (2003). Polymer property control in a continuous styrene polymerization reactor using model-on-demand predictive controller. *Korean Journal of Chemical Engineering* 20(1):14-21.

Koza, J. R. (1994) *Genetic Programming II*. MIT Press, Cambridge, Massachusetts, USA.

Li, S. J., Zhang, X. J., Qian, F. (2005). Soft sensing modeling via artificial neural

- network based on PSO-Aloplex. In: Proceedings of Machine Learning and Cybernetics, IEEE International Conference, vol. 7, pp. 4210-4215.
- Meleiro, L. A. D. C., Finho, R. M. (2000). A self-tuning adaptive control applied to an industrial large scale ethanol production. *Computers & Chemical Engineering* 24(2):925-930.
- Na, S. S., Rhee, H. K. (2000). Polynomial ARMA model identification for a continuous styrene polymerization reactor using on- line measurements of polymer properties. *Journal of applied polymer science* 76(13):1889-1901.
- Na, S. S., Rhee, H. K. (2002). An experimental study for property control in a continuous styrene polymerization reactor using a polynomial ARMA model. *Chemical engineering science* 57(7):1165-1173.
- Ogunnaike, B. A., Ray, W. H. (1994). *Process dynamics, modeling, and control*. Oxford University Press, New York.
- Richards, J. R., Congalidis, J. P. (2006). Measurement and control of polymerization reactors. *Computers & chemical engineering* 30(10):1447-1463.
- Schmidt, M., Lipson, H. (2009). Distilling free-form natural laws from experimental data. *Science* 324:81–85.
- Seborg, D. E., Edgar, T. F., Mellichamp, D. A. (2004). *Process Dynamics and Control*. Wiley & Sons, New York.
- Sotomayor, O. A., Odloak, D. (2005). Observer-based fault diagnosis in chemical plants. *Chemical Engineering Journal* 112(1):93-108.
- Srinivasan, R., Viswanathan, P., Vedam, H., Nochur, A. (2005). A framework for managing transitions in chemical plants. *Computers & chemical engineering* 29(2):305-322.
- Yan, W., Shao, H., Wang, X. (2004). Soft sensing modeling based on support vector machine and Bayesian model selection. *Computers & Chemical*

Engineering 28(8):1489-1498.

Yoon, W. J., Kim, Y. S., Kim, I. S., Choi, K. Y. (2004). Recent advances in polymer reaction engineering: modeling and control of polymer properties. Korean Journal of Chemical Engineering 21(1):147-167.

CHAPTER 6.

COMPUTATIONAL INTELLIGENCE BASED PERFORMANCE MODELING OF CO- GASIFICATION PROCESS USING DIFFERENT COAL-BIOMASS BLENDS

COMPUTATIONAL INTELLIGENCE BASED PERFORMANCE MODELING OF CO-GASIFICATION PROCESS USING DIFFERENT COAL-BIOMASS BLENDS

Abstract

Biomass, a renewable fuel is a cleaner and partial alternative to coal. Compared to commonly used combustion, *Co-gasification* (COG) is a relatively new technology that uses a binary fuel (coal-biomass) blend for generating a product gas mixture (syngas). It has found numerous industrial applications. A majority of currently mined Indian coals contain high ash percentages. To evaluate their efficacy for co-gasification, a rigorous experimental program was undertaken in a *fluidized-bed co-gasifier* (FBCOG) pilot plant operated at a gasification temperature between 850 and 1014°C. The high ash coals were blended with three types of biomasses, namely, rice husk, sawdust, and press-mud. Since it uses a blend of two naturally occurring solid fuels the FBCOG operation involves numerous solid and gas phase reactions, and associated complex and nonlinear mass and heat transfer phenomena. Thus, the process is difficult to model phenomenologically. Accordingly, process data as generated from the pilot-plant experiments were used for developing *computational intelligence* (CI)-based models, using three formalisms namely, *genetic programming* based *symbolic regression* (GPSR), *multilayer perceptron neural network* (MLPNN), and *support vector regression* (SVR). Each of these methods, was employed to predict four co-gasification performance variables, namely, the *total gas yield* (kg/kg fuel), *carbon conversion efficiency* (%), *heating value of the product gas* (MJ/Nm³) and *cold gas efficiency* (%). The developed CI-based models possess excellent prediction accuracy and generalization performance. The experimental data and the CI-based models developed in this study can be gainfully employed in the design and operation of the co-gasifiers using high ash coals.

6.0. INTRODUCTION

Coal is the most widely used fossil fuel for the production of thermal and electrical energies. However, its overconsumption is leading to increasing emissions of greenhouse gases and soil and water body contamination due to the storage and release of huge amounts of ash (post-combustion residue). This is affecting the climate and environment enormously (Zhou et al., 2016). In comparison, biomass, a renewable energy source, is considered to be a carbon-neutral fuel and thus has a significant potential to fulfill the future energy needs. Co-gasification takes advantages of the synergy between coal and biomass gasification, wherein the process not only leaves a low carbon footprint on the environment, but also improves the hydrogen-to-carbon monoxide generation ratio in the product gas as desirable for liquid fuel synthesis (Andre et al., 2005). Additionally, the inorganic matter (mainly alkaline earth metals) present in the biomass is known to catalyze the gasification of coal to produce clean gaseous fuels.

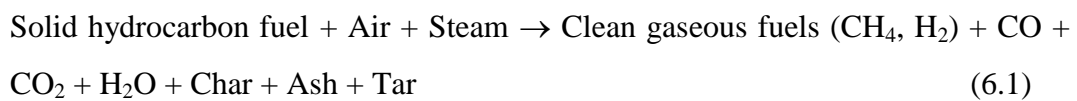
The tendency of producing higher carbon footprint, fast depleting reserves of fossil fuels, and accumulation of biomass wastes augmented by the growing energy requirements have created an urgent need for the development of *clean coal technologies* (CCTs). Such newer approaches to coal gasification basically aim at innovative, efficient and cleaner techniques by means of:

- Changes in the feed materials: The use of alternative fuel (non-coal) feedstock for gasification includes the use of biomasses and carbonaceous wastes or its blends with coal.
- Changes in the processing techniques: The main focus here is on modifications in the fuel firing techniques and bed operating techniques.

Co-gasification utilizes biomass-blended coal fuels as feedstock. Additionally, the fluidized bed operation of a co-gasification process ensures more control on the gasification reactions thereby contributing to increased efficiencies and reduced pollutions as discussed in the next sub-section.

6.0.1. Co-gasification (COG)

Co-gasification (COG) is a technology that thermally (at elevated temperatures) and chemically (under oxygen-starved conditions) converts the feed hydrocarbon materials into a product gas mixture (syngas) comprising mainly hydrogen, carbon monoxide, methane, carbon dioxide and nitrogen (when the air is used as gasifying agent) (Stevens, 2011). The process of conversion of the carbonaceous solid material to the desired gaseous mixture can be written in terms of material in and out as,



Co-gasification adds operational flexibility to the traditional coal-gasification process by enabling the use of biomass or other carbonaceous materials (industrial waste, sludge etc.) in varying proportions. Additionally, the COG process to some extent mitigates the problem of air and water pollution/contamination due to reduced release of the volatile organic compounds (VOCs) and less generation of tar as in the case of coal gasification (Rezaiyan and Cheremisinoff, 2005).

In a number of countries such as India, China, Australia, and Turkey large deposits of coals containing high percentages of ash (and in turn mineral matter) are present. The Indian thermal power stations invariably receive such type of coals and the pollution problems associated with them have become a major concern. For achieving the stringent pollution control targets, changes in the coal utilization practices have become imperative. Accordingly, co-gasification of coal-biomass blends is suggested to tackle the stated problematic issues. These measures are expected to result in a high conversion efficiency and lower environmental impact (Takematsu and Maude, 1991).

The use of fluidized bed gasifier for co-gasification has several advantages (Patil-Shinde et al., 2014; Chavan, 2012):

- The fluidized bed co-gasifier (FBCOG) allows flexibility of handling a wide range of coals and biomass materials (such as those obtained from

agricultural/industrial/forest sources including wastes, as well as their blends).

- It provides good mixing and contact between the fuel and gasifying agent leading to better heat and mass transfer in the reactor.
- Reduced formation of tar and phenols as compared to coal gasification.
- Allows usage of bigger particle sizes with higher residence times of the solid fuel particles.
- It is capable of handling high-sulfur coals without a need for a separate flue-gas desulfurization process, which increases the capital and operational costs.
- The FBCOG process operates at relatively lower temperatures, thus emitting lower amounts of nitrogen oxides (NO_x) and providing more energy efficiency.

Due to these several advantages, there are continuously increasing research studies involving COG processes in fixed/fluidized bed gasifiers utilizing region-specific coals and biomasses/carbonaceous materials/wastes (Jong et al., 1999; Sjostrom et al., 1999; Pan et al., 2000; Pinto et al., 2003; McLendon et al., 2004; Andre et al., 2005; Velez et al., 2009; Mastellone et al., 2010; Aigner et al., 2011; Sulaiman et al., 2017; Zhang and Bi, 2010). Such a large number of co-gasification studies are indicative of its emergence as a potential technology for the new generation of power plants in terms of higher efficiencies, lesser pollution and use of alternate fuels in replacement to the fast depleting fossil fuels.

The performance of a typical co-gasification process is strongly affected by co-gasification process operating conditions (mainly gasifier temperature and flow patterns), feed material flow-rates (fuel feed-rate, air feed-rate, steam feed-rate), and physical and chemical properties of the feedstock. The important performance measure parameters of the co-gasification process are:

(i) *Total gas yield (TGY)*: This is an indicator of the fuel mass conversion efficiency of the process as given by the equation,

$$TGY\left(\frac{\text{kg}}{\text{kg fuel}}\right) = \frac{\text{Total product gas produced (kg/h)}}{\text{Solid fuel feed rate (kg/h)}} \quad (6.2)$$

(ii) *Carbon conversion efficiency (CCE)*: As the name specifies this parameter measures the percent conversion of carbon in the process and is given by the expression,

$$CCE(\%) = \left(1 - \left(\frac{\text{Carbon in bottom and cyclone ash}}{\text{Carbon in fuel feed}}\right)\right) \times 100 \quad (6.3)$$

(iii) *Heating value (HV)*: This performance parameter is a measure of the energy content of the product gas as given by equation,

$$HV\left(\frac{\text{MJ}}{\text{Nm}^3}\right) = \sum_{i=1}^3 w_i \times CV_i \quad (6.4)$$

where, w_i denotes the weight percentage of the gas component ‘ i ’, ($i = 1$ for H_2 , 2 for CO and 3 for CH_4) and CV_i indicates the calorific value of the respective gas component.

(iv) *Cold gas efficiency (CGE)*: This performance parameter ultimately measures the energy efficiency of the process as given by the expression,

$$CGE(\%) = \left(\frac{TGY \times CV_P}{CV_F}\right) \times 100 \quad (6.5)$$

where, TGY is the total gas yield (Eqn. 6.2), CV_P and CV_F are the calorific values of product gas and feed solid fuel respectively.

6.0.2. Co-gasification Modeling for Process Performance Evaluation: Need for CI-Based Models

The performance analysis of the FBCOG process can be done through the use of appropriately validated process models. Commonly, a phenomenological (“first principles” or “mechanistic”) process model is developed based on a

thorough understanding of the underlying physicochemical phenomena. In general, phenomenological modeling of co-gasification processes is difficult due to following reasons: (i) nonlinear interplay of numerous process variables, (ii) complex throughput dependent process dynamics, (iii) exhaustive and cost-intensive experimentation is required for studying the effects of influential feed and operating parameters, (iv) unavailability of an in-depth knowledge of the underlying physicochemical phenomena (e.g., kinetics, heat and mass transport mechanisms), and (v) the fluidization effects in a fluidized bed co-gasification (FBCOG) process lead to complex momentum transfer effects, which are difficult to understand in totality (Żogała, 2014). Despite these difficulties, a few attempts were made for phenomenological modeling of the COG (see Table (6.1)) process by making some simplifying assumptions. In general, two types of simplified phenomenological models, namely *thermodynamic* (equilibrium) and *kinetic* (rate), have been developed for fluidized bed coal gasifiers (Lee, 2007). The models based on the process thermodynamics are independent of the gasifier type as they assume complete oxygen consumption. Being independent of the gasifier type, these models are not useful for examining the effects of the operating parameters on the gasifier performance. The kinetic models comprise mainly of the kinetics of various reactions occurring in the gasification operation. Given a set of gasifier operating conditions its kinetic model is capable of predicting the process performance in terms of, for instance, product composition, and temperature profiles. Some notable representative studies as also reviews pertaining to the phenomenological modeling of fluidized bed gasifiers are given by Rhinehart et al. (1987), Sett and Bhattacharya (1988), de Souza-Santos (1989), Gururajan et al. (1992), Donne et al. (1998), Moorea-Taha (2000), Villanueva et al. (2008), Mazumder (2010), Armstrong et al. (2011), Yang et al. (2012), Xiangdong et al. (2013) and Singh et al. (2014). Based on the simplified phenomenological modeling approaches described previously few researchers have attempted to develop *co-gasification* process models, the details of which are listed in Table (6.1).

Table 6.1: Existing phenomenological models of co-gasification processes

Author	Type of Model	Gasifier	Feedstock	Objective
Yuehong et al. (2006)	Thermodynamic Model- Minimization of Gibbs free energy	Shaft furnace type gasifier	Coke and natural gas	Composition of Syngas
Perez-Fortes et al. (2009)	Aspen-Hysys	Pressurized Entrained Flow Gasifier	Petroleum Coke and Olive Pomace (Orujilo)	Feasibility of COG, Gas Composition and Power Generation
Shen et al. (2012)	Aspen Plus (Minimization of total Gibbs free energy-Equilibrium Model)	Entrained bed gasifier	Coal (Sub-bituminous, Bituminous) and Petroleum coke	Concentration distribution of CO, H ₂ and CO ₂
Gartner et al. (2012)	1-D Kinetic Model	Entrained Flow gasification	Australian coal and char	Carbon conversion and cold gas efficiency
Cormos (2013)	ChemCAD and Thermoflex (Minimization of Gibbs Free energy-Equilibrium Model)	Siemens gasifier (entrained-flow type)	Coal (Romanian type) and biomass/solid waste	Co-gasification Performance indicators namely H ₂ , SNG and Liquid fuels

Commonly, empirical modeling in the form of linear/nonlinear regression is employed when phenomenological modeling is difficult. The *empirical* modeling approaches mainly include regression-based modeling strategies that avoid difficulties of first principles based modeling as they are exclusively based on the process data (data-driven modeling). The regression-based black-box modeling methods conventionally utilize common linear/nonlinear fitting functions such as linear, polynomial, power equations etc. A major difficulty in empirical modeling is pre-specification of an appropriate data-fitting function, which poses tremendous difficulties and requires a huge trail and error approach that very often doesn't succeed, especially for nonlinear and complex systems

such as FBCOG. The difficulties and complexities involved in the phenomenological and classical empirical modeling of FBCOG necessitate exploration of alternative nonlinear modeling strategies that do not require full details of the underlying physicochemical phenomena. Such an alternative has become available in recent years in the form of computational intelligence (CI) based exclusively data-driven modeling approaches. Earlier, several gasification (coal or biomass) modeling studies using CI-based methodologies (mostly ANNs) have been performed (Nougues et al., 2000; Mjalli and Al-Mfargi, 2008; Chavan et al., 2012; Puig-Arnavat et al., 2013; Patil-Shinde et al., 2014). Despite possessing attractive properties, CI-based methods have not been used in COG process modeling. Accordingly, this study presents results of FBCOG modeling using three CI-based methods, namely, *genetic programming based symbolic regression* (GPSR) (Koza, 1992), *multilayer perceptron neural network* (MLPNN) (Freeman and Skapura, 1991), and *support vector regression* (SVR) (Vapnik, 1995). Specifically, models have been developed using each of the stated CI-based method for predicting four FBCOG performance variables stated earlier. The CI-based models developed here, being inherently nonlinear, have accurately captured the complexities and nonlinearities involved in the fluidized bed co-gasification process.

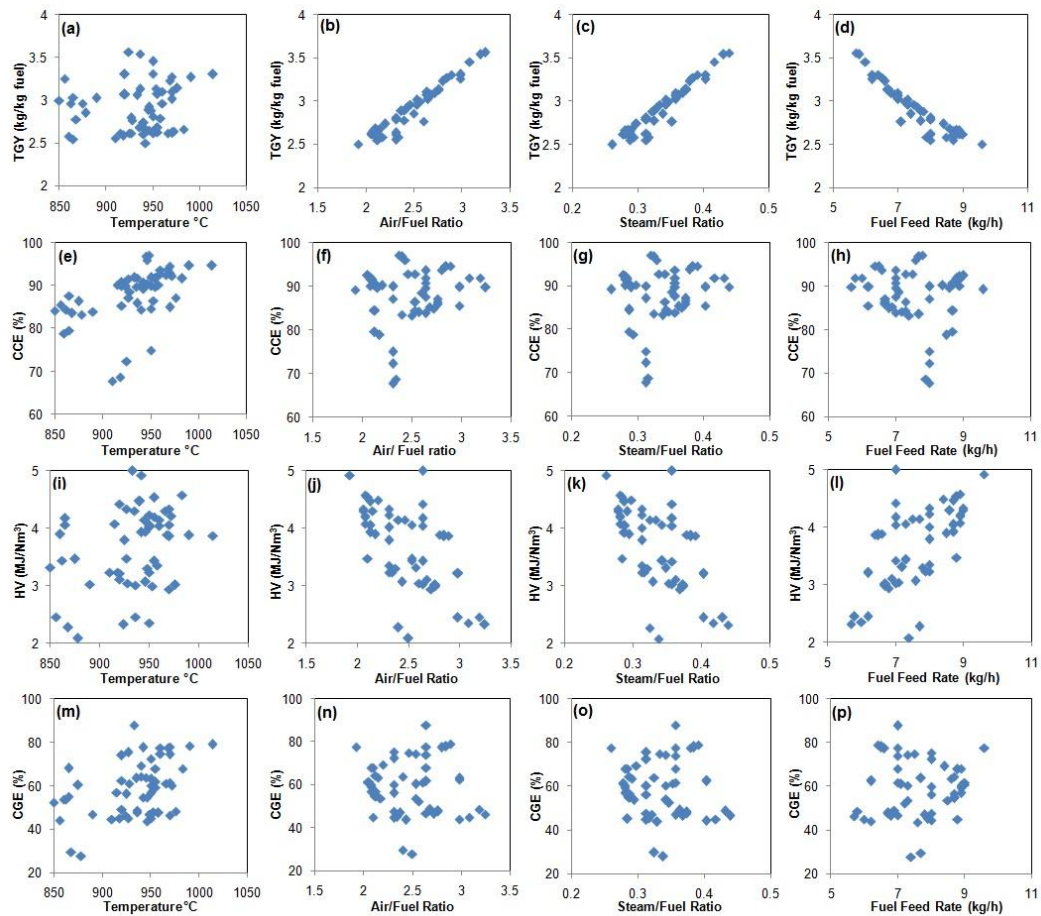


Figure 6.1: Scatter-plots of FBCOG performance variables (*TGY*, *CCE*, *HV* and *CGE*) versus important feed/process operating parameters (temperature, air/fuel ratio, steam/fuel ratio and fuel feed rate).

Nature of Dependencies between Performance and Important Operating Variables

Before conducting any data-driven modeling, it is advisable to examine the type of dependencies (linear/nonlinear) that exist between the dependent and independent (predictors) variables and parameters of the models. Accordingly, to analyze the relations between the four performance variables and the important independent (feed and operating) variables of the FBCOG process, scatter-plots were generated as shown in Figure (6.1) using the experimental dataset. Specifically, each of the four performance variables, namely *TGY*, *CCE*, *HV* and *CGE* was plotted vis-à-vis four important FBCOG process operating parameters, namely gasifier temperature (*T*), air/fuel ratio (*R_{AF}*), steam/fuel ratio (*R_{SF}*) and fuel

feed rate (f). From Figure (6.1) it is observed that TGY exhibits highly nonlinear behavior with respect to the reactor temperature and a mildly nonlinear behavior with respect to the fuel feed rate. From the variation of the remaining performance variables, namely, CCE , HV and CGE with respect to the four FBCOG process operating parameters considered while plotting the figure, high amount of nonlinear dependencies are observed (see Figure 6.1 panels ($e-p$)) between them. The stated nonlinearities can be effectively captured by the CI-based inherently nonlinear modeling methodologies, namely *genetic programming* based *symbolic regression* (GPSR), *multilayer perceptron neural network* (MLPNN), and *support vector regression* (SVR). Additionally, these being exclusively data-driven modeling methods can develop accurate models using the experimental dataset, without requiring the full details of the underlying mechanisms/phenomena. Accordingly, the principal objectives of this study are: (i) performance assessment of a fluidized bed co-gasification process (FBCOG) conducted in a pilot plant using high ash Indian coal and different biomasses blended in various proportions, (ii) development of CI-based process models for the prediction of four COG performance variables, namely, TGY (kg/kg fuel), CCE (%), HV (MJ/Nm³) and CGE (%) using thirteen feed and process operating parameters (these are described in detail below), (iii) to study the effect of three types of biomasses and their blending ratios with high ash coals on the performance of the FBCOG process.

The significance of a total of thirteen process feed and operating variables/parameters considered in the CI-based modeling is described below:

- *Gasifier temperature* (T) (°C) has a significant effect on the product gas generation per kg of fuel. Higher process temperatures result in faster pyrolysis, generating an increased amount of CO₂ gas that gets converted to CO via the Boudouard reaction.
- *Coal feed rate* (f_c) (kg/h) and *biomass feed rate* (f_b) (kg/h) define the flow rate of the carbonaceous constituents of the feed and, thus, affect the residence time of the feed solid fuel particles in the bed.
- Four feed variables, namely, *fixed carbon* (F_C), *volatile matter* (V_M), *ash* (A_S) and *moisture* (M), (all in weight percentages) belonging to the *proximate*

analysis, provide a primary assessment of the feed fuel's quality. While the extent of F_C positively affects the quality of the product gas in the co-gasification reaction, A_S negatively affects the reaction; V_M plays a significant role in determining the reactivity of the fuel.

- *Air-to-fuel ratio* (R_{AF}) (kg/kg fuel) is an important process parameter since the air-assisted oxidation of the carbon is one of the key reactions for attaining the desired co-gasification temperature; however, using a high air-to-fuel ratio decreases the heating value of the product gas thereby negatively affecting the performance of co-gasification process.
- The importance of *steam-to-fuel ratio* (R_{SF}) (kg/kg fuel) is that with its increase, the production of H_2 increases due to the elevated partial pressure of H_2O inside the gasification chamber, favoring the important water-gas, water-gas shift, and methane reformation reactions.
- The *ultimate analysis* of the feed represents the elemental composition of the organic portion of the solid fuel in terms of the weight percentages of major components, namely *carbon* (C), *hydrogen* (H), and *oxygen* (O), all in weight percentages.
- *Rate constant* (k): According to the Arrhenius law, the rate constant is significantly dependent on the gasification temperature; its magnitude increases with rise in the gasification temperature (see the experimental dataset in Appendix (6.A.1) on page 201).

The next section presents experimental details of the COG conducted in a fluidized bed pilot plant. This is followed by a brief description of the methodologies adopted for the development of the CI-based models. In the “Results and Discussion” section, the prediction performance of the developed CI-based models, predicting the four COG performance variables are compared numerically and graphically. Further, the best COG performance parameter predicting models were identified among the GPSR, MLPNN and SVR-based models. Finally, the “Conclusion” section summarizes the important finding of the study.

6.1. EXPERIMENTAL

6.1.1. Equipments and Materials

The pilot plant fluidized bed gasifier installed at the *Central Institute of Mining and Fuel Research CSIR-CIMFR*, Dhanbad, was utilized for conducting the co-gasification experiments. The schematic of the plant is shown in Figure (6.2) and it consists of the following sub-systems: (1) hopper feeding system for feeding the solid fuel, (2) mixer and super-heater for preparing desired blends of air-steam gasifying agent for feeding to the gasifier, (3) fluidized bed gasifier reactor, (4) cyclone separator for separating the solid particles generated during the gasification, and (5) quencher and scrubber for cooling and cleaning the product gas.

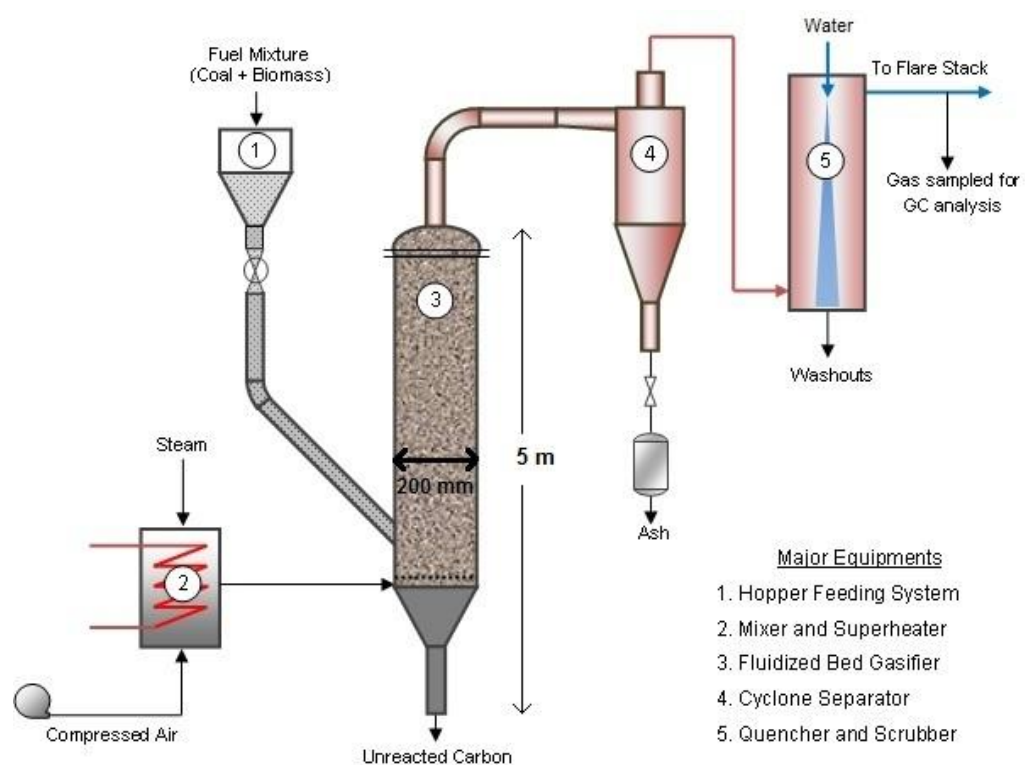


Figure 6.2: Schematic of the FBCOG pilot plant setup.

High ash coals were obtained from the Talcher coalfield located at Angul, Odisha, India. Three types of biomasses, namely, *rice husk*, *sawdust*, and *press-mud*, were obtained from the local sources. Prior to its use, the high ash coal, sawdust and press-mud were crushed and sieved to get a particle size of ~1 mm,

while the rice husk was obtained in the particle size of ~3mm. Steam and filtered air were used as obtained from an electric boiler and a compressor, respectively.

6.1.2. Experimental Procedure

The startup procedure of the FBCOG process involves feeding the solid fuel material (particles of a uniform size of 2 mm) into the gasifier and built-up of the bed in a fluidized form. The desired fluidization of the bed was attained by maintaining the up-flow of air at an adjusted rate. Next, the gasifier temperature is increased by an external electric heating system facility to conduct the co-gasification reaction.

To the FBCOG process, the feedstock containing the sieved binary blends of high ash coal and biomass (sawdust, press-mud or rice husk) were fed in varying proportions (10, 20 and 40 weight percent biomass except for rice husk). The same proportions for coal-biomass blends involving rice husk were 10, 20 and 35 percent biomass (by weight). The bed was charged with this feedstock and fluidization was attained according to the standard gasifier operating procedure mentioned above. The gasifier temperature was raised to $\approx 500^{\circ}\text{C}$, after which the secondary gasifying agents, i.e. superheated steam ($200\text{-}250^{\circ}\text{C}$) and the preheated fluidization air ($200\text{-}250^{\circ}\text{C}$) were fed. The ash generated by the partial combustion was discharged from the bottom and cooled to 40°C prior to discharging it in an ash bin. Once the desired gasifier temperature was reached, the blended feedstock was continuously fed into the gasifier. Depending on the maintained feed and operating conditions, the co-gasification process attained a steady-state in time ranging between 15 to 30 minutes. The fuel gases containing entrained particulates, as generated in the co-gasification were cleaned stage-wise by passing through (i) cyclone separator, and (ii) quenching and scrubbing column, as shown in the schematic in Figure (6.2). The composition of the resulting pure gas was analyzed by collecting the gas samples in glass pipettes and performing off-line gas chromatography (GC) of the samples. The final effluent clean gas was flared from the flare stack. The co-gasification rate constants and the corresponding activation energies were estimated using the laboratory scale

thermo-gravimetric analyzer (TGA). The procedure of the said determination can be found in Cakal et al. (2007).

Based on the stated procedure, 56 experimental runs were conducted in the FBCOG pilot plant by (i) systematically varying the process operating parameters, namely, fuel feed rate and process temperature, (ii) using three different biomasses namely, rice husk, press-mud and sawdust in the blend, and (iii) varying the blending ratios of the coal/biomass in the feedstock.

The main operating parameters of the FBCOG along with their ranges are listed in Table (6.2) and the complete experimental dataset of the 56 runs is given in Appendix (6.A.1) on page 201 of this chapter. Table (6.3) indicates the proximate and ultimate analyses of the solid fuels used in the experiment.

Table 6.2: Operating parameters of the FBCOG process

Operating Parameters	Variation Range
Fuel Feed rate	5.7-9.6 kg/h
Bottom Ash Discharge rate	2.8-4.6 kg/h
Temperature of the gasifier	850-1014°C
Air flow rate	200±20 LPM
Steam flow rate	2.5±0.2 kg/h
Bed Height	10 cm

Table 6.3: Analysis of the solid fuel feed samples

Sample	Proximate Analysis				Ultimate Analysis					HHV
	F_C	V_M	A_S	M	C	H	O	N	S	
Coal	27.8	26.7	40.8	4.7	39.43	3.54	10.25	0.78	0.5	3700
Rice husk	11.8	63.0	15	10.2	37.54	4.65	34.80	0.29	0.02	3480
Press-mud	9.5	61.7	18.5	10.3	33.73	3.92	41.49	2.36	0.01	4051
Sawdust	12.8	72.7	3.6	10.9	41.90	5.24	38.21	0.08	0.07	4000

F_C : Fixed carbon, V_M : Volatile matter, A_S : Ash, M : Moisture, C : Carbon, H : Hydrogen, O : Oxygen, N : Nitrogen, S : Sulphur all are in weight percentages, HHV: Higher heating value (Kcal/Kg)

6.2. RESULTS AND DISCUSSION

6.2.1. Effect of Biomass Types on the FBCOG Performance

Figure (6.3) shows the effect of variation in the coal-biomass ratio on the four FBCOG performance variables. Figure (6.3(a)), shows that for all three types of blends the *TGY* increases linearly as coal:biomass ratio increases although the slope of increase differs with the biomass used. Specifically, *TGY* shows a higher rise as the weight percentage of rice husk increases in the blend compared to other biomass blends. Demirbas (2004) has reported that a higher quantity of biomass in the coal-biomass blends enhances the reactivity of the co-gasification reaction leading to higher gas production. The stated higher reactivity may be attributed to the higher amount of hydrocarbons and lower ash content in biomasses as compared to the coal.

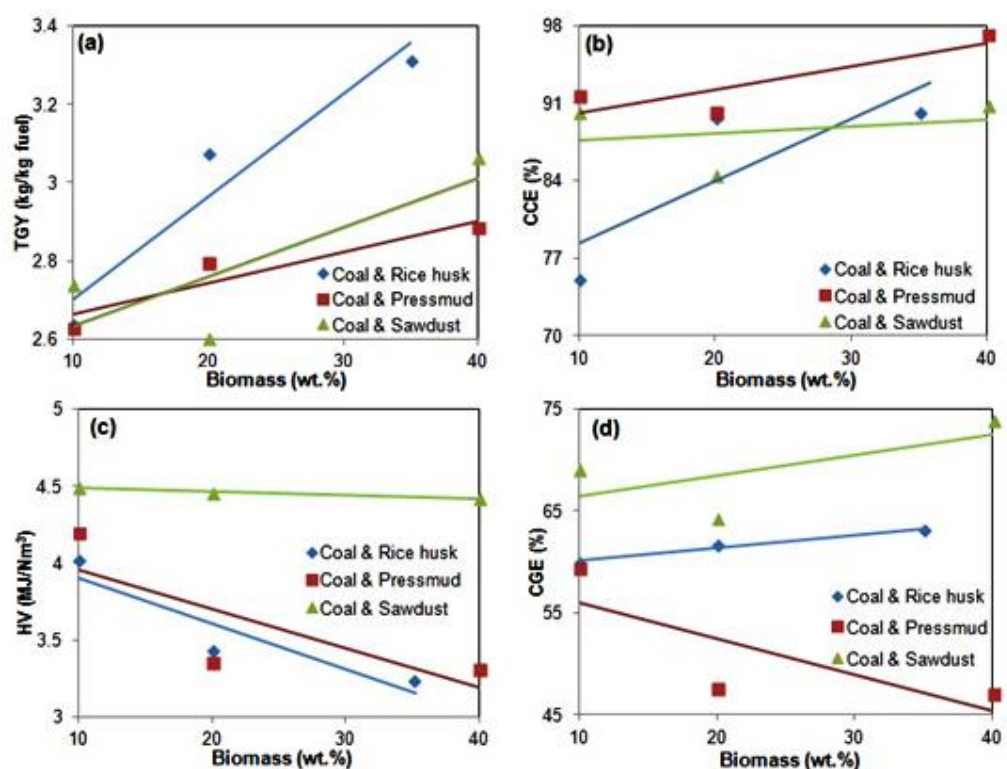


Figure 6.3: Effect of biomass type and its composition in coal-biomass feed blends on the co-gasification process, (a) *TGY* versus biomass percentages, (b) *CCE* versus biomass percentages, (c) *HV* versus biomass percentages, and (d) *CGE* versus biomass percentages.

There are also synergistic effects occurring between the biomass and coal owing to which the transfer of hydrogen radicals from biomass to coal causes more decomposition of coal leading to increase in the amount of gases (Zhang et al., 2007). Similar trends are witnessed in the behavior of carbon conversion efficiency (*CCE*) (see Figure 6.3(b)), due to the changes in the coal-biomass blending ratio wherein increasing biomass content in the blend causes an increase in *CCE*, albeit with varying slopes.

The behavior of the heating value (*HV*) of the effluent product gas towards changing coal-biomass ratio is different (see Figure 6.3(c)) than that displayed by *TGY* and *CCE*. Here, it is observed that *HV* decreases gradually with increasing biomass content of the blended fuel. This can be attributed to the reduction in generation of CO, H₂, and CH₄ in the product gas (Alzate et al., 2009).

A very different effect is seen in the behavior of *CGE* (cold gas efficiency) when coal:biomass ratio is changed (see Figure 6.3 (d)). Specifically, for coal-rice husk and coal-saw dust blends, *CGE* rises continuously as biomass content increases. However, *CGE* declines with increasing press-mud content in the blended feed. The increase in *CGE* witnessed for the coal-rice husk/sawdust blends is likely due to the increase in the *CCE* and *TGY* (see Eqn. 6.5) with increasing biomass content. The decrease in *CGE*, in the case of increasing press-mud content in the blend, may be due to higher oxygen content and lower carbon and hydrogen contents (see Table (6.3)) in the coal–press-mud blends. Also, the feed-rates of air and steam per kg of fuel were maintained high for conducting proper fluidization inside the gasifier for the higher press-mud contents. This reason may also be attributed to the generation of oxygen based non-fuel compounds in the effluent gas leading to a decline in *CGE*.

6.2.2. CI-Based FBCOG Process Modelling

(A) Principal Component Analysis

Principal component analysis (PCA) is a technique that performs dimensionality reduction of a multivariate linearly inter-correlated dataset (see Section (1.7) for details). This reduced dimensionality of the input space lowers

the computational load during GPSR, MLPNN and SVR-based model development. The FBCOG process contains several (here thirteen) feed and operating parameters that strongly affect the key performance variables of the process. In order to capture the variability in the process data set in a few linearly un-correlated variables and thereby reduce the thirteen-dimensional input space of the FBCOG process, PCA was performed using *RapidMiner* software package (Mierswa et al., 2006). For the PCA process, the FBCOG experimental dataset was normalized. The statistical normalization was done on the input and output variables using “z-score” technique given as,

$$\hat{x}_{i,j} = \frac{x_{i,j} - \bar{x}_i}{\sigma_i} \quad (6.5)$$

where, $\hat{x}_{i,j}$ denotes the normalized value of i^{th} input/output variable ($i = 1, 2, \dots, 13$) and j^{th} data-point ($j = 1, 2, \dots, 56$), with $x_{i,j}$ being the corresponding un-normalized element of a data-point. The variables \bar{x}_i and σ_i denote the mean and standard deviation of the i^{th} input/output variable. These values for the FBCOG process variables are listed in Table (6.4).

Table 6.4: Mean and standard deviation of thirteen inputs and four outputs

Variables		Mean (\bar{x})	Standard Deviation (σ)
Independent Variables	T	933.696	37.304
	f_c	5.938	1.346
	f_b	1.66	0.712
	F_C	24.128	1.808
	V_M	35.479	4.024
	A_S	34.398	3.057
	M	5.994	0.593
	R_{AF}	2.475	0.33
	R_{SF}	0.335	0.045
	C	39.073	0.853
	H	3.787	0.151
	O	16.436	3.082
	k	0.0072	0.0035
Dependent Variables	TGY	2.908	0.281
	CCE	88.065	6.287
	HV	3.658	0.7
	CGE	58.839	12.901

The first four principal components (PCs) yielded by PCA captured a large amount of variance of about 99.3% in the normalized FBCOG process data to be used as inputs during modeling. Thus, only first four PCs—instead of the original thirteen—were considered for the development of the CI-based models, thereby reducing the dimensionality of the models' input space from thirteen to four. The four principal components (P_1 , P_2 , P_3 and P_4) are related to the original thirteen FBCOG process/feed variables according to the following expressions,

$$P_1 = 0.133\hat{T} + 0.334\hat{f}_c - 0.32\hat{f}_b + 0.333\hat{F}_C - 0.355\hat{V}_M + 0.34\hat{A}_S - 0.356\hat{M} - 0.22\hat{R}_{AF} - 0.22\hat{R}_{SF} + 0.077\hat{C} - 0.251\hat{H} - 0.335\hat{O} + 0.108\hat{k} \quad (6.6)$$

$$P_2 = 0.058\hat{T} + 0.107\hat{f}_c + 0.289\hat{f}_b - 0.254\hat{F}_C + 0.022\hat{V}_M + 0.1\hat{A}_S + 0.109\hat{M} - 0.357\hat{R}_{AF} - 0.357\hat{R}_{SF} - 0.566\hat{C} - 0.407\hat{H} + 0.246\hat{O} + 0.113\hat{k} \quad (6.7)$$

$$P_3 = 0.573\hat{T} - 0.199\hat{f}_c - 0.054\hat{f}_b - 0.053\hat{F}_C - 0.05\hat{V}_M + 0.098\hat{A}_S - 0.006\hat{M} + 0.325\hat{R}_{AF} + 0.324\hat{R}_{SF} - 0.184\hat{C} - 0.109\hat{H} - 0.008\hat{O} + 0.599\hat{k} \quad (6.8)$$

$$P_4 = 0.399\hat{T} + 0.189\hat{f}_c + 0.138\hat{f}_b + 0.036\hat{F}_C + 0.189\hat{V}_M - 0.289\hat{A}_S + 0.098\hat{M} - 0.358\hat{R}_{AF} - 0.357\hat{R}_{SF} + 0.403\hat{C} + 0.347\hat{H} + 0.065\hat{O} + 0.336\hat{k} \quad (6.9)$$

The resulting PCA reduced dataset was split randomly in the training (75%) and test (25%) sets and was used for development of the CI-based models possessing good prediction accuracies and generalization capabilities as stated further.

(B) Development of GPSR-based FBCOG Process Models

The GPSR-based models, predicting the four FBCOG process performance variables (TGY , CCE , HV and CGE) were developed using *Eureqa Formulize* (Schmidt and Lipson, 2009) software package. The features and facilities offered by the *Eureqa Formulize* package can be found in Section (1.2.2). The prediction accuracies and generalization capabilities of the developed CI-based models were evaluated in terms of the *coefficient of correlation (CC)* and *root mean squared error (RMSE)*, between the experimental and the corresponding model-predicted values of each performance variable. Several GPSR runs were conducted by varying the GPSR algorithmic and procedural

parameters and the best models were chosen, based on the overall highest and lowest magnitudes of CC and $RMSE$ respectively for both training and test sets. The four GPSR-based models thus obtained are given below and the corresponding magnitudes of the training and test set CC s and $RMSE$ s are listed in Table (6.5).

(a) GPSR-Model 1 for TGY

$$TGY = 0.379P_3 + \frac{(-0.011P_4)}{(-0.078 - 0.048P_1P_4 - 0.405P_2P_3)} - 0.209P_1 - 0.316P_2 - 0.323P_4 \quad (6.10)$$

(b) GPSR-Model 2 for CCE

$$CCE = 0.45P_2 + 0.073P_1 + \left(\frac{0.45P_2 + 0.335P_1P_4 + 0.073P_1P_3P_2^2 - 0.335 - P_2P_4 - 0.795P_3}{P_1 - 2.57} \right) - 0.073P_1 \quad (6.11)$$

(c) GPSR-Model 3 for HV

$$HV = 0.883P_4 + 0.1404P_1 + \left(\frac{0.155}{P_1} \right) + \left(\frac{0.037}{0.193P_1 + P_2P_4} \right) - 0.083P_3 \quad (6.12)$$

(d) GPSR-Model 4 for CGE

$$CGE = 0.576P_4 + 1.582 \exp \left(\frac{0.476 - 0.129P_2 - 0.129P_2P_3}{P_1 + 0.204P_3 + 0.129P_2 - 0.736P_4 - 0.129P_2P_3} \right) - 1.67 - 0.261P_2 \quad (6.13)$$

From CC and $RMSE$ magnitudes listed in the table, it is clear the GPSR-based TGY (CC s > 0.97 , $RMSE$ s < 0.06) and CCE (CC s > 0.95 , $RMSE$ s < 1.7) prediction models possess an excellent prediction accuracy and generalization capability. Similarly, the GPSR-based models predicting the HV and CGE performance variables also possess good prediction accuracies and generalization capabilities with high and comparable magnitudes of CC s and low $RMSE$ s, although on a marginally lower scale compared to TGY and CCE predicting models. Figure (6.4) graphically depicts the prediction and generalization performances of the four GPSR-based models, wherein experimental versus the model-predicted values are plotted as parity plots in four panels ($a-d$). As can be

noticed from the panels that all the GPSR-based models exhibit low scatter; among the four parity plots those for *TGY* and *CCE* exhibit much lower scatter (than in the plots of *HV* and *CGE*) indicating better performance by the respective models. These plots support the *CC* and *RMSE* based observation that *TGY* and *CCE* predicting GPSR models have outperformed the remaining two models.

Table 6.5: Comparison of CI-based models for the prediction of four FBCOG performance variables

Performance Variables	Model	Training set		Test set	
		<i>CC</i>	<i>RMSE</i>	<i>CC</i>	<i>RMSE</i>
<i>TGY</i> (kg/kg fuel)	GPSR-Model 1	0.98	0.05	0.99	0.02
	MLPNN-Model 1	0.99	0.03	0.99	0.02
	SVR-Model 1	0.99	0.03	0.99	0.02
<i>CCE</i> (%)	GPSR-Model 2	0.97	1.61	0.96	1.02
	MLPNN-Model 2	0.99	0.73	0.92	1.41
	SVR-Model 2	0.91	3.03	0.91	2.42
<i>HV</i> (MJ/Nm ³)	GPSR-Model 3	0.95	0.20	0.95	0.25
	MLPNN-Model 3	0.94	0.22	0.93	0.28
	SVR-Model 3	0.97	0.15	0.95	0.25
<i>CGE</i> (%)	GPSR-Model 4	0.91	5.39	0.98	2.51
	MLPNN-Model 4	0.94	5.54	0.94	3.97
	SVR-Model 4	0.96	3.52	0.95	3.99

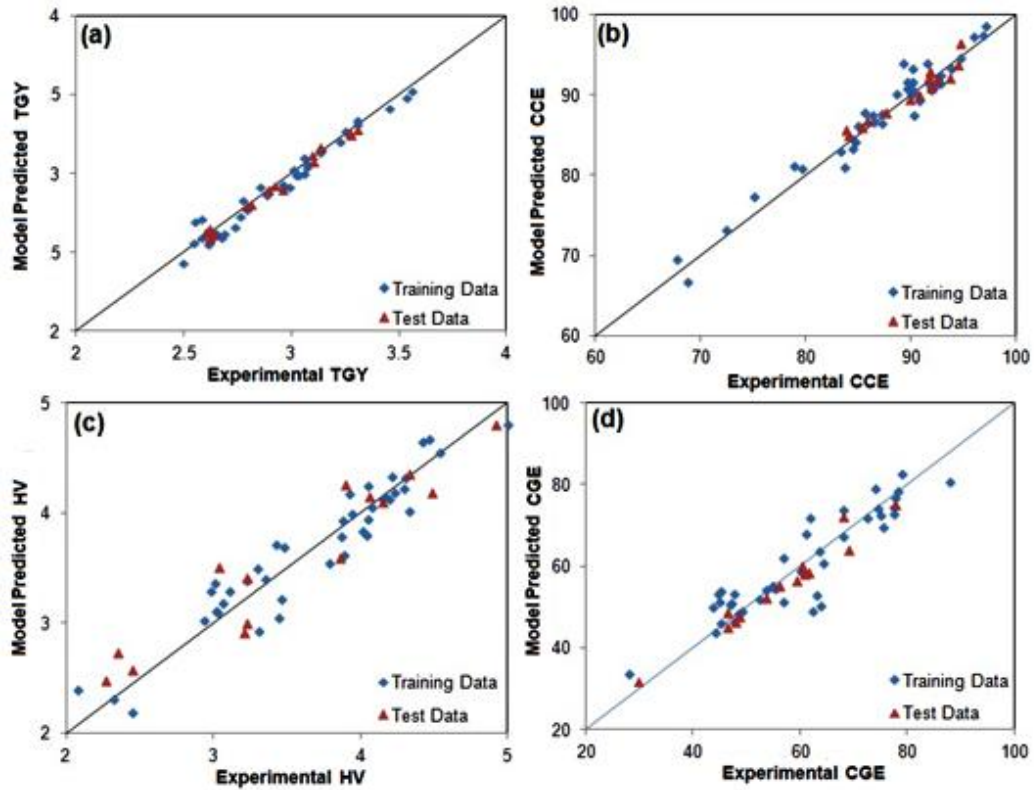


Figure 6.4: Parity plots of experimental versus GPSR-based model-predicted values of the FBCOG performance variables, namely *TGY* (kg/kg fuel) (panel a), *CCE* (%) (panel b), *HV* (MJ/Nm³) (panel c) and *CGE* (%) (panel d).

(C) Development of MLPNN-based FBCOG Process Models

Similar to the GPSR-based models, four MLPNN based models were developed using *RapidMiner* software (Mierswa et al., 2006). Each of the developed MLPNN-based models has a single hidden layer in its architecture, which was trained by the EBP algorithm (see Section 1.2.1(A)). To obtain the optimal MLPNN-based models, the influence of the network's structural parameters (i.e., the number of hidden layers and number of nodes in each hidden layer) and two error-back-propagation (EBP) algorithm specific parameters (*learning rate*, η and *momentum coefficient*, μ) on the model's prediction and generalization performance were systematically examined. The corresponding details of the optimal MLPNN models are given in Table (6.6).

Table 6.6: Architectural details of the MLPNN-based FBCOG models

Model	Output	Input Nodes	Hidden Layer Nodes	η	μ
MLPNN-Model 1	<i>TGY</i>	4	6	0.16	0.02
MLPNN-Model 2	<i>CCE</i>	4	5	0.19	0.09
MLPNN-Model 3	<i>HV</i>	4	5	0.26	0.097
MLPNN-Model 4	<i>CGE</i>	4	5	0.275	0.14

Hidden layer transfer function: Logistic Sigmoid, Output layer transfer function: Linear

The magnitudes of the two statistical measures (*CC* and *RMSE*) in respect of the predictions of the four MLPNN models are listed in Table (6.5). From these, it is seen that in general the prediction and generalization performances of all the four models are good. Similar trends are observed in the four parity plots (Figure (6.5)) of the experimental values of the performance variables versus those predicted by the corresponding MPLNN models. Among the four MLPNN-based models, the model predicting *TGY* (*CC* = 0.99 for both the training and test sets) possesses an excellent prediction accuracy and generalization ability. This can also be witnessed from the close match between experimental and model predicted values of the *TGY* in Figure (6.5) (panel (a)).

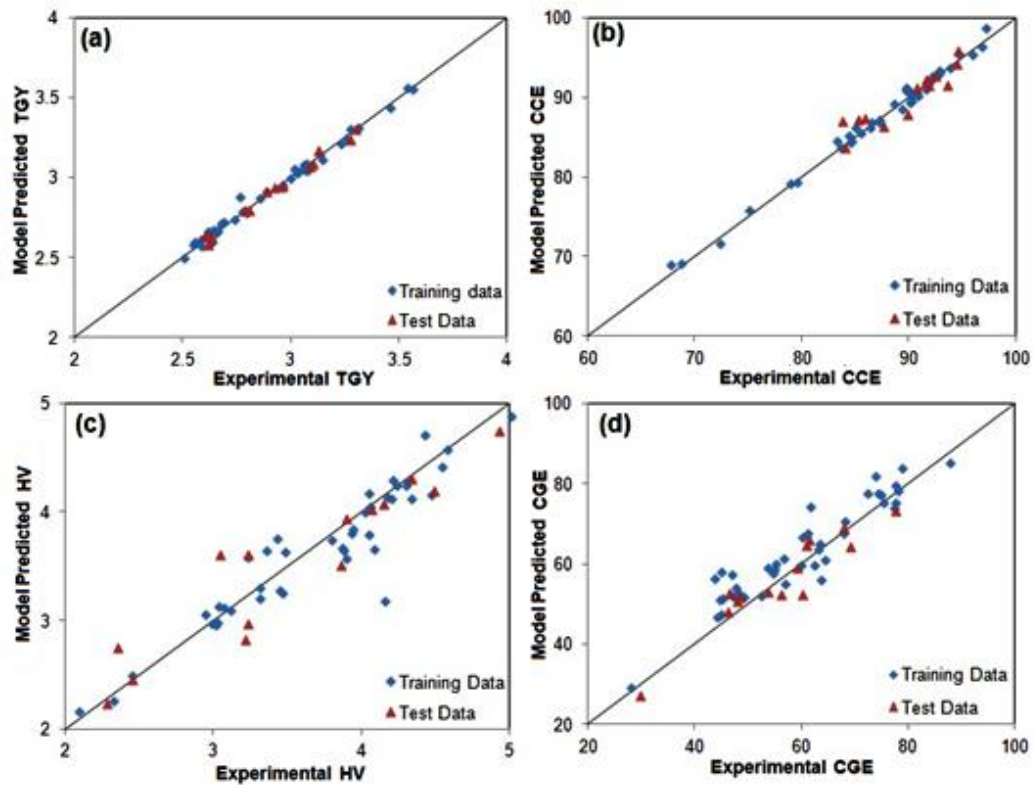


Figure 6.5: Parity plots of experimental versus MLPNN-based model-predicted values of the FBCOG performance variables, namely *TGY* (kg/kg fuel) (panel a), *CCE* (%) (panel b), *HV* (MJ/Nm³) (panel c) and *CGE* (%) (panel d).

(D) Development of SVR-based FBCOG Process Models

Utilizing the same datasets used in the development of GPSR and MLPNN-based modeling, four SVR-based optimal models were constructed for the prediction of four performance variables, namely, *TGY*, *CCE*, *HV* and *CGE*, using RapidMiner software (Mierswa et al., 2006). The *epsilon* (ϵ)-insensitive loss function was chosen for developing the best-performing SVR-based models; the SVR algorithm parameters leading to four optimal models are provided in Table (6.7). From the *CC* (high) and *RMSE* (low) values listed in Table (6.5), it can be observed that all four SVR-based models exhibit excellent prediction and generalization performance, although the SVR-Model 2 predicting the *CCE* is found to perform marginally inferior. Similar observations can be drawn by the lower scatter visible in the parity plots pertaining to the predictions of SVR-models 1, 3 and 4 and a marginally higher scatter (Figure 6.6 (b)) in the parity plot in respect of the SVR-Model 2 predictions.

Table 6.7: Details of SVR-based FBCOG models

Model	Output	Kernel function	Kernel gamma (γ)	Epsilon (ϵ)	Number of Support Vectors (SVs)
SVR-Model 1	<i>TGY</i>	ANOVA	0.159	0.1	27
SVR-Model 2	<i>CCE</i>	ANOVA	0.196	0.245	25
SVR-Model 3	<i>HV</i>	ANOVA	0.707	0.01	38
SVR-Model 4	<i>CGE</i>	ANOVA	0.88	0.132	36

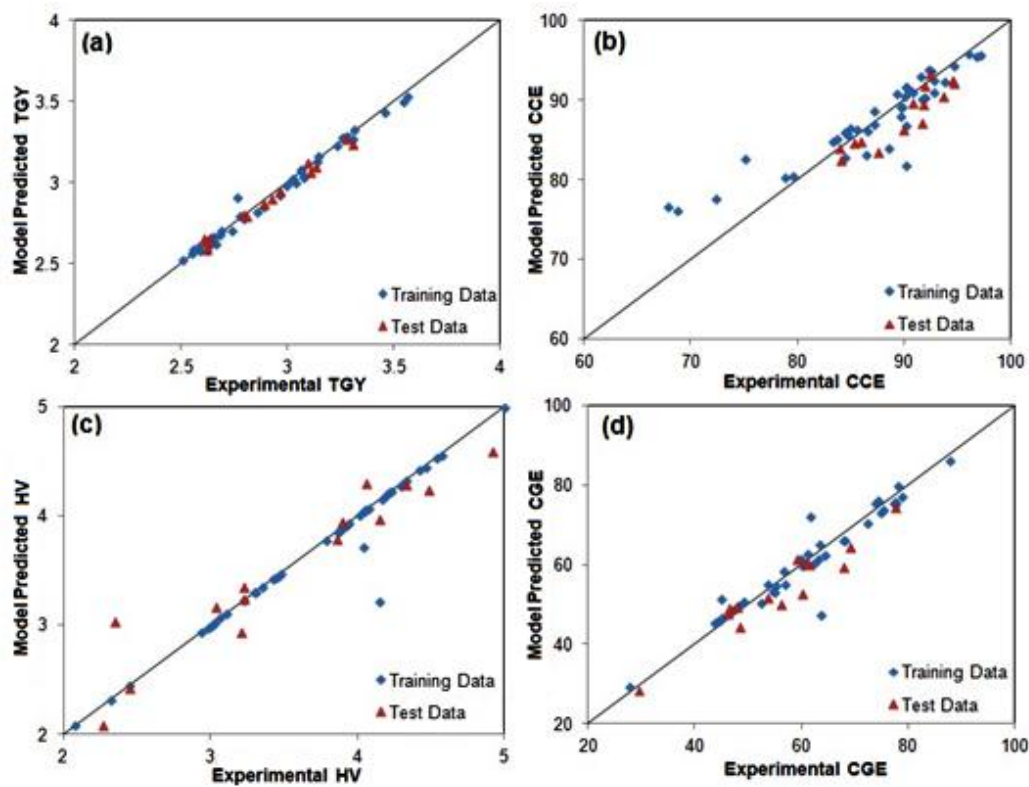


Figure 6.6: Parity plots of experimental versus SVR-based model-predicted values of the FBCOG performance variables, namely *TGY* (kg/kg fuel) (panel a), *CCE* (%) (panel b), *HV* (MJ/Nm³) (panel c) and *CGE* (%) (panel d).

6.2.3. Statistical Comparison of the CI-Based Models

The statistical test known as Steiger's z-test (Steiger, 1980) (see Section (1.8) for details) was conducted for identifying the best performing model from among the three CI-based models, constructed for predicting each of the four COG performance variables. Specifically, this test compares the z-scores

pertaining to the *CC* magnitudes resulting from the predictions of two competing models.

The Steiger's z-test was performed using *CC* values of different pairs of the CI-based models (GPSR, MLPNN, and SVR). Table (6.8) shows the results of this test, from which following observations can be made about the best performing model for predicting the four FBCOG performance variables,

- *TGY*: From the *p*-value (0.155) for the MLPNN-SVR pair of models (0.155) it can be inferred that both MLPNN and SVR models predicting *TGY* possess closely matching prediction and generalization performance which is better than that of the GPSR-Model 1.
- *CCE*: The Steiger's z-test carried out for all three pairs of CI models reject the null hypothesis. This indicates that there are significant differences in the prediction performances of all the three CI-based models. From the magnitudes of the *CC*s, the MLPNN-Model 2 (*CC* = 0.99) is found to yield better prediction performances than the GPSR and SVR-based models.
- *HV*: From the *p*-values, it is evident that the SVR and GPSR-based models have outperformed the MLPNN-Model 3.
- *CGE*: SVR-Model 4 possesses better prediction accuracy than the MLPNN and GPSR-based models predicting the *CGE*.

From this statistical comparison, it is clear that no single CI-based exclusively data-driven method from among GPSR, MLPNN, and SVR has consistently yielded best performance in predicting all the four FBCOG performance variables. Notwithstanding, this fact it is evident that the model determined to be best (among the competing three) statistically for the prediction of a performance variable possesses excellent prediction accuracy and generalization capability. The GPSR-based models though provide marginally inferior results in terms of prediction and generalization performance than the corresponding MLPNN and SVR models, their much reduced complexity and ease of deployment make them ideal candidates for use in practical settings. However, when accuracy of predictions is the sole criteria of selection, the best models determined by the statistical test should be given priority.

Table 6.8: Results of Steiger's z-test comparing correlation coefficients pertaining to CI-based models for co-gasification performance variables

Performance Variable	Model pair (B-C)	CC_{AB}	CC_{AC}	CC_{BC}	df	z	p -value	H_0
<i>TGY</i>	GPSR–MLPNN	0.987	0.995	0.982	53	-3.4	5.4×10^{-4}	Reject
	MLPNN–SVR	0.995	0.993	0.994	53	1.4	0.155	Accept
	SVR–GPSR	0.993	0.987	0.983	53	2.2	0.023	Reject
<i>CCE</i>	GPSR–MLPNN	0.974	0.99	0.97	53	-3.5	0.0004	Reject
	MLPNN–SVR	0.99	0.894	0.894	53	8.4	0	Reject
	SVR–GPSR	0.894	0.974	0.867	53	-4.9	7.7×10^{-7}	Reject
<i>HV</i>	GPSR–MLPNN	0.949	0.94	0.945	53	0.6	0.493	Accept
	MLPNN–SVR	0.94	0.966	0.965	53	-2.7	0.006	Reject
	SVR–GPSR	0.966	0.949	0.938	53	1.5	0.127	Accept
<i>CGE</i>	GPSR–MLPNN	0.926	0.934	0.953	53	-0.5	0.574	Accept
	MLPNN–SVR	0.934	0.959	0.964	53	-2.3	0.017	Reject
	SVR–GPSR	0.959	0.926	0.942	53	2.5	0.012	Reject

H_0 : $CC_{AB}=CC_{AC}$, where A denotes experimental values, B and C denote model-predicted values, df refers to the degrees of freedom; reject H_0 if p -value < 0.05.

6.3. CONCLUSION

Co-gasification (COG) is a newer fuel gas producing technology that to a significant extent mitigates the problems of air pollution and water-body contamination (caused by percolation/release of ash) due to coal combustion. It also reduces carbon foot-print since it uses a blend of coal and biomass, which is a renewable fuel. Moreover, use of biomass along with coal adds flexibility to the gasification process. As a result, co-gasification has emerged as an attractive technology for co-utilizing the available coal resources and cheaply available biomass resources. In this study, a difficult-to-model (using first principles) pilot

plant scale fluidized bed co-gasification (FBCOG) process exhibiting nonlinear behavior has been modeled using three computational intelligence (CI) based exclusively data-driven modeling approaches, namely, genetic programming based symbolic regression (GPSR), multi-layer perceptron neural network (MLPNN) and support vector regression (SVR). The extensive FBCOG experiments that were conducted, utilized blends of high coals (currently mined extensively in India) and three types of biomasses (available widely in India), namely sawdust, press-mud and rice husk. The resulting experimental dataset consisted of thirteen inputs (important FBCOG feed and operating parameters) and the corresponding four key process performance variables, namely, *total gas yield*, *carbon conversion efficiency*, *heating value of product gas* and *cold gas efficiency*. The thirteen-dimensional input space of the CI-based models was reduced to four dimensions by performing principal component analysis (PCA). This reduced dataset was then used to develop the CI-based models predicting the four performance variables. All the CI-based models developed possess good prediction and generalization performance as indicated by high training and test set correlation coefficient (0.908–0.996) and low root mean squared error (0.02–5.54) magnitudes. The CI-based models developed in this study can be used to design and operate co-gasifiers utilizing blends of high ash coals and three specific types of biomasses as co-feeds. Also, the modeling approach proposed in this study is less tedious and time-consuming compared to the first principles modeling and therefore can be easily extended to model COG processes operating with different configurations and using a wide variety of fuel blends not considered in this study.

APPENDIX 6.A: Experimental dataset of the pilot plant fluidized bed co-gasification process

Data	T	f_c	f_b	F_C	V_M	A_S	M	R_{AF}	R_{SF}	C	H	O	k	TGY	CCE	HV	CGE
C90 RH10	910	7.2	0.8	26.2	30.33	38.2	5.3	2.311	0.313	39.24	3.65	12.71	0.0047	2.554	67.75	3.227	44.59
	918	7.11	0.79					2.34	0.3164				0.0053	2.586	68.729	3.227 ^{&}	45.172
	925	7.2	0.8					2.311	0.313				0.006	2.618 [#]	72.318	3.789	56.155 [*]
	950	7.2	0.8					2.311	0.313				0.0085	2.634 [#]	74.987	4.012	59.964
C80 RH20-1	875	5.84	1.46	24.6	33.96	35.64	5.8	2.533	0.342	39.052	3.762	15.16	0.003	2.961	74.987	3.46	60.248 [*]
	928	5.68	1.42					2.604	0.352				0.007	2.762	88.574	3.038 ^{&}	61.003
	955	5.6	1.4					2.641	0.357				0.0094	3.074	89.648	3.425	61.699
	960	5.6	1.4					2.641	0.357				0.01	3.103 [#]	93.631 ^{\$}	4.043	77.464
	968	5.28	1.32					2.801	0.379				0.0112	3.23	93.777	3.886	77.348
	970	5.2	1.3					2.845	0.384				0.0114	3.273 [#]	94.5 ^{\$}	3.861 ^{&}	77.75 [*]
	990	5.2	1.3					2.845	0.384				0.014	3.27	94.623	3.876	78.165
	1014	5.12	1.28					2.889	0.391				0.0172	3.305	94.663 ^{\$}	3.867	78.952
C80 RH20-2	927	6.4	1.6	24.6	33.96	35.64	5.8	2.311	0.3125	39.052	3.762	15.16	0.0066	2.793	87.158	4.332	75.404
	942	7.68	1.92					1.926	0.26				0.008	2.501	89.245	4.923 ^{&}	77.616

Data	T	f_c	f_b	F_C	V_M	A_S	M	R_{AF}	R_{SF}	C	H	O	k	TGY	CCE	HV	CGE
	950	6.4	1.6					2.311	0.313				0.0089	2.809 [#]	89.959	4.227	72.311
	960	6	1.5					2.465	0.333				0.01	2.961	92.741	4.148 ^{&}	74.786
	970	5.84	1.46					2.533	0.342				0.0112	3.014	92.781	4.062 ^{&}	74.43
C65 RH35-1	856	4.03	2.17	22.2	39.405	31.76	6.63	2.982	0.403	38.77	3.93	18.843	0.0019	3.255	85.489	2.447	44.067
	920	4.03	2.17					2.982	0.403				0.0045	3.303 [#]	89.862 ^s	3.207 ^{&}	62.283
	950	4.03	2.17					2.982	0.403				0.0082	3.309	90.086	3.231 ^{&}	63.088
C65 RH35-2	924	3.705	1.995	22.2	39.405	31.76	6.63	3.244	0.439	38.77	3.93	18.843	0.0054	3.561	89.773	2.322	46.345 [*]
	936	3.77	2.03					3.188	0.431				0.0068	3.539	91.84	2.448 ^{&}	48.602 [*]
	950	3.9	2.1					3.0816	0.417				0.0082	3.455	91.914 ^s	2.35 ^{&}	44.922
C90 PM10	915	8.01	0.89	25.97	30.2	38.57	5.26	2.077	0.281	38.86	3.58	13.374	0.0044	2.617 [#]	90.113	4.075	56.883
	955	8.01	0.89					2.077	0.281				0.009	2.629	91.637	4.193	59.31 [*]
	966	8.1	0.9					2.054	0.278				0.0108	2.617	92.439 ^s	4.302	60.855 [*]
	970	8.1	0.9					2.054	0.278				0.011	2.622	92.539	4.329 ^{&}	61.559 [*]
	972	8.01	0.89					2.077	0.281				0.0114	2.641	92.209	4.207	60.166
C80 PM20	942	6.96	1.74	24.14	33.7	36.34	5.82	2.125	0.287	38.29	3.62	16.498	0.0066	2.655	90.771	3.925	54.663

Data	T	f_c	f_b	F_C	V_M	A_S	M	R_{AF}	R_{SF}	C	H	O	k	TGY	CCE	HV	CGE
	946	6.96	1.74					2.125	0.287				0.007	2.648	90.293	3.938	54.765
	958	6.4	1.6					2.311	0.313				0.0089	2.792 [#]	90.09	3.353	47.534
C60 PM40	868	4.62	3.08	20.48	40.7	31.88	6.94	2.401	0.325	37.15	3.69	22.746	0.0021	2.775	83.589	2.273 ^{&}	29.666 [*]
	878	4.44	2.96					2.499	0.338				0.0029	2.854	83.265	2.082	27.823
	927	5.28	3.52					2.101	0.284				0.007	2.617	91.498	3.475	44.973
	945	4.62	3.08					2.401	0.325				0.01	2.889 [#]	97.127	3.302	46.994
	946	4.56	3.04					2.433	0.329				0.0101	2.923 [#]	95.941	3.0676	43.70
	948	4.68	3.12					2.37	0.321				0.0102	2.885	97.127	3.302	46.994
C90 SD10	935	7.74	0.86	26.3	31.3	37.08	5.32	2.15	0.291	39.68	3.71	13.046	0.0056	2.687	89.656	4.293	63.452
	940	7.56	0.84					2.201	0.298				0.006	2.738	90.186	4.489 ^{&}	69.155 [*]
	955	7.92	0.88					2.101	0.284				0.0076	2.679	91.755 ^s	4.533	67.857
	983	8.01	0.89					2.077	0.281				0.0124	2.659	91.62 ^s	4.57	67.969 [*]
C80 SD20-1	850	5.76	1.44	24.8	35.9	33.36	5.94	2.568	0.347	39.9	3.88	15.84	0.0011	2.994	84.092 ^s	3.31	52.299
	862	5.84	1.46					2.533	0.342				0.0016	2.96 [#]	84.389	3.439	53.662 [*]
	890	5.6	1.4					2.641	0.357				0.0026	3.025	83.867 ^s	3.025	46.787

Data	T	f_c	f_b	F_C	V_M	A_S	M	R_{AF}	R_{SF}	C	H	O	k	TGY	CCE	HV	CGE
	920	5.52	1.38					2.68	0.362				0.0044	3.071	85.3 ^{12s}	3.111	49.136
	936	5.36	1.34					2.76	0.373				0.0056	3.13 [#]	85.883 ^s	3.01	47.937 [*]
	953	5.36	1.34					2.76	0.373				0.0076	3.133	86.466	2.98	47.552
	976	5.36	1.34					2.76	0.373				0.0112	3.142	87.162	3.015	48.204
C80 SD20-2	860	6.8	1.7	24.8	35.9	33.36	5.94	2.175	0.294	39.9	3.88	15.84	0.0015	2.582	78.828	3.892 ^{&}	53.637
	865	6.96	1.74					2.125	0.287				0.0017	2.548	79.540	4.052	55.216
	940	6.96	1.74					2.125	0.287				0.006	2.603 [#]	84.437	4.462	64.32
	950	6.96	1.74					2.125	0.287				0.0071	2.615	84.624	4.046	56.654
	970	5.44	1.36					2.719	0.368				0.01	3.094 [#]	84.94	2.938	46.47 [*]
C60 SD40	865	4.2	2.8	21.8	45.1	25.92	7.18	2.641	0.357	40.42	4.22	21.434	0.0019	3.033	87.595 ^s	4.171	68.04
	920	4.2	2.8					2.641	0.357				0.005	3.063	90.744 ^s	4.419	73.909
	933	4.2	2.8					2.641	0.357				0.0064	3.06	91.882	4.999	87.760

where, C90 RH10 represents Coal (90%), Rice husk (10%), C80 RH 20 represents Coal (80%), Rice husk (20%) and C65 RH35 represents Coal (65%) and Rice husk (35%). Similarly, C90 PM 10 represents Coal 90% Pressmud 10%, C90 SD 10 represents Coal 90% and Sawdust 10% #, \$, & and * - represents the test set of TGY , CCE , HV and CGE

NOMENCLATURE

A_S	Ash (wt%)
C	Carbon (wt%)
f_b	Biomass feed rate
f_c	Coal feed rate
F_C	Fixed carbon (wt%)
H	Hydrogen (wt%)
k	Rate constant
M	Moisture (wt%)
O	Oxygen (wt%)
p_i	i^{th} principal component
R_{AF}	Air-to-fuel ratio
R_{SF}	Steam-to-fuel ratio
T	Gasifier temperature
V_M	Volatile matter (wt%)
x	Un-normalized variable
$\hat{x}_{i,j}$	Normalized variable input/output variable
\bar{x}_i	Mean of variable 'x'

Greek letters

μ	Momentum coefficient in the EBP algorithm
ε	Epsilon (tube width); a precision parameter in SVR
η	Learning rate in the EBP algorithm

- γ Kernel gamma of radial basis function in SVR
- σ Standard deviation

REFERENCES

- Aigner, I., Pfeifer, C., Hofbauer, H. (2011). Co-gasification of coal and wood in a dual fluidized bed gasifier. *Fuel* 90(7):2404-2412.
- Alzate, C. A., Chejne, F., Valdés, C. F., Berrio, A., De La Cruz, J., Londoño, C. A. (2009). CO-gasification of pelletized wood residues. *Fuel* 88(3):437-445.
- André, R. N., Pinto, F., Franco, C., Dias, M., Gulyurtlu, I., Matos, M. A. A., Cabrita, I. (2005). Fluidised bed co-gasification of coal and olive oil industry wastes. *Fuel* 84(12):1635-1644.
- Armstrong, L. M., Gu, S., Luo, K. H. (2011). Parametric study of gasification processes in a BFB coal gasifier. *Industrial & Engineering Chemistry Research*, 50(10), 5959-5974.
- Çakal, G. Ö., Yücel, H., Gürüz, A. G. (2007). Physical and chemical properties of selected Turkish lignites and their pyrolysis and gasification rates determined by thermogravimetric analysis. *Journal of analytical and applied pyrolysis* 80(1):262-268.
- Chavan, P. D., Sharma, T., Mall, B. K., Rajurkar, B. D., Tambe, S. S., Sharma, B. K., Kulkarni, B. D. (2012). Development of data-driven models for fluidized-bed coal gasification process. *Fuel* 93:44-51.
- Cormos, C. C. (2013). Assessment of flexible energy vectors poly-generation based on coal and biomass/solid wastes co-gasification with carbon capture. *International journal of hydrogen energy* 38(19):7855-7866.
- de Jong, W., Andries, J., Hein, K. R. (1999). Coal/biomass co-gasification in a pressurised fluidised bed reactor. *Renewable Energy* 16(1-4):1110-1113.

- de Souza-Santos, M. L. (1989). Comprehensive modelling and simulation of fluidized bed boilers and gasifiers. *Fuel* 68(12):1507-1521.
- Demirbas, A. (2004). Combustion characteristics of different biomass fuels. *Progress in energy and combustion science* 30(2):219-230.
- Donne, M. S., Dixon, R., Pike, A. W., Odeku, A. J. L., Ricketts, B. E. (1998). Dynamic modelling of the ABGC prototype integrated plant. Harwell Laboratory, Coal R & D Programme.
- Freeman, J. A., Skapura, D. M. (1991). Algorithms, Applications, and Programming Techniques. Addison-Wesley Publishing Company, USA.
- Gärtner, L. E., Gräbner, M., Meyer, B. (2012). Influence of coal blend component kinetics on entrained flow gasification performance. In: 9th International Conference on CFD in the Minerals and Process Industries (CSIRO), http://www.cfd.com.au/cfd_conf12/PDFs/022GAR.pdf.
- Gururajan, V. S., Agarwal, P. K., Agnew, J. B. (1992). Mathematical modelling of fluidized bed coal gasifiers: Chemical reaction engineering. *Chemical engineering research & design* 70(A3):211-238.
- Koza, J. R. (1992). Genetic programming: on the programming of computers by means of natural selection (Vol. 1). MIT press.
- Lee, S.(2007). Gasification of coal. In: Handbook of Alternative Fuel Technologies; Lee, S., Speight, J. G., Loyalka, S. K., Eds.; CRC Press: Boca Raton, pp. 25–80.
- Li, K., Zhang, R., Bi, J. (2010). Experimental study on syngas production by co-gasification of coal and biomass in a fluidized bed. *International journal of hydrogen energy* 35(7):2722-2726.
- Mastellone, M. L., Zaccariello, L., Arena, U. (2010). Co-gasification of coal, plastic waste and wood in a bubbling fluidized bed reactor. *Fuel* 89(10):2991-3000.

- Mazumder, A. (2010). Development of a simulation model for fluidized bed mild gasifier. Thesis, Paper 101, University of New Orleans, New Orleans; <http://scholarworks.uno.edu/td/101/>. Accessed: 25 April 2015).
- McLendon, T. R., Lui, A. P., Pineault, R. L., Beer, S. K., Richardson, S. W. (2004). High-pressure co-gasification of coal and biomass in a fluidized bed. *Biomass and Bioenergy* 26(4):377-388.
- Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., Euler, T. (2006). YALE: rapid prototyping for complex data mining tasks, In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-06).
- Mjalli, F. S., Al-Mfargi, A. (2008). Artificial neural approach for modeling the heat and mass transfer characteristics in three-phase fluidized beds. *Industrial & Engineering Chemistry Research* 47(13):4542-4552.
- Moorea-Taha, R. (2000). Modeling and Simulation for Coal Gasification; IEA CoalResearch 2000; IEA Clean Coal: London; ISBN 92- 9029-354-3, pp. 1–50.
- Nougués, J. M., Pan, Y. G., Velo, E., Puigjaner, L. (2000). Identification of a pilot scale fluidised-bed coal gasification unit by using neural networks. *Applied thermal engineering* 20(15):1561-1575.
- Pan, Y. G., Velo, E., Roca, X., Manyà, J. J., Puigjaner, L. (2000). Fluidized-bed co-gasification of residual biomass/poor coal blends for fuel gas production. *Fuel* 79(11):1317-1326.
- Patil-Shinde, V., Kulkarni, T., Kulkarni, R., Chavan, P. D., Sharma, T., Sharma, B. K., Kulkarni, B. D. (2014). Artificial intelligence-based modeling of high ash coal gasification in a pilot plant scale fluidized bed gasifier. *Industrial & Engineering Chemistry Research* 53(49):18678-18689.
- Pérez-Fortes, M., Bojarski, A. D., Velo, E., Nougués, J. M., Puigjaner, L.

- (2009). Conceptual model and evaluation of generated power and emissions in an IGCC plant. *Energy*, 34(10):1721-1732.
- Pinto, F., Franco, C., Andre, R. N., Tavares, C., Dias, M., Gulyurtlu, I., Cabrita, I. (2003). Effect of experimental conditions on co-gasification of coal, biomass and plastics wastes with air/steam mixtures in a fluidized bed system. *Fuel* 82(15):1967-1976.
- Puig-Arnavat, M., Hernández, J. A., Bruno, J. C., Coronas, A. (2013). Artificial neural network models for biomass gasification in fluidized bed gasifiers. *biomass and bioenergy* 49:279-289.
- Rezaiyan, J., Cheremisinoff, N. P. (2005). *Gasification technologies: a primer for engineers and scientists*. CRC press, Taylor and Francis Group, New York, USA.
- Rhinehart, R. R., Felder, R. M., Ferrell, J. K. (1987). Dynamic modeling of a pilot-scale fluidized-bed coal gasification reactor. *Industrial & engineering chemistry research* 26(4):738-745.
- Schmidt, M., Lipson, H. (2009). Distilling free-form natural laws from experimental data. *Science* 324:81–85.
- Sett, A., Bhattacharya, S. C. (1988). Mathematical modelling of a fluidised-bed charcoal gasifier. *Applied energy* 30(3):161-186.
- Shen, C. H., Chen, W. H., Hsu, H. W., Sheu, J. Y., Hsieh, T. H. (2012). Co-gasification performance of coal and petroleum coke blends in a pilot-scale pressurized entrained-flow gasifier. *International Journal of Energy Research* 36(4):499-508.
- Singh, N., Raghavan, V., and Sundararajan, T. (2014). Mathematical modeling of gasification of high-ash Indian coals in moving bed gasification system. *International Journal of Energy Research* 38(6):737-754.
- Sjöström, K., Chen, G., Yu, Q., Brage, C., Rosén, C. (1999). Promoted reactivity of char in co-gasification of biomass and coal: synergies in the

- thermochemical process. *Fuel* 78(10): 1189-1194.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin* 87:245–251.
- Stevens, C. (2011). Thermochemical processing of biomass: conversion into fuels, chemicals and power. John Wiley & Sons Ltd, United Kingdom.
- Sulaiman, S. A., Roslan, R., Inayat, M., Naz, M. Y. (2017). Effect of blending ratio and catalyst loading on co-gasification of wood chips and coconut waste. *Journal of the Energy Institute*. In Press.
- Takematsu, T., Maude, C. (1991). Coal gasification for IGCC power generation. International Energy Agency, Coal Research: London.
- Vapnik, V. (1995). The nature of statistical learning theory. Springer Verlag, New York.
- Vélez, J. F., Chejne, F., Valdés, C. F., Emery, E. J., Londoño, C. A. (2009). Co-gasification of Colombian coal and biomass in fluidized bed: an experimental study. *Fuel* 88(3):424-430.
- Villanueva, A., Gómez-Barea, A., Revuelta, E., Campoy, M., Ollero, P. (2008). Guidelines for selection of gasifiers modeling strategies. In: Proceedings of 16th European Biomass Conference and exhibition; Valencia, Spain, pp 980–986.
- Xiangdong, K., Zhong, W., Wenli, D. U., Feng, Q. I. A. N. (2013). Three stage equilibrium model for coal gasification in entrained flow gasifiers based on Aspen Plus. *Chinese Journal of Chemical Engineering* 21(1):79-84.
- Yang, S., Yang, Q., Li, H., Jin, X., Li, X., Qian, Y. (2012). An integrated framework for modeling, synthesis, analysis, and optimization of coal gasification-based energy and chemical processes. *Industrial & Engineering Chemistry Research* 51(48):15763-15777.
- Yuehong, Z., Hao, W., Zhihong, X. (2006). Conceptual design and simulation

study of a co-gasification technology. *Energy Conversion and Management* 47(11):1416-1428.

Zhang, L., Xu, S., Zhao, W., Liu, S. (2007). Co-pyrolysis of biomass and coal in a free fall reactor. *Fuel* 86(3):353-359.

Zhou, L., Zhang, G., Schurz, M., Steffen, K., Meyer, B. (2016). Kinetic study on CO₂ gasification of brown coal and biomass chars: reaction order. *Fuel* 173:311-319.

Żogała, A. (2014). Equilibrium simulations of coal gasification—factors affecting syngas composition. *Journal of Sustainable Mining* 13(2):30-38.

CHAPTER 7.

DESIGN AND DEVELOPMENT OF COMPUTATIONAL INTELLIGENCE BASED METHODOLOGIES FOR MODELING AND OPTIMIZATION OF ADSORPTIVE REMOVAL OF CHROMIUM USING SYNTHETIC POLYMER RESINS

DESIGN AND DEVELOPMENT OF COMPUTATIONAL INTELLIGENCE BASED METHODOLOGIES FOR MODELING AND OPTIMIZATION OF ADSORPTIVE REMOVAL OF CHROMIUM USING SYNTHETIC POLYMER RESINS

Abstract

Chromium is a highly toxic heavy metal often present in an ionized form in the waste water of some industries. Thus, it is an environmental and health hazard. The Gallic acid-Formaldehyde-Ammonia (GFA) based synthetic polymer resin is an attractive substance for the treatment of contaminated waste water for the removal of chromium (Cr(VI)) ions by adsorption. Ammonia enhances the cross-linked structure of the final polymer resin thus improving its adsorption capacity. Accordingly, various grades of the GFA resin were synthesized by reacting varying proportions of formaldehyde and ammonia with a fixed quantity of gallic acid. The synthesized resins were employed successfully for the adsorptive removal of Cr(VI) ions from the contaminated water. Next, a computational intelligence (CI) based *hybrid approach* was used to model and optimize the stated contaminated-water treatment reaction for securing optimal reaction conditions. The hybrid approach first uses an exclusively data driven modeling strategy, namely *artificial neural network* (ANN) to predict the adsorption (%) of Cr(VI) on the GFA resins. The input space of the ANN-based model consisting mainly of the resin synthesis conditions was optimized using a novel CI-based stochastic nonlinear optimization method based on the *artificial immune system* (AIS); the objective of the said optimization was is to maximize the adsorption percentage of Cr(VI) ions on the resin. Finally, the set of optimal resins provided by the ANN-AIS modeling-optimization hybrid strategy were experimentally verified, which resulted in an improvement of 3% in Cr(VI) ion adsorption on the GFA resins. The ANN-AIS hybrid strategy introduced here can be gainfully utilized for modeling and optimization of similar types of heavy metal ion removal processes.

7.0. INTRODUCTION

Soil and water contamination by heavy metals in industrial effluents is one of the major contributors to environmental pollution. Most of these are transition metals with variable oxidation states and coordination numbers (Mohanty et al., 2005). Efficient removal of the traces of such heavy metals from effluent water has been a major research topic in environmental engineering/science. Various treatment methods based on chemical, physical, and biological treatments have been devised till date for heavy metal removal especially from water.

Chromium is a highly toxic heavy trace metal found in the waste water of certain industries such as electroplating, dichromate and basic chromium sulphate manufacturing, tannery, anodizing, cutting tools, and chrome mining (Mulani et al., 2013). In effluent, it is present in the ionized form in two oxidation states, namely, trivalent (III) and hexavalent (VI). The toxicity of the hexavalent Cr(VI) is approximately 100 times higher than that of the trivalent Cr ions. Also, the ionic state of Cr(VI) in water is found to be more stable and this forms a major hazard to the environment as it persists in the water for a long duration if untreated. The chromium concentration commonly found in industrial waste water effluents range from less than 1 ppm to 10 ppm, while the recommended limit is upto 0.1 ppm (American Water Works Association, 1990).

Adsorption is found to be a cleaner, effective, and relatively cheaper physical method for the removal of chromium from waste water. Over the years, numerous studies have explored this technique using multiple adsorbents. The type of adsorbents utilized for the process can be broadly categorized into following areas.

- (i) Biomass materials as abundant and cheaply available bio-sorbents (treated/untreated) have been exhaustively studied for heavy metal adsorption (Dakiky et al., 2002; Garg et al., 2007; Singh and Singh, 2012; Mulani et al., 2013; Shouman et al., 2013; Amiri et al., 2014; Elangovan et al., 2015),

- (ii) Recently the search has also shifted to use of specialized materials such as synthesized nano-sorbent materials (Huang et al., 2015), chelating reagents (Godea and Pehlivan, 2003) and metal oxides (Álvarez-Ayuso et al., 2007).
- (iii) Use of specially synthesized polymers using natural or synthetic materials as adsorbents (Wu et al., 2012; Vetriselvi and Santhi, 2015; Patil-Shinde et al., 2016).

Synthetic polymer resins have a high potential in the separation of heavy metals by adsorption as they can be designed for an economical operation, by appropriately altering the synthesis process towards achieving the following objectives: (i) to get a polymer resin with highly porous cross-linked structure for maximum adsorption capacity of the heavy metal ions under mild conditions; this being done by adjusting the surface and structural chemistry of the adsorbent, (ii) to prepare a polymer resin with good structural stability, (iii) to obtain a polymer resin with good regeneration capabilities for maximum reuse. Considering costs, structural and regeneration capability constraints the objective of maximum adsorption can be achieved by proper synthesis process design (Vetriselvi and Santhi, 2015).

Synthesized gallic acid-formaldehyde resin (with ammonia as a cross-linking agent) as an adsorbent was investigated for the removal of Cr(VI) ions from water (Mulani et al., 2013). Previously, gallic acid-formaldehyde resin was investigated for the adsorption studies of palladium (II) and rhodium (III), mainly for the determination of the underlying complex mechanisms (Can et al., 2012). The present study is aimed at optimization of the synthesis reaction of the polymeric resin, namely gallic acid-formaldehyde-ammonia (GFA), and the pH of adsorption for maximal pollutant (Cr(VI) ion) removal.

7.0.1. Need of CI-Based Modeling and Optimization for Heavy Metal Ion Adsorption on Synthetic Polymer Resins

Adsorption is a selective separation phenomenon involving surface forces in which the solute molecule gets attached physically to the surface of the solid

adsorbent. The kinetic rate and final equilibrium amount of adsorption depends mainly on the following factors:

- *Adsorption process conditions*: pH, temperature, turbulence, adsorbate concentration.
- *Adsorbent structural attributes*: Surface area and porosity of the adsorbent as also its structural stability.

Additionally, for the heavy metal pollutants in ionic form, such as Cr(VI), the charge effects between the solute ions and functional groups on the adsorbent also affect the adsorption mechanism (Can et al., 2012). Due to these multiple complex and nonlinear phenomena occurring during the adsorption of heavy metal ions, it is difficult task to capture the underlying mechanisms completely in a phenomenological model. To overcome the said difficulty, computational intelligence (CI)-based methodologies provide a good platform not only for the development of data-driven reaction models but also to optimize reaction conditions for achieving superior reaction performance.

Most of the previous work on modeling and optimization of chromium adsorption on various adsorbents involved correlating and optimizing the adsorption conditions such as solution pH, temperature, adsorbent dose, and initial pollutant concentration with the percent adsorption. This was commonly done by utilizing statistical experimental design methods such as response surface methodologies (RSM) (Brasil et al., 2006; Kiran et al., 2007; Sahu et al., 2009; Aydın and Aksoy, 2009). Recently, an increasing trend of the usage of *artificial neural networks* (ANNs) for modeling the adsorption process is observed due to the ability of ANNs to accurately capture the complex phenomena underlying the adsorption/reaction. Table (7.1) shows a listing of the ANN-based modeling studies on Cr(VI) adsorption using different adsorbents. Being inherently nonlinear (and non-smooth functions), the ANN-based reaction models need stochastic strategies to optimize their input space. This requirement is also fulfilled by CI-based optimization formalisms such as *genetic algorithm* (GA), ant colony system, particle swarm, and more recent *artificial immune system* (AIS).

Synthetic polymer resins as adsorbents possess following adsorption-affecting attributes (apart from the common process conditions mentioned previously):

- The internal porous structure of the polymer resin,
- Distribution of the functional groups (active adsorption sites) on the adsorbent polymer resin,
- The exposed surface area and stability of the resin structure.

Termed as “architectural features”, these mainly vary according to the type and amount of cross-linking in the polymer during its synthesis. In this study, a CI-based hybrid modeling-optimization strategy has been implemented successfully in two stages:

- (I) In the first stage the widely used *multi-layer perceptron neural network* (MLPNN) is used for modeling the reaction conditions of the GFA polymer resins, wherein the molar composition of ingredients used in the resin synthesis and the key adsorption process parameter, i.e., adsorption pH formed model inputs (predictors) to predict the percent adsorption of the Cr(VI) ions.
- (II) A relatively recent and rarely used CI-based stochastic optimization methodology, namely *clonal selection algorithm* (CLONALG) belonging to the Artificial Immune Systems (AIS) class of algorithms is used for determination of the optimal GFA resin synthesis and resin-based adsorption conditions for the removal of Cr(VI) ions from contaminated waste-water (hereafter termed as the *optimum synthesis-adsorption conditions*).
- (III) Simultaneously, to afford a comparison of the performance of the AIS method, the popular CI-based methodology, namely, genetic algorithm (GA) was utilized to conduct optimization of the resin synthesis-adsorption conditions.

The optimized resin synthesis-adsorption conditions given by AIS were comparable with those obtained using GA. The AIS-based optimized resin synthesis-adsorption conditions were also successfully validated experimentally.

The above-stated modeling-optimization effort resulted in synthesizing a GFA resin possessing maximum Cr(VI) ion adsorption capacity for a set of specified conditions.

The next section discusses the resin preparation and experiments conducted for the resin-based adsorptive removal of Cr(VI) ions from contaminated water. The “Results and Discussion” section first describes the results of the Cr(VI) adsorption experiments using various grades of GFA resin, followed by the development of the ANN-based model. The same section explains how the ANN-based model of the reaction-adsorption is used to determine the optimal resin synthesis conditions in the framework of AIS-based and GA optimization methodologies. The section ends with the results of the experimental validation of the AIS-based sets of the optimum resin synthesis conditions yielded by the ANN-GA and the ANN-AIS hybrid strategies. The final section, “Conclusion” summarizes the principal findings of the study.

Table 7.1: Existing ANN-based models for predicting the amount of Cr(VI) ions adsorption on different adsorbent materials

Sr. No.	Adsorbent Materials	Objective	Reference
1	Brown seaweed, Ecklonia Biomass	Modeling of the adsorption breakthrough curves of packed bed adsorption column operations.	Park et al. (2006)
2	Zea mays	Modeling of batch sorption processes for the prediction of Cr(III) and Cr(VI) removal efficiency.	Kardam et al. (2011)
3	<i>Bacillus</i> sp.	Developing models predicting the biosorption efficiency for the removal of Cr, with inputs as pH, contact time and initial concentration of Cr.	Masood et al. (2012)
4	Zeolite prepared from raw fly ash	Modeling and optimization of the adsorption process of Cr(VI) ions using process variables such as the initial pH, adsorbent dosage, temperature and contact time.	Asl et al. (2013)
5	Isolate Bacillus aryabhatai	Optimization of the adsorption process variables (pH, temperature, biomass dose and initial Cr(VI) ion concentration) using the developed ANN-based model through genetic algorithm.	Verma et al. (2014)
6	Activated granular carbon from coconut shell	Process modeling of adsorptive removal of Cr(VI) from batch sorption experiments and optimization of parameters viz. pH, contact time, adsorbent dose, initial concentration, and temperature using response surface methodology.	Halder et al. (2015)
7	Powdered activated carbon	Process modeling of removal of Cr(VI) on powdered activated carbon with process parameters, namely the solution pH, adsorbent dose, initial pollutant concentration and contact time.	Anupam et al. (2016)

7.1. EXPERIMENTAL SECTION

7.1.1. Preparation of Gallic Acid-Formaldehyde-Ammonia (GFA) Resin

For the preparation of various grades of the GFA resin, 4 g of gallic acid (LOBA CHEMIE, Pvt. Ltd. Mumbai, India) and required amount (2.5, 5 and 10 mL) of 37% formaldehyde solution (QUALIGENS, India) were added and stirred for five minutes to achieve uniform mixing. To this mixture, 25 wt% ammonia solution of desired amount (10 and 20 mL) was added and stirring was continued for further five minutes. The resulting reaction mixture was kept for fifteen days, after which a precipitate was observed at the bottom of the plastic container. The reaction mixture was neutralized with 1 M hydrochloric acid solution and the resulting suspension was filtered through an ordinary filter paper. The precipitate was treated with 1.2 M hydrochloric acid to make it insoluble in both acidic and basic medium. Finally, hydrochloric acid solution was filtered and the residue was washed with de-ionized water, followed by drying at 80°C for 8 hours to obtain the insoluble GFA resin. The five resin grades along with the initial composition of the ingredients used for their synthesis are given in Table (7.A.1). These resins were then utilized for the adsorption of Cr(VI) ions from contaminated water in batch equilibrium adsorption experiments discussed next.

7.1.2. Equilibrium Adsorption Experiments of Cr(VI) on GFA Resins

Equilibrium batch adsorption experiments were conducted to investigate the adsorption of Cr(VI) ions by GFA resins. In these experiments, the effect of pH on Cr(VI) adsorption was rigorously studied. Here, 250 mg of resin was immersed into 10 mL (52.8 ppm) Cr(VI) stock solutions with differing pH values as indicated in Table (A.7.1). The stock solution of Cr(VI) was prepared by dissolving accurately weighed potassium dichromate ($K_2Cr_2O_7$, LOBA CHEMIE Pvt. Ltd. Mumbai, India) in 100 mL deionized water. Milli-Q water purification system (MILLIPORE, USA) was used to get the deionised water. The test mixture was stirred at room temperature for 24 hours. The concentration of Cr(VI) ions in the effluent was determined spectrophotometrically by the development of a purple color with 1,5-diphenyl carbazide in an acidic solution as the complexing agent. A solution (0.025%) of 1,5-diphenylcarbazide (SIGMA-ALDRICH, USA)

was prepared by dissolving 25 mg of 1,5-diphenylcarbazine in 5 mL acetone (MERCK) and 10 mL, 5 M H₂SO₄. The resulting mixture was diluted to 100 mL with deionised water and stored in a glass bottle. The absorbance of the purple colored solution was observed at a wavelength of 540 nm after 20 minutes against a blank. Here, Beer's law holds well in the concentration range 0.08-0.51 mg/mL with the regression equation of $a = 4.332 c$, possessing a correlation coefficient of 0.997, where, a is the absorbance and c is the concentration of the Cr(VI) ions in water.

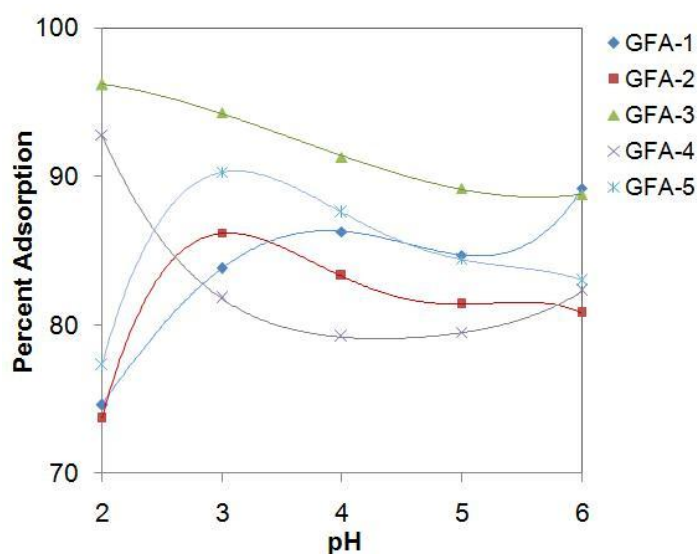


Figure 7.1: Variation of percent adsorption of Cr(VI) ions on various grades of GFA resins with pH of the solution.

The results of the batch equilibrium adsorption experiments conducted for the removal of Cr(VI) from the aqueous solutions are tabulated in Table (7.A.1) for various grades of the GFA resin. This table indicates the grades in terms of the varying amounts (moles) of ammonia and formaldehyde (with a fixed amount of gallic acid) used in synthesizing the specific resin grade. Figure (7.1) shows the variation in the adsorption (%) of Cr(VI) ions on various grades of GFA resins at different solution pH magnitudes. The figure clearly indicate a nonlinear trend in the adsorption of Cr(VI) ions for all grades of the resin and varying solution pH magnitudes. This nonlinear character creates a need to explore CI-based inherently nonlinear, data-driven methodologies such as ANNs for modeling of the resin-based Cr(VI) adsorption.

7.2. RESULTS AND DISCUSSION

7.2.1. MLPNN Based Modeling of Adsorption of Cr(VI) on GFA Resins

For ANN based modeling of the resin synthesis reaction and Cr(VI) adsorption the input (predictor) space consisted of following resin and adsorption variables/parameters: moles of formaldehyde (x_1) and moles of ammonia (x_2) used for the resin synthesis and the adsorption pH (x_3), while the model output was the percent adsorption (y) of Cr(VI) ions on the resin. For developing the multilayer perceptron neural network (MLPNN)-based model, the GFA synthesis reaction and adsorption experimental data were randomly shuffled and divided into training and test sets in the ratio 75:25. The model was developed using *Rapidminer* software package (Mierswa et al., 2006). The architecture and parameters of the developed MLPNN model were optimized heuristically; the best model was selected on the basis of high values of *coefficient of correlation (CC)* and low values of *root mean squared error (RMSE)* corresponding to the model predictions with regard to both training and test set data. Table (7.2) lists the architectural details and the parameters used for developing the best MLPNN-based model. The prediction performance and generalization capability of the model are listed in Table (7.3). The high *CCs* (≥ 0.93) and low *RMSEs* (≤ 2.88) magnitudes listed in the table are indicative of MLPNN model's good prediction and generalization performance. Figure (7.2) shows the parity plot of the experimental versus model predicted Cr(VI) adsorption (%) values in which a reasonably good match can be seen between the two adsorption quantities for both the training and test datasets.

Table 7.2: Structural details of the developed MLPNN-based model predicting the percent adsorption of Cr(VI) ions on the GFA resin

Input nodes	Training cycles	Hidden layer nodes	η	μ	TF for hidden layer	TF for output layer
3	700	4	0.2	0.029	Logistic Sigmoid	Linear

η : Learning rate, μ : Momentum coefficient, TF: Transfer function

Table 7.3: Statistical performances of MLPNN-based model predicting the percent adsorption of Cr(VI) on the GFA resin

Resin Type	Training-set		Test-set	
	CC_{Trn}	$RMSE_{Trn}$	CC_{Tst}	$RMSE_{Tst}$
GFA	0.93	2.13	0.94	2.88

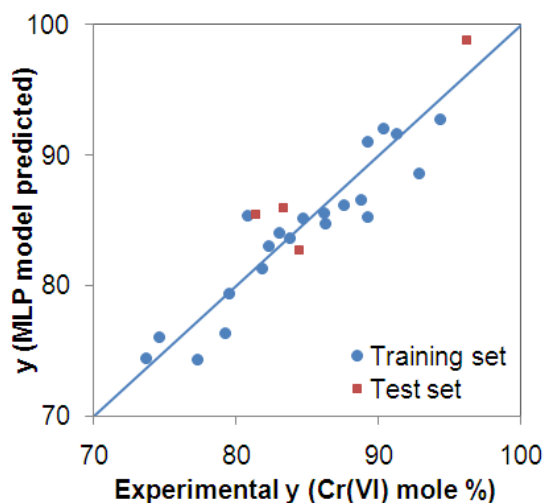


Figure 7.2: Parity plots of experimental versus MLPNN model predicted values of the percent adsorption of Cr(VI) ions on various grades of the GFA resin at different pH.

7.2.2. CI-Based Stochastic Optimization of Adsorption of Cr(VI) on GFA Resins

The artificial immune system optimization algorithms are based on the functioning of biological immune system (BIS). The BIS involves several interrelated biological mechanisms capable of counteracting the invasion of pathogenic microorganisms (antigens) by producing antigen-specific antibodies in living beings. The “clonal selection” mechanism of the BIS mainly comprises *cloning* and *hypermutation* of initially generated random antibodies to drive them to maturity—producing high-affinity antibodies—in an iterative manner. The AIS analog of the “clonal selection” mechanism is termed *clonal selection algorithm* (CLONALG) (Burnet, 1978), and is used for performing stochastic optimization and pattern recognition (de Castro and Zuben, 2001). More details of the AIS and CLONALG can be found in Section (1.4.2 (A)).

The MLPNN-based Cr(VI) ion adsorption (percentage) predicting model was used as the objective function (to be maximized) in the AIS-CLONALG optimization scheme to secure the optimized values of the model inputs (moles of formaldehyde, moles of ammonia and pH) resulting in the maximum adsorption of Cr(VI) ions on the resin. The binary AIS-CLONALG optimization algorithm developed in MATLAB[®] (MATLAB, 2000) (see the Code 7.A.1 for the MATLAB code in Appendix 7.A on page 229) was designed specifically for the stated optimization. The optimization objective and the associated constraints are given as:

$$\text{Maximize } y = f(x_1, x_2, x_3) \quad (7.1)$$

$$\text{Subject to } 0 < y < 100; \quad x_{iL} < x_i < x_{iU} \quad (7.2)$$

where, y = percent adsorption of Cr(VI) ions, x_1 = moles of formaldehyde, x_2 = moles of ammonia, x_3 = pH, x_{iL} = lower bound of variable x_i ($i = 1, 2, 3$) and x_{iU} = upper bound of variable x_i . The function ' f ' represents the objective function that appropriately includes the MLPNN-based model of the process under study. The constraints for the optimization are: $x_{1L} = 0.0308$ gmole, $x_{1U} = 0.1232$ gmole, $x_{2L} = 0.1467$ gmole, $x_{2U} = 0.2935$ gmole, $x_{3L} = 2$, $x_{3U} = 6$; output ' y ' constrained between 0 and 100%. In order to compare the performance of CLONALG optimization method, the mentioned optimization task defined in Eqns (7.1) and (7.2) was also performed using the widely used CI-based stochastic optimization methodology, namely genetic algorithm (GA). The GA-based optimization was performed using a GA add-in for Microsoft excel, namely *Mendelsoft* (MendelSolve, 2016). For more details of the GA formalism please refer to Section (1.4.1).

Several runs of both the CI-based stochastic optimization methods were performed in order to secure the best optimal values of the inputs. The parameter values used in the optimization runs using AIS-CLONALG and GA methods are given in Tables (7.4) and (7.5), respectively. Both AIS-CLONALG and GA yielded comparable optimized values of the three model inputs (x_1, x_2, x_3) that would lead to maximum percent adsorption of Cr(VI) ions (see Table 7.6). These

optimized solutions were subjected to experimental validation the results of which are also presented in Table (7.6). As can be seen, the AIS-CLONALG based optimum values of the three inputs yielded marginally better Cr(VI) percent adsorption value (99.2%) as compared to that given by the GA (99%). It is thus clear that the usage of ANN-AIS hybrid modeling strategy has resulted in improving the Cr(VI) adsorption percentage from 96.2 (experiment number 24 in Table 7.A.1 on page 227) to 99.2.

Table 7.4: Selected parameters of AIS-CLONALG implemented for the grade optimization for Cr(VI) ions adsorption on GFA reins

N	N_{kept}	B	ρ	E	G
20	10	1	2	2	20

N : Population, N_{kept} : Population Selected, B : Cloning factor, ρ : Hypermutation coefficient, E : Elite count, G : Generations

Table 7.5: Selected parameters of GA implemented for the grade optimization for Cr(VI) ions adsorption on GFA reins.

N	Trails	Crossover rate	Mutation rate	G
20	2	1	0.03125	25

N : Population, G : Generations

Table 7.6: Optimized grade conditions for the GFA resin synthesis and adsorption process pH for total adsorption of the Cr(VI) ions by the resin

Formaldehyde moles (x_1)		Ammonia moles (x_2)		pH (x_3)		Experimentally validated Cr(VI) adsorption (%)
AIS	GA	AIS	GA	AIS	GA	
0.11	0.10	0.15	0.15	4	3.9	99.2

*Adsorbent dose: 250 gm, Initial Chromium concentration in solution: 52.8 ppm

7.3. CONCLUSION

Chromium in the form of Cr(VI) ions creates a major pollution hazard due to its presence in the effluent water of some chemical industries. Adsorption is an

economical and clean treatment technique to tackle the said chromium pollution problem. It is known that cross-linking in the polymer resins improves the adsorption capacity of the resins by creating a highly porous internal structure with more active adsorption sites becoming accessible to the adsorbate. Accordingly, in the current study, various grades of the gallic acid-formaldehyde-ammonia (GFA) polymer resins (having a cross-linked structure) were investigated for the adsorptive removal of Cr(VI) ions from the contaminated water. With ammonia as the cross-linking agent, the main emphasis was on modeling and optimization of the GFA synthesis reaction conditions for determination of the optimal GFA grade for maximum adsorption of Cr(VI) ions. For this purpose, a CI-based hybrid modeling-optimization approach was used to model and optimize the GFA resin synthesis reaction conditions, namely, molar composition of the ingredients used in the reaction, and pH of the Cr(VI) ion adsorption on the GFA resins. The CI-based modeling of the stated system resulted in an MLPNN-based model predicting Cr(VI) adsorption (%) with high prediction accuracy and generalization capability ($CCs \geq 0.93$, $RMSEs \leq 2.88$). The input space of the developed MLPNN-based nonlinear model consisting of three predictors namely moles of formaldehyde, moles of ammonia used in the GFA polymer synthesis and pH of the adsorption of Cr(VI) ions on the resin, was then optimized using two CI-based stochastic optimization schemes namely artificial immune system based clonal selection algorithm (AIS-CLONALG) and genetic algorithm (GA). The principal advantage of the two hybrid MLPNN-AIS and MLPNN-GA modeling-optimization approaches is that the optimization can be performed exclusively from the data pertaining to the resin synthesis reaction and Cr(VI) adsorption on the resin without invoking the detailed knowledge of the underlying physicochemical phenomena. Both the optimization methods yielded similar solutions. The optimized conditions when verified experimentally, resulted in 3% increase in the GFA-based adsorption (%) of Cr(VI) metal ions. These results notably demonstrate the potential of CI-based methodologies for (a) developing a single model representing resin synthesis and Cr(VI) adsorption, and (b) optimization of the reactant composition of the resin synthesis reaction and Cr(VI) adsorption conditions. The proposed modeling-optimization approach can be fruitfully extended to similar waste-water treatment processes.

APPENDIX 7.A

Table 7.A.1: Experimental dataset of Cr(VI) adsorption on different grades of GFA resins

Sr. no.	Resin grade code*	Formaldehyde moles ^a (x_1)	Ammonia moles ^b (x_2)	pH (x_3)	Cr(VI) Adsorption (%) (y)
1	GFA-1	0.0308	0.2935	2	74.6
2	GFA-1	0.0308	0.2935	3	83.8
3	GFA-1	0.0308	0.2935	4	86.3
4	GFA-1	0.0308	0.2935	5	84.7
5	GFA-1	0.0308	0.2935	6	89.2
6	GFA-2	0.0616	0.2935	2	73.7
7	GFA-2	0.0616	0.2935	3	86.2
8	GFA-2	0.0616	0.2935	6	80.8
9	GFA-3	0.1232	0.2935	3	94.3
10	GFA-3	0.1232	0.2935	4	91.3
11	GFA-3	0.1232	0.2935	5	89.2
12	GFA-3	0.1232	0.2935	6	88.8
13	GFA-4	0.0308	0.1467	2	92.8
14	GFA-4	0.0308	0.1467	3	81.8
15	GFA-4	0.0308	0.1467	4	79.2
16	GFA-4	0.0308	0.1467	5	79.5
17	GFA-4	0.0308	0.1467	6	82.3
18	GFA-5	0.0616	0.1467	2	77.3
19	GFA-5	0.0616	0.1467	3	90.3
20	GFA-5	0.0616	0.1467	4	87.6

Sr. no.	Resin grade code*	Formaldehyde moles^a (x_1)	Ammonia moles^b (x_2)	pH (x_3)	Cr(VI) Adsorption (%) (y)
21	GFA-5	0.0616	0.1467	6	83
22 [#]	GFA-2	0.0616	0.2935	4	83.3
23 [#]	GFA-2	0.0616	0.2935	5	81.4
24 [#]	GFA-3	0.1232	0.2935	2	96.2
25 [#]	GFA-5	0.0616	0.1467	5	84.4

*Gallic acid: 4 gm, ^aFormaldehyde quantities used: 2.5, 5, 10 mL, ^bAmmonia quantities used: 10, 20 mL, [#]Test data; Adsorbent dose: 250 gm, Initial Chromium concentration in solution: 52.8 ppm.

Code 7.A.1: MATLAB code for CLONALG-based optimization of percent adsorption of Cr(VI) by GFA resin

Code Description:

The following CLONALG-based MATLAB (version 7) code is written for the computation of the optimum adsorption conditions for maximum adsorption of the Cr(VI) ions by the GFA resins. The main function starts with the “main_file.m” MATLAB file, which invokes a list of CLONALG-based functions in an iterative manner till the termination criteria is met.

main_file.m

```
clear;                                %AIS-CLONALG for optimization

clc;

global N_BITS N_VARS N_POP AB LENGHT BETA N_C N_M N_R SEL_P
HMUTATE_COEFF N_ROWS_KEPT N_ROWS_FILL N_ELITE X_MIN X_MAX
X_RANGE X_DiffMin;

N_BITS=10; N_VARS=3; N_POP=20;    %Naming convention:
%n=number, bin:Binary, y,x,X=decimal col vector/matrix,
%IC:Initial Conditions, _P:probability %number of bits per
%variables

SEL_P=0.4; CROSS_P = 0.8; HMUTATE_COEFF = 2.5;
%SEL_P=Population kept

AB LENGHT = N_BITS * N_VARS; N_ROWS_KEPT =
round(SEL_P*N_POP);    %N_ROWS_KEPT...high affinity antibodies

N_M = N_ROWS_KEPT;

BETA = 1;

X_RANGE = [0.0308 0.1232;0.1468 0.2935;2 6];

%<Enter this: Ranges for [x1(Formaldehyde); x2(Ammonia);
%x3 (pH) ]
```

```

X_DiffMin = X_RANGE;

X_DiffMin(:,1) = X_RANGE(:,2) - X_RANGE(:,1);

X_DiffMin(:,2) = X_RANGE(:,1);

N_ELITE=2;

N_ROWS_FILL = N_POP - N_ROWS_KEPT;
%N_pop_offsprings

n_gen=20;

IC_bin_pop_x=round(rand(N_POP,AB LENGHT)); % Generate random
                                         %initial population:

next_gen_bin_x=IC_bin_pop_x;

i=1;

while i <= n_gen

    X_DecBin=BintoDecColumns(next_gen_bin_x);

    X_norm = X_DecBin/((2^N_BITS) - 1);          %Normalize

    X=denormalize_matGA_Cr_Ads(X_norm);        %DeNormailze

    y=funcgenGA_Cr_MolarGFAANN(X);

    yXXbin = [y X next_gen_bin_x];

    sorted_yXXbin = sortrows(yXXbin);

    sel_rows_for_nextpopyyXXbin = sorted_yXXbin(1:N_POP,:);

    sel_rows_for_cloninyXXbin =
    sel_rows_for_nextpopyyXXbin(1:N_ROWS_KEPT,:); % y,x,

    [clonedABsXbin n_c] =
    cloningABs(sel_rows_for_cloninyXXbin);      % returns
    %the main matrix and number of clones

```

```

maturatedClonesXbin =
hypermutateABs(sel_rows_for_cloningyXXbin, n_c,
clonedABsXbin);

random_Xbin = round(rand(N_POP,AB LENGHT)); % Generate
%random initial population

next_gen_bin_x = [maturatedClonesXbin;random_Xbin];

i = i + 1

end

gen = i - 1

sel_rows_for_nextpopyXXbin(:,1:4)

```

BintoDecColumns.m

```

function x=BintoDecColumns(pop,n) %n=n_var

global N_VARS,

[M,N]=size(pop);

bitlength=N/N_VARS;

x(M,N_VARS)=0;

for row=1:M

    for col=1:N_VARS

        k=bitlength*(col-1)+1; %k=1,9,17,...for 8-bit
%length

        y=0;

        p=bitlength-1;

        for i=k:1:k+bitlength-1 %increment bit-index

```

```

        y=2^p*(pop(row,i))+y;

        p=p-1;

        end

        x(row,col)=y;

        end

end

```

denormalize_matGA_Cr_Ads.m

```

function Xcol = denormalize_matGA_Cr_Ads(X_norm)
                                %DeNormalize a col matrix

global N_POP N_VARS X_DiffMin;

[a,b]=size(X_norm);

Xcol=zeros(a,b);

for i=1:N_VARS

Xcol(:,i) = X_norm(:,i).*X_DiffMin(i,1) + X_DiffMin(i,2);

end

```

funcgenGA_Cr_MolarGFAANN.m

```

function [p c] = funcgenGA_Cr_MolarGFAANN(x)

xsize=size(x);

```

```

p=zeros(xsize(1),1);

OP=zeros(xsize(1),1);

f=zeros(xsize(1),1);

load Cr_GA_GFA_ANNDDataLAT.mat

x_norm=x;

for i=1:xsize(1)

    x_norm(i,1) = 2*(x(i,1) -
    min(ds(1:21,1)))/(max(ds(1:21,1))-min(ds(1:21,1))) -
    1;

    x_norm(i,2) = 2*(x(i,2) -
    min(ds(1:21,2)))/(max(ds(1:21,2))-min(ds(1:21,2))) -
    1;

    x_norm(i,3) = 2*(x(i,3) -
    min(ds(1:21,3)))/(max(ds(1:21,3))-min(ds(1:21,3))) -
    1;

    y1=x_norm(i,:);

    y1=[y1 1];

    y2=y1*L1;

    y2=logsig(y2);

    y2=[y2 1];

    y3=y2*L2;

    y3=purelin(y3);

    OP(i,:)=y3;

end

p=postmnmx(OP,73.7,94.3);

f = abs(100 - p);

```

```
p = f;
```

cloningABs.m

```
function [p n_c] = cloningABs(sel_rows_for_cloningyXXbin)
```

```
global AB LENGHT N_VARS BETA N_POP N_C N_M N_R;
```

```
ABs_selected =
```

```
sel_rows_for_cloningyXXbin(:, (1+N_VARS+1):end);
```

```
j=1;
```

```
n = 1 + N_VARS + AB LENGHT;
```

```
AB_clones = 0.*ABs_selected;
```

```
Nc=zeros(N_M,1);
```

```
for i = 1:N_M
```

```
    Nc(i) = round((BETA*N_POP)/i);
```

```
end
```

```
nnc = sum(Nc);
```

```
AB_clones_all = zeros(nnc,AB LENGHT);
```

```
s = 0;
```

```
for i = 1:N_M
```

```
    for m = 1:Nc(i)
```

```
        AB_clones(m,:) = ABs_selected(i,:);    %append all  
                                                %clones in single matrix
```

```
        s = s + 1;
```

```
        AB_clones_all(s,:)=AB_clones(m,:);
```

```

        end

    end

    p = AB_clones_all;

    n_c = Nc;

    -----

```

hypermutateABs.m

```

    -----

    function p = hypermutateABs(sel_rows_for_cloningyXXbin, Nc,
    next_gen_xbinUM)

    global HMUTATE_COEFF AB LENGHT N_ROWS_KEPT N_ELITE;

    [A,B] = size(sel_rows_for_cloningyXXbin);

    y = sel_rows_for_cloningyXXbin(:,1);

    n = N_ROWS_KEPT;

    m = 0;

    h = zeros(n,1);

    hmutp = zeros(n,1);

    ymin = min(y); ymax = max(y); range = ymax-ymin;

    for i = 1:n

        h(i) = abs(1-((y(i)-ymin)/range));

        hmutp(i) = exp(-HMUTATE_COEFF*h(i));

    end

    c = [sel_rows_for_cloningyXXbin h hmutp];

    for i = 1:n

        for j = 1:Nc(i)

```

```

        m = m + 1;

        r = rand(1,1)*100;

        if m > N_ELITE

            if r > c(i,end)

                rand_mutation_pt=ceil(rand(1,1)*AB LENGHT);
                next_gen_xbinUM(m,rand_mutation_pt)=abs(next_gen_xbinUM
                M(m,rand_mutation_pt)-1);

            end

        end

    end

end

end

end

p=next_gen_xbinUM;

```

NOMENCLATURE

N	Population size
N_{kept}	Population Selected
X	Dependent variable
G	Generations
Y	Independent variable (here % adsorption of Cr(VI) ions)

Greek letters

μ	Momentum coefficient in the EBP algorithm
ρ	Hyper-mutation coefficient

- β Cloning factor in CLONALG
- η Learning rate in the EBP algorithm

REFERENCES

- Alvarez-Ayuso, E., Garcia-Sanchez, A., Querol, X. (2007). Adsorption of Cr (VI) from synthetic solutions and electroplating wastewaters on amorphous aluminium oxide. *Journal of Hazardous Materials* 142(1):191-198.
- Amiri, M. J., Fadaei, E., Baghvand, A., Ezadkhasty, Z. (2014). Removal of heavy metals Cr (VI), Cd (II) and Ni (II) from aqueous solution by bioadsorption of *Elaeagnus angustifolia*. *International Journal of Environmental Research* 8(2):411-420.
- Anupam, K., Dutta, S., Bhattacharjee, C., Datta, S. (2016). Artificial neural network modelling for removal of chromium (VI) from wastewater using physisorption onto powdered activated carbon. *Desalination and Water Treatment* 57(8):3632-3641.
- Asl, S. H., Ahmadi, M., Ghiasvand, M., Tardast, A., Katal, R. (2013). Artificial neural network (ANN) approach for modeling of Cr (VI) adsorption from aqueous solution by zeolite prepared from raw fly ash (ZFA). *Journal of Industrial and Engineering Chemistry* 19(3):1044-1055.
- Aydin, Y. A., Aksoy, N. D. (2009). Adsorption of chromium on chitosan: Optimization, kinetics and thermodynamics. *Chemical Engineering Journal* 151(1):188-194.
- Brasil, J. L., Ev, R. R., Milcharek, C. D., Martins, L. C., Pavan, F. A., Dos Santos, A. A., Dias, S. L. P., Dupont, J., Norena, C. P. Z., Lima, E. C. (2006). Statistical design of experiments as a tool for optimizing the batch conditions to Cr (VI) biosorption on *Araucaria angustifolia* wastes. *Journal of Hazardous Materials* 133(1):143-153.
- Burnet, F. M. (1978). Clonal selection and after. In: *Theoretical Immunology*,

Bell, G. I., Perelson, A. S., Pimbley, G. H. Eds. New York: Marcel Dekker, pp. 63–85.

Can, M., Bulut, E., Özacar, M. (2012). Synthesis and characterization of gallic acid resin and its interaction with palladium (II), rhodium (III) chloro complexes. *Industrial & Engineering Chemistry Research* 51(17):6052-6063.

Dakiky, M., Khamis, M., Manassra, A., Mer'eb, M. (2002). Selective adsorption of chromium (VI) in industrial wastewater using low-cost abundantly available adsorbents. *Advances in environmental research* 6(4):533-540.

De Castro, L. N., Von Zuben, F. J. (2002). Learning and optimization using the clonal selection principle. *IEEE transactions on evolutionary computation* 6(3):239-251.

Elangovan, N. S., Lavanya, V., Arunthathi, S. (2015). Removal of Chromium From Ground water using Neem Leaves as Adsorbent. *International Journal of Environmental Research* 9(2):439-444.

Garg, U. K., Kaur, M. P., Garg, V. K., Sud, D. (2007). Removal of hexavalent chromium from aqueous solution by agricultural waste biomass. *Journal of Hazardous Materials* 140(1):60-68.

Gode, F., Pehlivan, E. (2003). A comparative study of two chelating ion-exchange resins for the removal of chromium (III) from aqueous solution. *Journal of Hazardous Materials* 100(1):231-243.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. 1st ed., Reading: Addison-Wesley, New York, USA.

Gregory, R., Zabel, T. (1990). *Water Quality and Treatment: A Handbook of Community Water Supplies*, American Water Works Association, McGraw-Hill, New York, USA.

Halder, G., Dhawane, S., Barai, P. K., Das, A. (2015). Optimizing chromium (VI) adsorption onto superheated steam activated granular carbon through

response surface methodology and artificial neural network. *Environmental Progress & Sustainable Energy* 34(3):638-647.

Holland, J. H. (1975). *Adaptation in natural and artificial systems. An introductory analysis with application to biology, control, and artificial intelligence.* University of Michigan Press, Ann Arbor, USA.

Huang, Z. N., Wang, X. L., Yang, D. S. (2015). Adsorption of Cr (VI) in wastewater using magnetic multi-wall carbon nanotubes. *Water Science and Engineering* 8(3):226-232.

Kardam, A., Raj, K. R., Arora, J. K., Srivastava, S. (2011). ANN modeling on predictions of biosorption efficiency of zea mays for the removal of Cr (III) and Cr (VI) from waste water. *International Journal of Mathematics Trends and Technology* 2(1):23-29.

Kiran, B., Kaushik, A., Kaushik, C. P. (2007). Response surface methodological approach for optimizing removal of Cr (VI) from aqueous solution using immobilized cyanobacterium. *Chemical Engineering Journal* 126(2):147-153.

Masood, F., Ahmad, M., Ansari, M. A., Malik, A. (2012). Prediction of biosorption of total chromium by *Bacillus* sp. using artificial neural network. *Bulletin of environmental contamination and toxicology* 88(4):563-570.

MATLAB (2000) The MathWorks Inc., Natick, MA, www.mathworks.com/Products/MATLAB.

MendelSolve (2016) A Genetic Algorithm Solver. <http://www.bluestretch.com/mendelsolve/index.htm>. Accessed: 17 June 2016

Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., Euler, T. (2006). YALE: rapid prototyping for complex data mining tasks, In: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery*

and Data Mining (KDD-06).

- Mohanty, K., Jha, M., Meikap, B. C., Biswas, M. N. (2005). Removal of chromium (VI) from dilute aqueous solutions by activated carbon developed from Terminalia arjuna nuts activated with zinc chloride. *Chemical Engineering Science* 60(11):3049-3059.
- Mulani, K., Daniels, S., Rajdeo, K., Tambe, S., Chavan, N. (2013). Adsorption of chromium (VI) from aqueous solutions by coffee polyphenol-formaldehyde/acetaldehyde resins. *Journal of Polymers*, Article ID: 798368.
- Park, D., Yun, Y. S., Lee, D. S., Lim, S. R., Park, J. M. (2006). Column study on Cr (VI)-reduction using the brown seaweed Ecklonia biomass. *Journal of hazardous materials* 137(3):1377-1384.
- Patil-Shinde, V., Mulani, K. B., Donde, K., Chavan, N. N., Ponrathnam, S., Tambe, S. S. (2016). The Removal of arsenite [As (III)] and arsenate [As (V)] ions from wastewater using TFA and TAFA resins: Computational intelligence based reaction modeling and optimization. *Journal of Environmental Chemical Engineering* 4(4):4275-4286.
- Sahu, J. N., Acharya, J., Meikap, B. C. (2009). Response surface modeling and optimization of chromium (VI) removal from aqueous solution using Tamarind wood activated carbon in batch process. *Journal of hazardous materials* 172(2):818-825.
- Shouman, M. A., Fathy, N. A., Khedr, S. A., Attia, A. A. (2013). Comparative biosorption studies of hexavalent chromium ion onto raw and modified palm branches. *Advances in Physical Chemistry*, Article ID: 159712.
- Singh, S. R., Singh, A. P. (2012). Treatment of water containing chromium (VI) using rice husk carbon as a new low cost adsorbent. *International Journal of Environmental Research* 6(4):917-924.
- Verma, D. K., Hasan, S. H., Singh, D. K., Singh, S., Singh, Y. (2014). Enhanced

biosorptive remediation of hexavalent chromium using chemotailored biomass of a novel soil isolate *Bacillus aryabhatai* ITBHU02: process variables optimization through artificial neural network linked genetic algorithm. *Industrial & Engineering Chemistry Research* 53(9):3669-3681.

Vetriselvi, V., Santhi, R. J. (2015). Redox polymer as an adsorbent for the removal of chromium (VI) and lead (II) from the tannery effluents. *Water Resources and Industry* 10:39-52.

Wu, Z., Li, S., Wan, J., Wang, Y. (2012). Cr (VI) adsorption on an improved synthesized cross-linked chitosan resin. *Journal of Molecular Liquids* 170:25-29.

CHAPTER 8.

DEVELOPMENT OF COMPUTATIONAL INTELLIGENCE BASED CLUSTERING MODEL FOR FAULT DETECTION AND DIAGNOSIS IN BIOCHEMICAL PROCESS

DEVELOPMENT OF COMPUTATIONAL INTELLIGENCE BASED CLUSTERING MODEL FOR FAULT DETECTION AND DIAGNOSIS IN BIOCHEMICAL PROCESS

Abstract

Process fault detection and diagnosis (FDD) (i.e. identifying the *normal* and *abnormal* behavior of a process and causes thereof) are the major challenges in all types of chemical industries. In any process, a malfunction or a fault may develop due to significant deviations in the operating variables and/or parameters. The causes of these deviations are, for instance, malfunctioning sensors, and/or faulty equipment such as pumps, heaters, controllers, valves, etc. For conducting FDD, an accurate model is needed using which it is possible to differentiate between “normal” and “faulty” process behavior speedily. Biochemical processes are highly sensitive to the changes in the process operating conditions and even small changes in these may result in a process/batch failure. Accordingly, this study presents results of fault detection and diagnosis of a continuously stirred controlled bioreactor using a computational intelligence (CI)-based methodology, namely *fuzzy c-means clustering* (FCC) and the conventional *c-means clustering* methods. The process under study generates biomass from a substrate. The process data comprising the normal and faulty operations of the biochemical process was used to model four types of faults. The clustering methods use unsupervised learning and depending upon the type of fault, cluster the corresponding data into one of the five possible clusters (one each for four types of faults and the normal process operation). The results of this study clearly indicate that the FCC-based process model could successfully categorize the normal and faulty bioreactor behavior into appropriate clusters with better accuracy than the conventional c-means algorithm.

8.0. INTRODUCTION

A process fault indicates those process symptoms, which result from undesired physical changes in the process. In chemical process industries, occurrences of faults (equipment malfunctions/abnormal process conditions) may lead to unwanted situations such as deterioration in the process performance (e.g., lower product quality, reduced conversion/yield, etc.), damage to process equipment, increase in the environmental pollution, and in worst cases, catastrophic accidents coupled with injuries and/or loss of human life. These scenarios may potentially lead to severe financial losses and damage/injuries to the plant and even general public. Hence, over the years, online fault detection and diagnosis have been given a serious attention by researchers and chemical engineers (Caccavale and Pierri, 2009). In a *fault detection and diagnosis* (FDD) task, the first objective is to recognize non-permissible deviations from the set points, disturbances or equipment/sensor/actuator malfunction followed by investigating the fault cause-effect relationship and extent of the malfunction. Overall, it is a three-step process:

1. Recognition of the occurrence of a fault/malfunction.
2. Locating the fault origin.
3. Determination of the extent of fault.

The first activity is referred to as *fault detection*, while the other two activities are part of *fault diagnosis*. Traditionally, process history data were used in statistical monitoring of the process with threshold checking and/or trend checking in the FDD task. Broadly, the FDD methods are classified as *model* and *history* based (Dash and Venkatasubramanian, 2000). During the past few decades, various CI-model based FDD methods have been extensively developed and successfully applied to engineering systems (Himmelblau, 1978; Pau et al., 1981; Tambe et al., 1996; Sobhani-Tehrani and Khorasani, 2009).

Model-based FDD techniques involve use of an accurate model of the process for monitoring the actual state of the process. It is based on the idea that if the inputs, as given to the process are also supplied to the process model, the

model and actual process outputs must match within a certain limit. A non-permissible deviation between the model-predicted output values and the actual process values indicates a fault/malfunction. Traditionally, phenomenological and conventional empirical models (such as regression based) are used for the model based FDD. However, due to the lack of an in-depth understanding of the process under study (necessary in “first principles” modeling) and the enormous heuristic effort needed in empirical modeling, these approaches are often found unwieldy for model-based FDD. Thus, the trend is shifting towards CI-based FDD approaches due to the numerous attractive features of the CI-based modeling methods (see Chapter 1).

Once the fault is detected and identified, its intensity can be inferred from the measurement of the extent of deviation of the affected process variable from its normal (set) value. Typically, depending on the normal and faulty process behavior, the corresponding process variable values can be categorized into distinct ‘groups’ or ‘clusters’. This type of grouping is possible using unsupervised clustering (pattern recognition) modeling. Having identified the clusters corresponding to the normal process behavior and various faults, the constructed clustering model can be used in the prediction mode for detecting and diagnosing faults.

8.0.1. Pattern Recognition/Clustering for Process FDD

A pattern is a particular arrangement of objects under study. It can also consist of real-valued numbers. Pattern recognition involves identification of distinct patterns/groups/classes/categories of similar objects; in process FDD studies an object is a vector ‘ \mathbf{x} ’ of the process variable/parameter values, x_i ($i = 1, 2, \dots, N$). The representation space denoted by \mathbf{X} is the set of all possible \mathbf{x} . Mathematically, pattern recognition, defined as, $\phi: \mathbf{X} \rightarrow \mathbf{C}$ is learning the mapping between the representation space \mathbf{X} and the classification space \mathbf{C} , where $\mathbf{C} = (c_1, c_2, \dots, c_M)$ is a set of all possible classes/clusters. A process malfunction is visible in the measurements of the plant’s sensors, which show abnormal readings. These sensors could be hardware or software (known as “soft-sensor”) based. In the pattern recognition based FDD, the sets of plant sensor measurements are

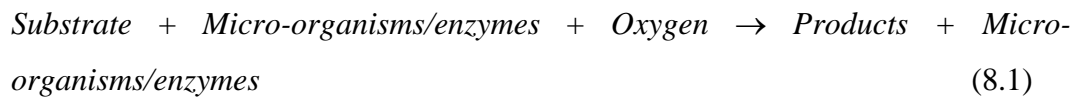
associated with the corresponding process faults. The task of FDD then involves classifying the sets of symptoms (sensor readings) into various categories of faults and the “normal” process behavior. One major benefit of such methods is that a phenomenological process model is not required for the associated FDD. Various clustering methods (conventional non-CI and CI-based) are used for process FDD by pattern recognition (see Section (1.5) and (1.6)). Conventionally, the *c*-means (also known as k-means) method is used for clustering in FDD applications. In the present study, the CI-based *fuzzy c-means clustering* (FCC) method is used for FDD of a biochemical process. For affording a comparison, the FDD problem has also been addressed using the conventional *c*-means clustering method.

Fuzzy c-means clustering (FCC) (Bezdek et al., 1984) is a soft clustering methodology that uses unsupervised learning for classifying the data-points to one of the *c*-clusters. FCC tries to determine soft (i.e., fuzzy) *c*-partitions of a data-set **X**. The idea behind the working of fuzzy clustering was laid by Zadeh (1965), which is to represent the similarity a point shares with each cluster with a value ranging from 0 to 1 (known as membership value). Further details of the FCC methodology are given in Section (1.6.1).

The *c*-means is a popular conventional centroid-based hard clustering algorithm that clusters a given data-set, **X** into the pre-specified *c*-clusters. It does this by iteratively minimizing an objective function that includes the Euclidean distance of each data-point from the *c*-centroids, which are initialized randomly (MacQueen, 1967). The disadvantages of the *c*-means algorithm are that it does not necessarily find the most optimal configuration, corresponding to the global objective function minimum and it is highly sensitive to the initial randomly generated cluster centroid locations. For more details of the *c*-means clustering methodology the reader can refer the section Section (1.5.1).

8.0.2. Fault Detection and Diagnosis of Biochemical Process

The biochemical process considered here involves microorganisms or enzymes secreted by microorganisms that act as catalysts in the transformation of the high molecular weight organic raw materials (substrate) to low molecular weight organic products. Such a reaction, is written as,



Biochemical processes possess an important inherent advantage of mild operating conditions. A variety of important organic chemicals are now manufactured through a biochemical route rather than the traditional chemical route due to the stated reason, thus saving large energy costs required in typical unit processes (synthesis reactions). These mainly include pharmaceuticals, food, and beverage processes (Bequette, 1998; Levenspiel, 1999).

Biochemical processes are sensitive to the changes in the operating conditions; even small changes in process conditions may result in a process malfunction or a complete failure. The major reason for this behavior is that the activity of most microorganisms/enzymes is highly temperature dependent and cannot sustain significant deviations in the reaction temperature. Thus, FDD studies of biochemical processes are important for avoiding batch/process malfunctions and complete failures. Table (8.1) lists a few recent FDD studies of bioprocesses using different methods.

Table 8.1: Existing fault detection and diagnosis studies pertaining to bioprocesses

Sr. No.	Based on	Method(s) used	Objectives	Reference
1.	Model-based	ANN	FDD of a biotechnological process in waste water treatment	Fuente and Vega (1999)
		Fuzzy logic	FDD using fuzzy residual based module for a pilot-scale anaerobic digestion reactor	Genovesi et al. (1999)
		High gain nonlinear observers	High-gain nonlinear observer for FDD of a bioreactor	Martínez-Guerra et al. (2001)
		Adaptive observers	Fault detection and isolation in an alcoholic fermentation process	Kabbaj et al. (2001)
		Adaptive observer based process model	FDD of ethanol production bioprocess with a two-stage bioreactor	Aceves-Lara et al. (2012)
		Particle Swarm Optimization (PSO), DE, Covariance Matrix Adaptation Evolution Strategy (CMA-ES)	FDD of a nonlinear bioreactor	Díaz et al. (2016)
2.	History-based	Expert systems	Use the enlarged state concept in FDD of bioreactors	Halme (1989)
		PCA and ANN	FDD of penicillin production and protein synthesis processes	Kulkarni et al. (2004)
		PCA	FDD of a Industrial Fed-Batch Cell Culture Process	Gunther et al. (2007)

		PCA, auto-regressive PVA and multi-way PCA	FDD of Biochemical batch process of penicillin fermentation.	Van den Kerkhof et al. (2012)
		PCA, Partial least squares (PLS) and Principal Component Regression (PCR)	FDD of batch and fed-batch bioreactor system for production of carotenoids via fermentation.	Karim et al. (2015)
		Random forest (RF), ANN and SVR	FDD of bioreactor for penicillin production process	Shrivastava et al. (2017)

The commonly used kinetic models for simple biochemical reactions are Michaelis-Menten (1913) and its simplified version by Monod (Monod, 1949). Both these are nonlinear and describe saturation growth kinetics of the biomass with respect to substrate concentration. In addition to the saturation-growth, phenomena such as competitive/non-competitive inhibition by a foreign substance, substrate inhibition and/or product inhibition also occur in biochemical reactions. These features often make the reactions complex in character and, thus, difficult to understand and model phenomenologically. Consequently, process data based FDD needs to be resorted to.

8.1. BIOCHEMICAL PROCESS FOR BIOMASS GENERATION

To generate process data comprising faulty and normal process operations, the phenomenological model of the bioprocess was simulated along with the control laws. It may be noted that in the absence of real process data, the process behavior has been simulated using a phenomenological model. In actual practice, data compiled from a real physical process operation should be used in FDD.

The continuous stirred bioreactor operating under the substrate-inhibition kinetic mechanism, considered in this FDD study has been analyzed by Bequette

(1998). Figure (8.1) shows the block diagram of this bioprocess. As indicated in the figure, the biomass concentration (X) is the main variable of the process controlled by a proportional-integral-derivative (PID) controller; the corresponding manipulating variable is the feed flow rate (F_0). The bioreactor temperature (T) is controlled by a proportional (P) controller (due to mild heat effects in the bio-reaction) by manipulating the jacket coolant flow rate (F_{J0}) (Coulson and Richardson, 1999).

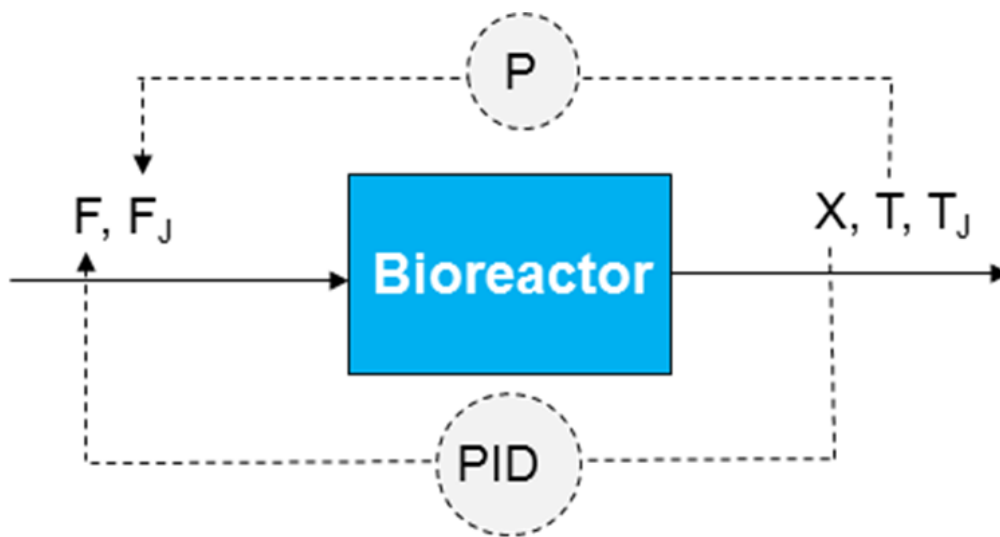


Figure 8.1: Block diagram of the continuous bioreactor generating biomass with controllers.

The phenomenological process model based on the substrate-inhibition kinetics of the bio-reaction is given by the following dynamical equations (Bequette, 1998),

$$\frac{dX}{dt} = \left(\mu - \frac{F}{V} \right) X \quad (8.2)$$

$$\frac{dS}{dt} = \left(\frac{F}{V} \right) (S_F - S) - \frac{\mu X}{Y} \quad (8.3)$$

$$\frac{dT}{dt} = \left(\frac{FT_0}{V} \right) - \left(\frac{FT}{V} \right) + \left(\frac{\mu X}{Y_\delta \rho C_p} \right) - \left(\frac{UA(T - T_J)}{V \rho C_p} \right) \quad (8.4)$$

$$\frac{dT_J}{dt} = \left(\frac{F_{J0}T_{J0}}{V_J} \right) - \left(\frac{F_{J0}T_J}{V_J} \right) + \left(\frac{UA(T - T_J)}{V_J \rho_J C_{pJ}} \right) \quad (8.5)$$

$$\mu = \left(\frac{\mu_{\max} S \alpha^{(T-25)}}{k_m + S + k_1 S^2} \right) \quad (8.6)$$

Table (8.2) describes the variables used in the above equations (Eqns. 8.2-8.6) along with their steady-state values used in the simulation.

Table 8.2: Process parameters and variables of the phenomenological bioreactor model

Sr. no.	Variable	Description	Steady-state value	Unit
1	A	Heat exchange area	0.1	cm^2
2	C_p	Heat capacity of reactor liquid	4.18	$\text{cal/g } ^\circ\text{C}$
3	C_{pJ}	Heat capacity of jacket liquid	4.18	$\text{cal/g } ^\circ\text{C}$
4	F_0	Volumetric feedrate	0.3	L/hr
5	F_{J0}	Jacket-side coolant flowrate	0.07177	L/hr
6	k_I	Inhibition kinetic constant 1	0.4545	
7	k_m	Inhibition kinetic constant 2	0.12	
8	S	Substrate concentration in reactor	0.175	g/L
9	S_F	Substrate concentration in feed	4	g/L
10	T	Temperature of reactor mass	24.95	$^\circ\text{C}$
11	T_0	Feed temperature	25	$^\circ\text{C}$
12	T_J	Temperature of jacket-side fluid	23	$^\circ\text{C}$
13	T_{J0}	Jacket-side feed temperature	23	$^\circ\text{C}$
14	U	Heat transfer coefficient	350	$\text{cal/cm}^2\text{hr } ^\circ\text{C}$
15	V	Reactor liquid volume	1	L
16	V_J	Jacket liquid volume	0.25	L
17	X	Biomass concentration in	1.530163	g/L

Sr. no.	Variable	Description	Steady-state value	Unit
		reactor		
18	Y	Substrate yield factor	0.4	g/g
19	Y_{δ}	Thermal yield factor	10000	g/cal
20	α	Temperature Coefficient	1.04	
21	ρ	Density of reaction mass	1000	g/L
22	ρ_J	Density of jacket fluid	1000	g/L
23	μ_{max}	Specific growth rate (maximum) of biomass	0.53	hr ⁻¹

8.1.1. Simulation of the Phenomenological Model

Kumar et al. (2008) have developed a robust PID controller scheme using genetic algorithm (GA) for the optimal feedback control of biomass generation process. The PID parameters proposed in the said study were used in the simulations for generating process data for FDD. Here, we have considered four representative single faults that can occur during the steady-state operation of the bioreactor when process parameters deviate from their normal (set) values. The specific faults which can majorly influence the steady-state bioreactor behavior are: (i) input flow rate (F_0) high, (ii) input flow rate low, (iii) input flow rate of coolant to the jacket (F_{J0}) high, and (iv) input flow rate of coolant to the jacket low. It is assumed that these faults occur in a mutually exclusive manner; that is, only one of the four types of faults can occur at any given time. Any high/low deviation in the two process parameters (F_0 , F_{J0}) is considered to eventually affect the steady-state values of the bioreactor state variables (X , T and T_J); the magnitudes of these state variables are commonly measured during a typical process operation.

For development of the FDD models of the bioreactor, a training set consisting of 601 data-points was generated by varying F and F_{J0} separately in the [0.0 to $\pm 15\%$] range from their nominal set values. The steady-state data were generated by successively incrementing (decrementing) the parameter value by 0.1% from its nominal value, thus generating 150 data-points for each of the four faults. The data containing the parameter and steady state values corresponding to the normal bioprocess operation were added to this set containing 600 data points

making the training data set of size of 601 patterns. To test the generalization capabilities of the CI and non-CI based clustering methods a test set was generated by the same method by successively incrementing (decrementing) the parameter values (F_0 , F_{j0}) by 0.01% in the same [0.0 to $\pm 15\%$] range. The data structure of the resultant sets is given in Table (8.3). The datasets consists of 601 (training set) and 6001 (test set) patterns representing four single faults and the normal process behavior. Conventionally, the deviations in process variables/parameters within $\pm 5\%$ threshold are considered to be normal and are not treated as faults (Venkatasubramanian et al., 1990). Therefore, deviations up to 5% are assumed to be the cases of normal operation. Table (8.4) list the corresponding faulty/normal operation values of 5 and 15% deviations. Figure (8.2) shows a two-dimensional projection of the training dataset. The plots show the variation of the steady-state values of bioreactor output variables, namely X , T and T_j with respect to the deviations (faults) imparted in the F_0 and F_{j0} input variables.

Table 8.3: Data structure for the training and test sets

Sr. No.	Nature of Fault	Fault code	Training-Data		Test-Data	
			Pattern no(s)	Deviation(%)#	Pattern no(s)	Deviation(%)#
1.	Input flow rate low	Fault 1	1-150	(-) 0.1-15.0	1-1500	(-) 0.1-15.0
2.	Input flow rate high	Fault 2	151-300	(+) 0.1-15.0	1501-3000	(+) 0.1-15.0
3.	Coolant flow rate low	Fault 3	301-450	(-) 0.1-15.0	3001-4500	(-) 0.1-15.0
4.	Coolant flow rate high	Fault 4	451-600	(+) 0.1-15.0	4501-6000	(+) 0.1-15.0
5.	Normal operation	Normal	601		6001	

Training and Test data were generated at the deviation intervals of 0.1% and 0.01%, respectively.

Table 8.4: Magnitudes of state variables in faulty and normal conditions

Fault code	Fault Type	Fault magnitude, %	Steady-state values of process variables		
			X (g/L)	T (°C)	T_J (°C)
Fault 1	F_{0-}	5	1.539	24.949	23.205
		15	1.553	24.943	23.212
Fault 2	F_{0+}	5	1.5182	24.953	23.2
		15	1.4788	24.957	23.195
Fault 3	F_{J0-}	5	1.53	24.952	23.213
		15	1.53	24.952	23.236
Fault 4	F_{J0+}	5	1.53	24.951	23.195
		15	1.53	24.951	23.18
Normal		0	1.5302	24.9511	23.2024

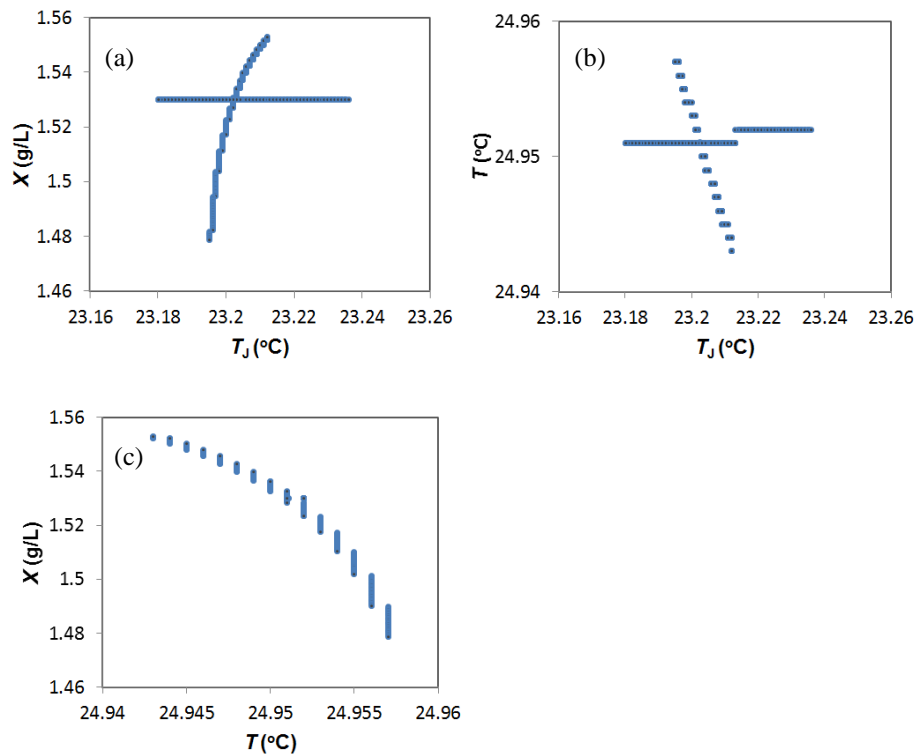


Figure 8.2: Nonlinear projection of 3-dimensional steady-state bioreactor data in 2-dimensions: (a) Biomass concentration (X) versus Jacket temperature (T_J), (b) Bioreactor temperature (T) versus Jacket temperature (T_J), (c) Biomass concentration (X) versus Bioreactor temperature (T).

8.2. RESULTS AND DISCUSSION

8.2.1. c-means Based FDD Modeling of the Biochemical Process

The data consisting values of 601 training set data-points pertaining to faulty and normal process behavior were subjected to clustering using the c-means clustering feature offered by *Rapidminer* software package (Mierswa et al., 2006). Parameters and other settings of the algorithm used for developing the bioprocess FDD-based model are given in Table (8.5), while its results are presented in Table (8.7). As can be seen, the c-means method has clustered the training (93.67% classification accuracy) and the test (93.66% classification accuracy) sets with good accuracies.

8.2.2. FCC Based FDD Modeling of the Biochemical Process

The bioreactor datasets used in the c-means clustering exercise were also used to develop and assess the generalization capability of the FCC-based model using the FCC module available in the *Fuzzy Logic* toolbox of MATLAB software package (MATLAB, 2000). Table (8.6) lists the parameters used for the FCC-based clustering of the bioprocess data-set (see Eqn. 1.33 in Section (1.6.1) for details).

Table 8.5: Parameters and settings used for the c-means clustering of the bioprocess data-set.

Dataset	Data-points		c*	Divergence method	Distance measure
	Training-set	Test-set			
Bioprocess	601	6001	5	Bregman divergence	Squared Euclidean Distance

* c (no. of clusters)

Table 8.6: Parameters used for the FCC-based clustering of the bioprocess data-set.

Dataset	Data-points		c*	m*
	Training-set	Test-set		
Bioprocess	601	6001	5	2

* c (no. of clusters); m (weighing exponent)

The true classification accuracy magnitude is calculated using the expression given below.

$$\text{True classification} = \frac{\text{Datapoints correctly classified into respective clusters}}{\text{Total datapoints}} \times 100$$

(8.7)

Table 8.7: Comparative results of c-means and FCC-based clustering of the bioprocess data-set.

Method	True classification (%)	
	Training-set	Test-set
c-means	93.67	93.66
FCC	95.33	95.43

The FCC-based clustering exercise for fault detection and diagnosis of the bioprocess generated a matrix of clustered data-points containing five cluster centers and U-matrix. The U-matrix contains membership values (in the range of 0 to 1) of each of the 601 (6001) data-points. The U-matrix pertaining to the training and test datasets are plotted in Figure (8.3), wherein it can be observed that in each subsection there is a dominant membership curve. In all the four sections approximately first fifty points correspond to the “normal” process behavior (indicated by blue curves); the said points have higher membership values than remaining points. After the first fifty points, the curve representing the faulty process behavior dominates. Specifically, green, red, light blue and violet colored curves dominate after around fifty points in sections (i), (ii), (iii), and (iv) respectively. These curves represent data pertaining to faults 1, 2, 3 and 4, respectively. Note that the individual lines in U-matrix plot represent a cluster. After clustering the training data, predictions for the test data were made. Figure (8.3(b)) shows the U-matrix pertaining to the test data.

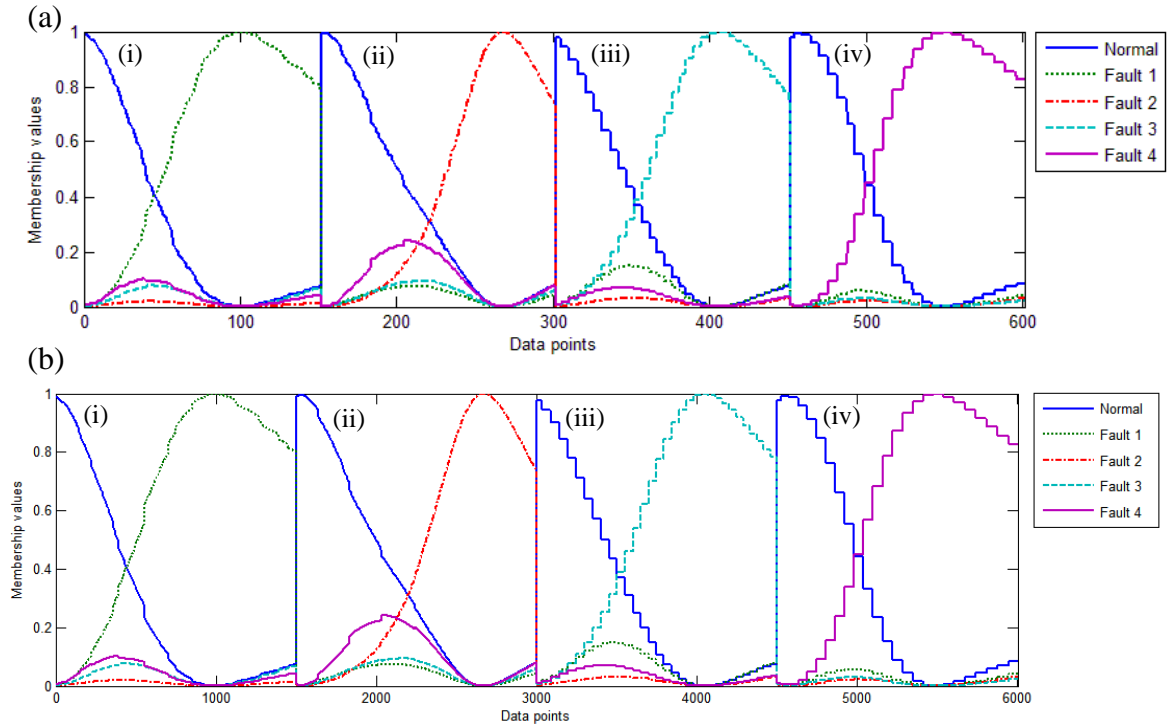


Figure 8.3: Plots of membership values of the U-matrix of the developed FCC-based bioreactor FDD model: (a) Training-set data, (b) Test-set data.

The classification accuracy corresponding to the test data has been found to be 95.43 percent (see Table (8.7)). It can thus be seen that the FCC-based model detected the faulty (non-normal) or normal bioreactor operation as also identified (diagnosed) the type of fault based on the cluster to which the data-point belongs to. From the results presented in the table it is also clear that faulty/normal behavior classification and generalization accuracies of the FCC methodology (95.33% and 95.43%, respectively) are better than that (93.67% and 93.66%) of the conventional c-means method. This indicates that the FCC-model has outperformed the conventional c-means clustering model in detecting and diagnosing four faults of the biochemical process.

8.3. CONCLUSION

Biochemical processes are typically run under mild operating conditions; generally these are also highly sensitive to the changes in their operating conditions. Thus, small offsets in the operating conditions/parameters caused by malfunctioning equipment/sensor/actuator may drive the bioprocess towards a

partial/complete failure. An early and accurate detection and diagnosis of such faults is necessary to avoid process failures. Accordingly, in this study, a CI-based methodology, namely *fuzzy c-means clustering* (FCC) has been successfully utilized to develop fault detection and diagnosis model for a continuous biochemical reactor operating under a substrate-inhibition kinetic mechanism. FCC is a clustering method based on the principles of fuzzy logic and clusters the data-points according to membership values assigned to each data-point in the dataset. To afford a comparison, the stated FDD task was also performed using standard c-means clustering algorithm. The FDD potential of both the methods was assessed using a simulated dataset comprising of faulty and normal operation of the bioreactor substrate-inhibition kinetic mechanism. The developed FCC-based model could successfully identify four types of faulty regions (clusters) and the normal operating region (cluster) with 95.33% prediction accuracy and 95.43% generalization accuracy. The corresponding figures for the c-means clustering method are 93.67% and 93.66%. It can thus be seen that the FCC method showed a improvement of around 2% in the prediction and 2% in the generalization performance over c-means method while performing FDD of the bioreactor. The results of this study suggest that the FCC is an attractive strategy for fault detection and diagnosis of nonlinear processes such as bioreactors and can be easily extended to similar type of complex processes.

NOMENCLATURE

A	Heat exchange area
c	Class/Cluster
C	Classification space
C_p	Heat capacity of reactor liquid
C_{pJ}	Heat capacity of jacket liquid
F_0	Volumetric feedrate
F_{J0}	Jacket-side coolant flowrate

k_I	Inhibition kinetic constant 1
k_m	Inhibition kinetic constant 2
m	Weighing exponent in FCC
S	Substrate concentration in reactor
S_F	Substrate concentration in feed
T	Temperature of reactor mass
T_0	Feed temperature
T_J	Temperature of jacket-side fluid
T_{J0}	Jacket-side feed temperature
U	Heat transfer coefficient
V	Reactor liquid volume
V_J	Jacket liquid volume
x	Process variable
\mathbf{x}	Vector of process variable values
\mathbf{X}	Process variable representation space
X	Biomass concentration in reactor
Y	Substrate yield factor
Y_δ	Thermal yield factor

Greek letters

α	Temperature Coefficient
ρ	Density of reaction mass
ρ_J	Density of jacket fluid
μ_{max}	Specific growth rate (maximum) of biomass

REFERENCES

- Aceves-Lara, C. A., Fragkoulis, D., Roux, G., Dahhou, B. (2012). Single and multiple faults in system actuators and sensors for ethanol production. In: IFAC Proceedings vol. 45(20):228-233.
- Bequette, W. B. (1998). Process Dynamics Modeling, Analysis, and Simulation. Prentice Hall, Upper Saddle River, New Jersey, USA.
- Caccavale, F., Pierri, F. (2009). Fault Diagnosis for a Class of Chemical Batch Processes. In: American Control Conference, Hyatt Regency Riverfront, St. Louis, MO, USA.
- Coulson, Richardson. (1999). Chemical & Biochemical Reactors & Process Control, Chemical Engineering vol. 3, eds. Richardson, J. F., Peacock, D. G., 3rd ed., Elsevier Butterworth-Heinemann, MA.
- Dash, S., Venkatasubramanian, V. (2000). Challenges in the industrial applications of fault diagnostic systems. Computers and Chemical Engineering 24:785-791.
- Díaz, C. A., Echevarría, L. C., Prieto-Moreno, A., Neto, A. J. S., Llanes-Santiago, O. (2016). A model-based fault diagnosis in a nonlinear bioreactor using an inverse problem approach and evolutionary algorithms. Chemical Engineering Research and Design 114:18-29.
- Fuente, M. J., Vega, P. (1999). Neural networks applied to fault detection of a biotechnological process. Engineering applications of artificial intelligence 12(5):569-584.
- Genovesi, A., Harmand, J., Steyer, J. P. (1999). A fuzzy logic based diagnosis system for the on-line supervision of an anaerobic digester pilot-plant. Biochemical engineering journal 3(3):171-183.
- Gunther, J. C., Conner, J. S., Seborg, D. E. (2007). Fault Detection and Diagnosis in an Industrial Fed-Batch Cell Culture Process. Biotechnology progress 23(4):851-857.

- Halme, A. (1989). Expert System Approach to Recognize the State of Fermentation and to Diagnose Faults in Bioreactors. In: Computer applications in fermentation technology: modelling and control of biotechnological processes, Springer Netherlands. pp. 159-168.
- Himmelblau, D. M. (1978). Fault detection and diagnosis in chemical and petrochemical processes. vol. 8, Elsevier Science Ltd.
- Kabbaj, N., Polit, M., Dahhou, B., Roux, G. (2001). Adaptive observers based fault detection and isolation for an alcoholic fermentation process. In: 8th IEEE International Conference Emerging Technologies and Factory Automation, vol. 2, pp. 669-673.
- Karim, M. N., Botre, C., Gomez, M. S. A., Nounou, M., Nounou, H., Mansouri, M. (2015). Fault Detection and Diagnosis in Batch and Fed-Batch Bioreactor System Using PCA, PCR, PLS with GLR Method. In: Proceedings of 2015 AIChE Annual Meeting on Systems and Process Control.
- Kulkarni, S. G., Chaudhary, A. K., Nandi, S., Tambe, S. S., Kulkarni, B. D. (2004). Modeling and monitoring of batch processes using principal component analysis (PCA) assisted generalized regression neural networks (GRNN). *Biochemical Engineering Journal* 18(3):193-210.
- Kumar, S. M. G., Jain, R., Anantharaman, N., Dharmalingam, V., Begum, K. M. M. S. (2008) Genetic Algorithm Based PID Controller Tuning for a Model Bioreactor, *Indian Chemical Engineer* 50:214–226.
- Levenspiel, O. (1999). *Chemical Reaction Engineering*. Third Edition, John Wiley & Sons, New York.
- Martínez-Guerra, R., Garridot, R., Osorio-Mirónt, A. (2001). High-gain nonlinear observer for the fault detection problem: Application to a bioreactor. *IFAC Proceedings Volumes* 34(6):1567-1572.
- Michaelis, L., Menten, M. L. (1913). *Biochemische Zeitschrift* 49:333.

- Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., Euler, T. (2006). YALE: rapid prototyping for complex data mining tasks. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-06).
- Monod, J.(1949). Ann. Reu. Microbiology, 3, 371.
- Pau, L. F.(1981). Failure diagnosis and performance monitoring. Marcel Dekker, New York.
- Shrivastava, R., Mahalingam, H., Dutta, N. N. (2017). Application and Evaluation of Random Forest Classifier Technique for Fault Detection in Bioreactor Operation. Chemical Engineering Communications 204(5):591-598.
- Sobhani-Tehrani, E., Khorasani, K. (2009). Fault diagnosis of nonlinear systems using a hybrid approach. Springer, Berlin, Germany.
- Tambe, S. S., Kulkarni, B. D., Deshpande, P. B. (1996). Elements of artificial neural networks with selected applications in chemical engineering, and chemical & biological sciences. Simulation & Advanced Controls, Louisville KY.
- Van den Kerkhof, P., Gins, G., Vanlaer, J., Van Impe, J. F. (2012). Dynamic model-based fault diagnosis for (bio) chemical batch processes.Computers & Chemical Engineering 40:12-21.

CHAPTER 9.

THESIS CONCLUSION

THESIS CONCLUSION

9.0 INTRODUCTION

This chapter concludes the thesis with the presentation of principal results and conclusions reached thereof in the studies described in chapters 2 to 8. Finally, future scope of the research related to this thesis is discussed.

Mathematical models are essential for numerous process-engineering related tasks such as monitoring, control, equipment design, operation, scale-up, optimization and, fault detection and diagnosis. The *phenomenological (first principles/mechanistic)* modeling approach is a traditional approach that is based on the complete knowledge of the physicochemical phenomena underlying the process. In this approach, the basic mass, energy and momentum balance equations governing the process are considered for modeling the process under study. Finally, such phenomenological models involve numerous parameters that are determined from an extensive experimentation. However, the modern day chemical processes invariably involve complex phenomena and a multitude of nonlinear interactions among the process variables, whose mechanisms are hard to determine. Thus, the phenomenological modeling of modern chemical processes is tedious, time-consuming and costly.

The *empirical* modeling is an attractive alternative to the phenomenological modeling approach, wherein a pre-specified linear/nonlinear function (model) is fitted to the reaction/reactor/process data. Thus, in empirical modeling a detailed knowledge of the process is not required. However, it has its own limitations such as the requirement of an appropriately pre-specified linear/nonlinear function with known form and estimation of the function parameters. Generally, this is a demanding task since in many chemical processes multiple variables interact nonlinearly thus requiring a tremendous trial and error effort in pre-specifying the data-fitting model and estimating its parameters.

In chemical engineering/technology optimization of process parameters is critical for maximization of conversion/yield, minimization of operating cost, etc. Conventionally, *deterministic gradient-based* methods are used in process

optimization. The basic requirement for these methods to work is that the objective function (to be maximized/minimized) must be continuous, differentiable and smooth. Among these, the requirement of objective function to be smooth is difficult to meet for exclusively data-driven reaction/process models developed using for instance, genetic programming, ANN and SVR methods. The final solutions determined by the gradient-based optimization methods are also highly sensitive to the initial guess specified for the decision variables appearing in the objective function. Often, improper initial guesses drive the solution trajectories to get stuck in a local optimum leading to sub-optimal solutions.

Computational intelligence (CI) offers numerous state-of-the-art techniques for data-driven modeling and optimization of complex nonlinear processes/reactions/reactors. Notably, they overcome the above-stated limitations of phenomenological/empirical modeling and the deterministic optimization approaches. Accordingly, in this thesis computational intelligence (CI) based modeling methodologies, namely, *artificial neural networks* (ANN), *genetic programming* (GP), *fuzzy c-means clustering* (FCC), and *support vector regression* (SVR) have been employed for tasks such as property prediction, process identification and control, process/reaction modeling, process/reaction optimization, and process fault detection and diagnosis. The modeling studies conducted and presented in this thesis belong to the diverse fields, namely, fuel science and engineering, environmental engineering, process dynamics and control, and process safety.

Process models are required in process optimization. Accordingly, for optimization of a resin based adsorptive pollutant removal the corresponding ANN model has been optimized by two stochastic CI-based optimization methodologies; these are *artificial immune system* (AIS) based *clonal selection algorithm* (CLONALG) and *genetic algorithms* (GA).

Additionally, wherever the input space of the model was large, a dimensionality reducing linear algebra technique termed *principal component analysis* (PCA) has been utilized. Chapter 1 begins with a description of the need for process modeling and optimization as important process engineering tasks.

Next, it gives a detailed review of computational intelligence and its methodologies that have been used/have the potential of use for these process engineering tasks. Specifically, the CI-based methodologies applied to various chemical engineering and technology applications in the current study are elaborated in much detail along with their mathematical background, applications, advantages, disadvantages and software tools available for their implementation. The following section provides the rationale, salient features, and highlights of the research work reported in chapters 2 to 8.

9.1 OVERALL CONCLUSION

Chapter 2 introduces the development of models predicting the higher heating value (HHV) of solid biomass fuels using the CI-based methodologies namely *multi-layer perceptron neural network* (MLPNN) and *genetic programming based symbolic regression* (GPSR). The important/novel features of this study are as follows.

- The stated CI-based modeling has been introduced for developing nonlinear models encompassing a large size proximate and ultimate analyses dataset consisting of 382 and 536 biomasses respectively, used in developing the models.
- When compared with their existing counterparts, the four optimal CI-based biomass HHV models exhibit excellent prediction and generalization capabilities.
- The results of this study indicate the high potential of CI-based nonlinear modeling approach for property modeling in fuel science and engineering.

The knowledge of elemental composition of a biomass fuel is important in the efficient design and operation of combustion and gasification systems using these fuels, and pollution control thereof. The ultimate analysis of biomass fuels is a finer analysis that includes this elemental composition expressed in terms of weight percentages of major elements, namely *carbon (C)*, *hydrogen (H)*, *oxygen*

(O) and some minor elements. The experimental determination of the elemental composition of a biomass fuel is time-consuming and expensive. The proximate analysis, which determines fixed carbon, ash, volatile matter and moisture content of a biomass fuel, is a cruder albeit easier to perform analysis. Accordingly, the objective of Chapter 3 was to present CI-based modeling approach for determination of the best correlations existing between the elemental composition of solid biomass fuels and their proximate analysis. The salient features of this study are as follows.

- A literature survey indicated that only two linear models are available for the prediction of the said elemental composition although the actual relationships are nonlinear.
- The present study utilizes three computational intelligence (CI) methods, namely, GPSR, MLPNN, and SVR, for developing nonlinear models for the prediction of *C*, *H*, and *O* percentages of solid biomass fuels.
- All the nonlinear CI-based correlations predicting the elemental compositions of biomass significantly and consistently outperformed their linear counterparts in terms of higher prediction accuracies and generalization capabilities.

Chapter 4 introduces *genetic programming based symbolic regression* (GPSR) for the development of high performing models for predicting the higher heating value of coals of different ranks, and from diverse geographies of the world. The principal features of this study are as follows.

- An extensive literature survey indicated most of the previously developed coal HHV models are linear regression based models. Since a decade it was determined that the HHV of coal nonlinearly depends on few constituents of their proximate and/or ultimate analysis and thus nonlinear modeling of coal's HHV was emphasized.
- Accordingly, CI-based nonlinear models were developed recently, based on the ANNs, *co-active neuro-fuzzy adaptive networks*

(CANFIS), *alternating conditional expectation* (ACE), and SVR formalisms. However, these modeling attempts were aimed at coals of a specific rank and/or from specific geographical regions based on limited amounts of data.

- Accordingly, in this study five GPSR-based nonlinear, high performing models are developed for the prediction of coal HHV. The said models are developed using a huge-sized dataset comprising coals of various ranks and from diverse geographies of the world and thus the models are highly generalized.
- When compared with their existing counterparts statistically, the five optimal GPSR-based coal HHV models exhibit excellent prediction and generalization capabilities.
- The advantage of using GPSR-based models for this task is that due to their lower complexity, these models are easier to understand and deploy in a practical setting as compared to the previously developed CI-based coal HHV models.

In Chapter 5, *genetic programming based symbolic regression* (GPSR)-based modeling approach has been suggested for *process identification* of a polymerization reactor in the form of *soft sensor* models. The notable features of this study are as follows.

- The styrene polymerization process involves a sequence of reactions with complex mechanisms often dominated by heat, mass and momentum transfer effects, thus creating modeling difficulties by the traditional phenomenological approach.
- In view of these difficulties, two GPSR-based soft-sensor models were introduced for dynamic modeling of the styrene polymerization reactor. The developed process models dynamically relate the polymer molecular weight (output) to the jacket coolant flowrate (input).

- Both the GPSR-based models predicting the effluent polystyrene molecular weight exhibit excellent prediction and generalization performances.
- The best GPSR-based model is successfully implemented in a *model predictive control* (MPC) scheme to control the effluent polymer quality in terms of its molecular weight. When compared with the traditional proportional-integral-derivative (PID) controller the MPC scheme using the GPSR-based predictive model exhibited superior control performance in both set-point tracking and disturbance rejection cases.
- The GPSR-based models owing to their simplicity impart much lesser computational load during the software-based MPC implementation. The proposed approach can be extended for process identification and control of industrial processes exhibiting similar difficulties.

Co-gasification (COG) of coal with biomass is an important step towards clean coal technology as it operates with reduced pollution and possesses operational flexibility. Accordingly, in Chapter 6 for the performance analysis of a pilot plant fluidized bed co-gasifier (FBCOG), CI-based modeling of the process is proposed. The important/novel features of this study are as follows.

- The FBCOG process comprises of multiple synergetic reactions, which poses phenomenological modeling difficulties due to the existence of complex nonlinear phenomena created by heat, mass and momentum transfer as also due to the multiple reaction effects.
- Extensive FBCOG experiments using various blends of high ash Indian coal and biomasses (rice husk, press-mud and sawdust) were conducted and the resulting dataset was utilized in the CI-based modeling of the process.
- Specifically, the three CI-based modeling methodologies, namely GPSR, MLPNN and SVR were used to develop FBCOG process models predicting the four significant co-gasification performance parameters, viz. *total gas yield* (TGY) (kg/kg fuel), *carbon conversion*

efficiency (CCE) (%), *heating value of product gas (HV) (MJ/Nm³)*, and *cold gas efficiency (CGE) (%)*. The models inputs were and as inputs, predict

- The original experimental dataset consisting of thirteen inputs characterizing the feed characteristic and the process operating conditions was compressed to obtain four linearly uncorrelated inputs using *principal component analysis (PCA)*. This compressed dataset was used for developing the CI-based models.
- All the developed CI-based FBCOG process models exhibit good prediction accuracy and generalization performances. Finally, statistical Steiger's tests were conducted for comparing the developed CI-based models to identify the best ones predicting the four COG performance variables.
- The stated approach can be used to model industrial scale fluidized bed co-gasifiers which can be gainfully employed in the optimal design and operation of such processes.

In Chapter 7, the environmental problem of chromium contamination in ground water is addressed by utilizing a CI-based hybrid strategy. The salient features of this study are as follows.

- Different grades of the gallic acid-formaldehyde-ammonia (GFA) polymer resin were synthesized and batch equilibrium adsorption experiments were conducted for evaluating adsorption capacity of chromium Cr(VI) ions on the resins at different pH.
- Highly nonlinear relations were witnessed between the percent adsorption of Cr(VI) and adsorption pH while using different grades—dictated by the synthesis reaction conditions—of the GFA resin for adsorption.
- Accordingly, the CI-based nonlinear methodology, *multilayer perceptron neural network (MLPNN)* is used to develop model predicting the percentage adsorption of Cr(VI) ions using the synthesis

reaction conditions of the polymer resin and pH of the adsorption process as inputs.

- The developed nonlinear MLPNN-based model exhibit excellent prediction and generalization performance. For optimization of reaction conditions and the adsorption pH, CI-based optimization formalisms were used in a hybrid setting. Specifically, the CI-based stochastic optimizers based on the novel *artificial immune system* (AIS), namely, *clonal selection algorithm* (CLONALG) and *genetic algorithm* (GA) were used for the said optimization.
- The CI-based hybrid approach—especially using the AIS-based optimizer—led to a 3% improvement in the resin based adsorptive removal of Cr(VI) ions over that observed in experiments before performing the reaction optimization.
- The proposed CI-based hybrid strategy can be extended for mitigating similar environmental problems.

Chapter 8 reports the use of CI-based *fuzzy c-means clustering* (FCC) algorithm for fault detection and diagnosis of a biochemical process. The FCC method is an unsupervised learning algorithm that uses the process history data to generate a model that appropriately clusters the dataset into a pre-specified number of clusters. The notable features of this study are as follows.

- A continuous bioreactor generating biomass was simulated for four types of faults, namely increase/decrease in the bioreactor feed and jacket fluid flow rates. The resultant process history data was captured in terms of the bioreactor state variables, namely the bioreactor biomass concentration, bioreactor temperature and the jacket fluid temperature, which was used to develop the FCC-based model.
- The developed FCC-based model is capable of accurately identifying all four types of faults as also the normal process behaviour.

- Simultaneously the conventional c-means clustering algorithm was also applied to model the faulty and normal operations of the bioreactor.
- When compared with the c-means model, the FCC-based model exhibit better cluster prediction and generalization capability for the said fault detection and diagnosis application.
- The proposed CI-based methodology can be extended for fault detection and diagnosis of any full-scale bioreactor from its process history data.

9.2 FUTURE RESEARCH SCOPE

The major research in artificial/computational intelligence (AI/CI) began around the middle of the nineteenth century and gained momentum in 1990s with the tremendous advancement of the computing hardware and software technologies. Since then new efficient methodologies are being continuously developed and old ones are modified for improved performance. Today AI/CI based methodologies are being successfully applied in almost all branches of engineering, technology and science and soon AI/CI will touch all aspects of our lives.

While it is impossible to imagine futuristic applications of AI/CI in chemical engineering and technology, those which have a high probability of succeeding are given below.

- *Applications in modeling of huge/big datasets:* Today's highly automated complex mega-scale processes generate process history data at high speeds. CI-based methods have capabilities to extract important features/knowledge of interest from such huge datasets. This is specifically required in process monitoring and fault detection. In this thesis the newer GPSR-based methodology have been successfully utilized to extract useful expressions from large datasets, such as for property predictions and process identification. Recent CI-based approaches can be gainfully

utilized for such tasks. For example, the recently researched *evolutionary deep intelligence*, wherein *deep neural networks* are evolved to generate highly efficient and powerful solutions to problems involving huge data can be utilized for process monitoring, fault detection and diagnosis as also fluid flow modeling. Deep neural networks (also termed as *deep learning networks*) are “stacked” neural networks consisting of around thousands of processing nodes that are efficiently used for pattern (image, voice and motion) recognition in computer science/engineering. These can be utilized for image recognition/classification of flows (laminar/transitional/turbulent) in reactors.

- *Utilization of newer CI-based methodologies*: Newer CI-based methods may give improved results when applied for process engineering tasks. The major modeling work in this thesis is based on the utilization of GP for chemical engineering and technology applications; the reason being its rare use, in spite of its numerous advantages over other CI-based formalisms. The use of GP can be further extended to multiple input – multiple output (MIMO) systems, i.e., for process identification, monitoring, and fault detection and diagnosis. Apart from process engineering tasks the GP-based methods have a huge scope in chemical engineering for applications in property prediction such as quantitative structure-property/activity relationship (QSPR/ QSAR) studies, molecular modeling for novel drug synthesis, modeling of multi-component equilibrium systems, etc.
- Artificial Immune System (AIS) is a novel CI-based paradigm offering numerous modeling and optimization algorithms such as the clonal selection algorithm (CLONALG), negative selection algorithms (NSAs), and algorithms based on the danger theory and the immune network models for various process applications. Currently, it is being researched heavily for computer and internet security (cyber security) applications due to its several immune-based advantages. There is a vast scope of its utilization for fault detection and diagnosis of modern complex multivariable processes as also in process optimization.

- *CI-based hybrid approaches*: Use of CI-based hybrid approaches can give better performances as such an approach works on the strategy that the synergy between two methods can be harnessed usefully. In this thesis the AIS-based stochastic optimizer has been successfully used for reaction/process optimization as modeled by the highly nonlinear ANN-based model. Similar hybrid approaches can be applied to get improved results.

There exists a large number of modeling and optimization algorithms in CI that are based on the intelligent behavior seen in nature, such as ANNs, evolutionary algorithms, genetic programming, ant colony methods, particle swarm, bee algorithm, etc. There are many more intelligent processes in nature that are followed by trees, plants, animals (land and aquatic), birds, microorganisms, viruses. These can be simulated on computer to develop more efficient, fast and numerically light weight modeling, optimization, security and image/audio/video processing/recognition applications.

To summarize, the increasing popularity of CI-based formalisms is alluded to their potential in efficient modeling, pattern recognition and optimization of highly complex and/or data intensive nonlinear processes/systems. The future smart process industries for their efficient and smooth operation will require a proper integration of the advancing computing technologies such as the process centric *internet of things* (IoT), cloud computing and computational intelligence based approaches/strategies.

List of Publications

Publications Received from the Work Presented in the Thesis

1. Ghugare, S. B., Tiwary, S., Elangovan, V., Tambe, S. S. (2014). Prediction of Higher Heating Value of Solid Biomass Fuels Using Artificial Intelligence Formalisms. *Bioenergy Research* 7:681–692. DOI: 10.1007/s12155-013-9393-5
2. Ghugare, S. B., Tiwary, S., Tambe, S. S. (2014). Computational intelligence based models for prediction of elemental composition of solid biomass fuels from proximate analysis. *International Journal of System Assurance Engineering and Management* 1-14. DOI: 10.1007/s13198-014-0324-4
3. Ghugare, S. B., Tambe, S. S. (2017). Genetic Programming based High Performing Correlations for Prediction of Higher Heating Value of Coals of Different Ranks and from Diverse Geographies. *Journal of the Energy Institute*, 90(3):476-484. DOI: 10.1016/j.joei.2016.03.002
4. Ghugare, S. B., Tambe, S. S. (2016). Development of Genetic Programming based Softsensor Model for Styrene Polymerization Process and its Application in Model Based Control. *Indian Control Conference (ICC), IEEE Xplore Digital Library*, pp. 238-244.
5. Tiwary, S., Ghugare, S. B., Chavan, P. D., Saha, S., Datta, S., Sahu, G., Tambe, S. S. Co-gasification of high ash coal and biomass: Development of Computational Intelligence based Models. (Communicated).