



**Modelling and optimisation of post-combustion carbon capture
process integrated with coal-fired power plant using
computational intelligence techniques**

Fei Li

A Thesis submitted for the degree of Doctor of Philosophy

School of Chemical Engineering and Advanced Materials

Newcastle University

United Kingdom

January 2018

Abstract

Coal-fired power plants are the major source of CO₂ emission which contributes significantly to global climate change. An effective way to reduce CO₂ emission in coal-fired power plants is post-combustion carbon dioxide (CO₂) capture (PCC) with chemical absorption. The aim of this project is to carry out some research in model development, process analysis, controller design and process optimization for reliable, optimal design and control of coal-fired supercritical power plant integrated with post-combustion carbon capture plant. In this thesis, three different advanced neural network models are developed: bootstrap aggregated neural networks (BANNs) model, bootstrap aggregated extreme learning machine (BAELM) model and deep belief networks (DBN) model. The bootstrap aggregated model can offer more accurate predictions than a single neural network, as well as provide model prediction confidence bounds. However, both BANNs and BAELM have a shallow architecture, which is limited to represent complex, highly-varying relationship and easy to converge to local optima. To resolve the problem, the DBN model is proposed. The unsupervised training procedure is helpful to get the optimal solution of supervised training. The purpose of developing neural network models is to find a best model which can be used in the optimization of the CO₂ capture process precisely.

This thesis also presents a comparison of centralized and decentralized control structures for post-combustion CO₂ capture plant with chemical absorption. As for centralized configuration, a dynamic multivariate model predictive control (MPC) technique is used to control the post-combustion CO₂ capture plant attached to a coal-fired power plant. When consider the decentralized control structures based on multi-loop proportional-integral-derivative (PID) controllers, two different control schemes are designed using relative disturbance gain (RDG) analysis and dynamic relative gain array (DRGA) analysis, respectively. By comparing the two control structures, the MPC structure performs better in terms of closed-loop settling time, integral squared error, and disturbance injection.

Acknowledgement

I would like to express my very great appreciation to Dr. Jie Zhang, my research supervisor, for his, patient guidance, enthusiastic encouragement, valuable and constructive suggestions during the planning and development of this research work. I would also like to thank Dr. Jonathan Lee for his help in doing the HYSIS simulation. My grateful thanks are also extended to Prof. Meihong Wang for his help in perfecting my mechanistic knowledge frame and good understanding of the process, to Dr. Eni Oko, who generate data from gPROMS simulation software. Advice given by Prof. Dexian Huang, a professor in Department of Automation, Tsinghua University, China, has been a great help in developing my research work. I would like to offer my special thanks to Dr Chao Shang of Department of Automation, Tsinghua University, China, for his help in writing deep learning codes in MATLAB.

I would like to extend my thanks to the staffs and students of the School of Chemical Engineering and Advanced Materials for their help in offering me the resources in running the program. My special thanks are given to Mrs Justine McGruther, who gives a lot of advices and assistance in keeping my progress on schedule.

Many thanks to all friends who were staying with me through my 5 years life in UK. Because of you, I have the sense of belonging here and do not feel lonely. I would also like to acknowledge the emotional support from my best friends in China. Every time when I was falling down, you give me confidence and encourage me never give up. So many friends I cannot state all your names. Life is not easy, but it becomes better with you, my friends.

Last but not least, I wish to thank my parents, my sister and all my family members for their support and encouragement throughout my study.

Table of Contents

Abstract.....	i
Acknowledgement	iii
Chapter 1. Introduction	1
1.1 Background.....	1
1.2 Motivation.....	4
1.3 Aims and Objectives	4
1.4 Contribution	5
1.5 Structure of the thesis	5
1.6 Publications.....	6
Chapter 2. Literature review.....	8
2.1 Separation technologies of Post-combustion CO ₂ capture	8
2.1.1 Adsorption	8
2.1.2 Physical absorption.....	9
2.1.3 Cryogenics separation.....	9
2.1.4 Membrane absorption	11
2.1.5 Membrane-based gas separation.....	12
2.1.6 Chemical absorption	13
2.1.7 Summary.....	14
2.2 Solvents of post-combustion CO ₂ capture with chemical absorption.....	15
2.2.1 Amines.....	15
2.2.2 Aqueous ammonia	20
2.2.3 Ionic liquids	21
2.2.4 Potassium-based solvents	22
2.2.5 Summary.....	23
2.3 Modelling of post-combustion CO ₂ capture with chemical absorption: a state-of-the-art review.....	24

2.3.1	Mechanistic models.....	24
2.3.2	Regression model.....	29
2.3.3	Artificial intelligence based models.....	31
2.4	Process Control.....	36
2.4.1	Decentralized Control strategy.....	37
2.4.2	Centralized control strategies.....	44
2.5	Process optimisation.....	47
2.6	Conclusions	51
Chapter 3. Modelling of a Post-combustion CO ₂ Capture Process Using Bootstrap		
Aggregated Neural Networks.....		
3.1	Introduction	52
3.2	Overview of post-combustion CO ₂ capture plant.....	54
3.2.1	gPROMS simulator description	54
3.2.2	Variables selection	56
3.3	BANN model.....	57
3.3.1	Construct of BA-NNs models with sensitivity analysis (SA).....	57
3.3.2	Results and discussions of static model	61
3.3.3	Results and discussions of dynamic BANN model	66
3.4	Optimisation of CO ₂ capture plant	72
3.4.1	Optimisation problem statement	72
3.4.2	Optimal operation	73
3.5	Conclusions	74
Chapter 4. Modelling of a Post-combustion CO ₂ Capture Process Using Extreme Learning		
Machin		
4.1	Introduction	75
4.2	Development of BA-ELM.....	76
4.2.1	Single hidden neural networks.....	76
4.2.2	BA-ELM	78

4.2.3	Characteristic comparison between SLFNNs and ELM.....	80
4.3	Performance evaluation of static BA-ELM model	81
4.4	Performance evaluation of dynamic BA-ELM model	84
4.5	Conclusions.....	92
Chapter 5.	Modelling of a Post-combustion CO ₂ Capture Process Using Deep Belief Network	94
5.1	Introduction.....	94
5.2	Deep Belief Networks.....	95
5.2.1	Restricted Boltzmann Machines	95
5.2.2	Learning algorithm for RBMs	97
5.2.3	Unsupervised pre-training and supervised fine tuning of DBN	100
5.3	Case study: Post-combustion CO ₂ capture plant with chemical absorption.....	102
5.4	Results and discussions.....	102
5.5	Conclusions.....	109
Chapter 6.	Post-combustion CO ₂ capture plant control structure selection	110
6.1	Introduction.....	110
6.2	Multivariate decentralized control analysis	111
6.2.1	RGA range analysis via optimisation	111
6.2.2	RDGA range analysis via optimisation	115
6.2.3	DRGA analysis	117
6.3	Results and analysis of decentralized control structure	118
6.4	MPC control analysis.....	124
6.4.1	State space model for MPC controller.....	124
6.4.2	Objective function	126
6.4.3	Tuning rules of MPC	126
6.5	Results and analysis of MPC control.....	127
6.6	Conclusions.....	134
Chapter 7.	Conclusions and Recommendations for Future Works	135

7.1	Conclusions	135
7.2	Recommendations for future work.....	137
	References	139

List of Figures

Figure 1.1: World net electricity generation by fuel, 2012-2040. (Unit: trillion kilowatt-hours) (EIA, 2016).....	1
Figure 1.2: Total annual anthropogenic GHG emission by gases 1970-2010 (Anderson, 2016).	2
Figure 1.3: A simple schematic diagram of CCS.	3
Figure 2.1: Post-combustion carbon capture by physical adsorption.	9
Figure 2.2: Post-combustion carbon capture with cryogenics separation.	10
Figure 2.3: The schematic diagram of membrane gas absorption.	11
Figure 2.4: The schematic diagram of membrane absorption.	12
Figure 2.5: Simplified process flow diagram of chemical absorption process for post-combustion capture.	14
Figure 2.6: Molecular structures of primary amine, secondary amine and tertiary amine.	15
Figure 2.7: Primary oxidation pathways for MEA (S. B. Fredriksen1, 2013). (a) Autoxidation pathways; (b) Oxidation in the presence of metal ions.	18
Figure 2.8 Thermal degradation pathways for MEA (S. B. Fredriksen1, 2013).	18
Figure 2.9: Commonly used anions and cations of ionic liquids.....	22
Figure 2.10: Level of complexity. Source: Kenig <i>et al.</i> (2001).....	25
Figure 2.11: A diagram of two-film theory.	26
Figure 2.12: The schematic structure of SHLFNNs.	33
Figure 2.13: Architecture of ANFIS.	35
Figure 2.14: Flow chart of controllability analysis.....	36
Figure 2.15: A simplified block diagram of SISO feedback control system.....	41
Figure 2.16: Ultimate cycle tuning procedure. Source: (Love, 2007).....	42
Figure 2.17: Schematic representation of the closed loop system of RTO and MPC controller.	48
Figure 3.1: Simplified process flow diagram of chemical absorption process for post-combustion capture.	54
Figure 3.2 BA-NNs model establishment process.	57
Figure 3.3: An aggregated neural network.	59
Figure 3.4: Static process operation data (inlet gas flow rate (a), CO ₂ mass fraction in inlet flow gas (b), inlet gas flow pressure (c), inlet gas flow temperature (d), lean solvent circulation rate (e), MEA concentration (f), lean solution temperature (g) and capture level (h)).	62

Figure 3.5: Number of hidden neurons in individual neural networks.	63
Figure 3.6: MSE of CO ₂ capture level for individual neural networks.....	64
Figure 3.7: MSE of CO ₂ capture level for aggregated neural networks.	65
Figure 3.8: Static model predictions for CO ₂ capture level on unseen validation data.....	65
Figure 3.9: The time series plot of input variables (inlet gas flow rate, CO ₂ mass fraction in inlet flow gas, inlet gas flow temperature, inlet gas flow pressure, lean solvent circulation rate, lean solvent temperature, reboiler temperature, lean loading).....	67
Figure 3.10: The time series plots of output variables (CO ₂ capture level and CO ₂ production rate).	68
Figure 3.11: MSE of CO ₂ production rate for individual neural networks.....	70
Figure 3.12: MSE of CO ₂ production rate for aggregated bootstrap neural networks.....	70
Figure 3.13: Dynamic model prediction of CO ₂ production rate by BA-NNs on Batch 7.	71
Figure 3.14: Dynamic model prediction of CO ₂ capture level by BA-NNs on Batch 2.....	71
Figure 3.15: Steady-state flow sheet in Aspen Hysys.	72
Figure 3.16. Optimal lean solvent flow rate at different capture level.	74
Figure 4.1: The structure of single hidden layer feedforward networks.	77
Figure 4.2: MSE of CO ₂ capture level for individual ELM.....	81
Figure 4.3: MSE of CO ₂ capture level for aggregated ELM.	82
Figure 4.4: Static BA-ELM model predictions for CO ₂ capture level on unseen validation data.	83
Figure 4.5: MSE of CO ₂ production rate for individual ELM models on training & testing data (a) and validation data (b).	86
Figure 4.6: MSE of CO ₂ production rate for bootstrap aggregated ELM training & testing data (a) and validation data (b).	87
Figure 4.7: Dynamic model predictions of CO ₂ production rate on Batch 7 using BA-ELM (a) and BA-NNs (b).....	88
Figure 4.8: MSE of CO ₂ capture level for individual ELM models on training & testing data (a) and validation data (b).	89
Figure 4.9: MSE of CO ₂ capture level for individual ELM models on training & testing data (a) and validation data (b).	89
Figure 4.10: Dynamic model prediction of CO ₂ capture level using BA-ELM (a) and BA-NNs (b).	91
Figure 5.1: The structure of Restricted Boltzmann Machines.	96
Figure 5.2: Markov chain in two-stage Gibbs sampler and CD for training RBM.....	99
Figure 5.3: The structure of DBN with continuous-valued inputs.....	101

Figure 5.4: Multi-rate sampling in soft sensor modelling. Source: Shang <i>et al.</i> (2014).	101
Figure 5.5: One-step-ahead predictions of CO ₂ production rate by SLNNs (top) and DBN (bottom).	104
Figure 5.6: Multi-step-ahead predictions of CO ₂ production rate by SLNNs (top) and DBN (bottom).	105
Figure 5.7: One-step-ahead predictions of CO ₂ capture level by SLNNs (top) and DBN (bottom).	107
Figure 5.8: Multi-step-ahead predictions of CO ₂ capture level by SLNNs (top) and DBN (bottom).	108
Figure 6.1: A simplified PCC block diagram.	111
Figure 6.2: Frequency-dependent RGA analysis.	117
Figure 6.3: The control structure based on RGA analysis using PID controller.	118
Figure 6.4: Magnitudes of <i>Lcm max</i> with different frequencies.	120
Figure 6.5: Response to set point changes in CO ₂ capture level at 100s.	121
Figure 6.6: Response to set point changes in lean loading at 200s.	122
Figure 6.7: Response of CO ₂ capture level (a) and lean loading (b) to disturbance in inlet flue gas flowrate.	124
Figure 6.8: Simulink model with MPC controller.	127
Figure 6.9: Response to set point changes in CO ₂ capture level at 100s by MPC control. ...	128
Figure 6.10: Changes in manipulated variables during the set point changes of CO ₂ capture level.	129
Figure 6.11: Response to set point changes in lean loading at 200s by MPC control.	130
Figure 6.12: Changes in manipulated variables during the set point changes of lean loading.	131
Figure 6.13: Response to step change in inlet flow gas rate at 100s by MPC control.	132
Figure 6.14: Changes in manipulated variables during the set point changes of inlet flow gas flow rate.	133

List of Table

Table 2.1: Summary of various promoters.....	23
Table 2.2: List of manipulated variables and controlled variables. Source: (Nittaya <i>et al.</i> , 2014).	44
Table 2.3: Control loops of three control structures. Source: (Nittaya <i>et al.</i> , 2014).....	44
Table 3.1: Equipment Specification.....	55
Table 3.2: The dynamic sampling information from simulator and to develop BANN model.	55
Table 4.1: Performance comparison of BA-ELM and BA-NNs for CO ₂ production rate.....	85
Table 4.2: Performance comparison of BA-ELM and BA-NNs for CO ₂ capture level.....	92
Table 5.1: Comparison of modelling results of DBN and SLNNs on CO ₂ production rate. .	103
Table 5.2: Comparison of modelling results of DBN and SLNNs on CO ₂ capture level.	106
Table 6.1: Lower band and upper band of RGA and their corresponding steady state gains for $\alpha = 0.01$	113
Table 6.2: Lower band and upper band of RGA and their corresponding steady state gains for $\alpha = 0.1$	114
Table 6.3: lower band and upper band of RGA and their corresponding steady state gains for $\alpha = 0.25$	115
Table 6.4: Recommended pairing for a frequency range.....	118

Nomenclatures

a	coefficients of input variables of regression models
a_i	the width of the Gaussian function
A_i	the linguistic labels associated to input m
A	an appropriate matrix of size $(2n^2) \times (n^2)$
b	bias of hidden nodes
B_i	the linguistic labels associated to input n ,
c_i	the centre of the Gaussian function.
c	a vector of size $(2n^2) \times 1$ containing the lower and upper bounds of K
d	is a disturbance matrix
F	an indicator represents the extent of accuracy the model has improved
G	gas flow rate
G	the process gain matrix
G_d	the disturbance gain matrix
G_{diag}	a diagonal matrix of G .
$G_c(s)$	PID controller transfer function
G_s	controller setting gain
H	hidden layer output matrix
K_1	forward reaction rate
K_{-1}	reverse reaction rate
K	temperature
K	a steady state gain matrix

K_d	disturbance gain matrix
K_p	proportional gain
K_i	integral gain
K_d	derivatives gain
K_{jk}	the jk th element of K^{-1}
K_u	the ultimate gain using ultimate cycle tuning rules
L	lean solvent flow rate
m	mass fraction
o_i	output of a standard SLFNs
$O_{1,i}$	the degree to which the inputs m is related to A_i
$O_{1,j}$	the degree to which the input n is related to B_j
P_u	the period of constant oscillation using ultimate cycle tuning rules
Q_{cond}	condenser heat duty
Q_{tank}	reboiler heat duty
r	correlation coefficient
R	multiple correlation coefficient
S	the standard error
t	an indicator represents the extent of input contribution to predicted variable
T	a vector of the target values of network outputs.
T_{cond}	temperature of condenser
v_k	measure disturbance variables
v	flow rate
w	weights between hidden nodes and input nodes

w_k	unmeasured disturbance variables.
X_n	input variable of regression model
Y	output variable of regression model
y_{sp}	the desired output in control loop
y	the controlled output in control loop
%CC	capture efficiency

Greek letters

β	weight linking the i th hidden node and the output node
β_{ij}	a function of \mathbf{K} and \mathbf{K}_d
ε_i	difference between predicted and actual output values.
λ	the values of elements in RGA matrix
λ_{Dij}	i,j th element of the DRGA
μ	a vector of manipulated variables
μ	control signal
τ_i	integral time constant
τ_D	derivative time constant
η	production rate

Mathematical operators

Σ	summation
$ $	modulus
exp	exponential function
ln	natural logarithm function
log	10 base logarithm function

Abbreviations

AMP	adenosine monophosphate
AC	activated carbon
ANN	artificial neural networks
ANFIS	adaptive-network-based fuzzy inference system
BP	back propagation
CCS	carbon capture and sequestration
CV	controlled variable
DEA	diethanolamine
DIPA	Diisopropylamine
DRGA	dynamic relative gain array
DE	differential evolution
DBN	deep belief networks
ESP	electrostatic precipitators
EDTA	ethylenediaminetetracetic acid
ELM	extreme learning machine
FOPDT	first-order plus dead time
GHG	greenhouse gas
GRDG	general relative disturbance gain
IMC	Internal Model Control
LM	Levenberg-Marquardt
MOF	Metal Organic Frameworks
MEA	monoethanolamine

MLR	multiple linear regression
MSE	mean squared error
MV	manipulated variable
MPC	model predictive control
MIC	Morari index of integral controllability
PCC	post-combustion carbon capture
PZ	Porphyrizin
PID	proportional–integral–derivative
PCR	principal component regression
RG	relative gain array
RDGA	relative disturbance gain array
RDG	relative disturbance gain
RBM	restricted Boltzmann machine
SCNR	Selective Noncatalytic Reduction
SHLFNNs	single hidden-layer feedforward neural networks
SGC	Scaled Conjugate Gradient
SISO	single input single output
SA	sensitivity analysis
SVM	support vector machine
TEA	triethanolamine
VLSE	vapor-liquid-solid thermodynamic equilibrium

Chapter 1. Introduction

1.1 Background

Until now, fossil fuel, especially coal, plays a vital role in electricity generation worldwide. Specifically, as seen from figure 1.1, coal-fired power plants currently supply 40% of global electricity (EIA, 2016). As reported, China and Southeast Asia have added a large number of coal-fired power plants from 1990-2040 (Minchener, 2016). This is because coal is characterized as an inexpensive and affordable power source. Beyond that, coal-fired power plants can offer some advantages to operators, such as its flexible operation to changes in supply and demand (Lawal *et al.*, 2010). However, the process of burning coal would possibly release a large amount of hazardous gas emissions, such as sulphide, nitride and carbide etc. Especially, the amount of CO₂ emission per unit of electricity released by coal-fired power plants is twice as much as their natural gas counterparts (Lawal *et al.*, 2009b). Proverbially, CO₂ is considered as the main component of GHG gas, which leads to the global climate change.

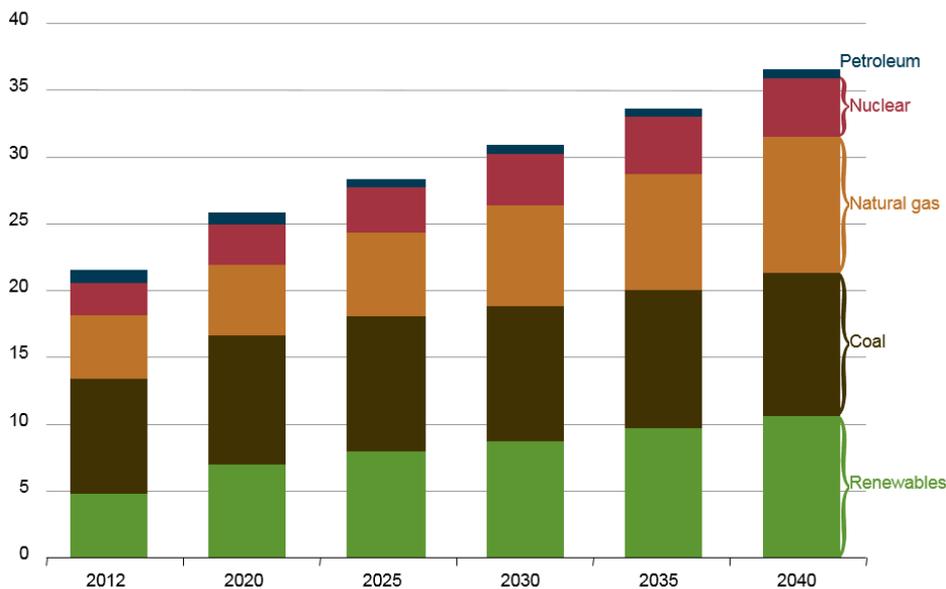


Figure 1.1: World net electricity generation by fuel, 2012-2040. (Unit: trillion kilowatt-hours) (EIA, 2016).

Global climate change, especially temperature increase, has become a key concern of our society. Due to the accelerated build-up of greenhouse gas (GHG) emission in atmosphere, people have to take steps to prevent the situation. In the past a few decades, numerous climate change policies were launched, but nonetheless as it can be seen from figure 1.2, annual GHG

emission increased by 1.0 GtCO₂-eq (Giga-tons carbon dioxide equivalent) (2.2%) per year from 2000 to 2010, compared to 0.4 GtCO₂-eq(1.3%) per year, from 1970 to 2000(Anderson, 2016). A rapidly growing population plus industrialization, with corresponding increase in energy demand, is likely to lead to increasing amount of GHG emission. Under the circumstances, the Intergovernmental Panel on Climate Change advocated that, compared to the emission levels in 1990, a 50% reduction of CO₂ emission is needed in 2050. (Metz and Intergovernmental Panel on Climate Change. Working Group III., 2007).

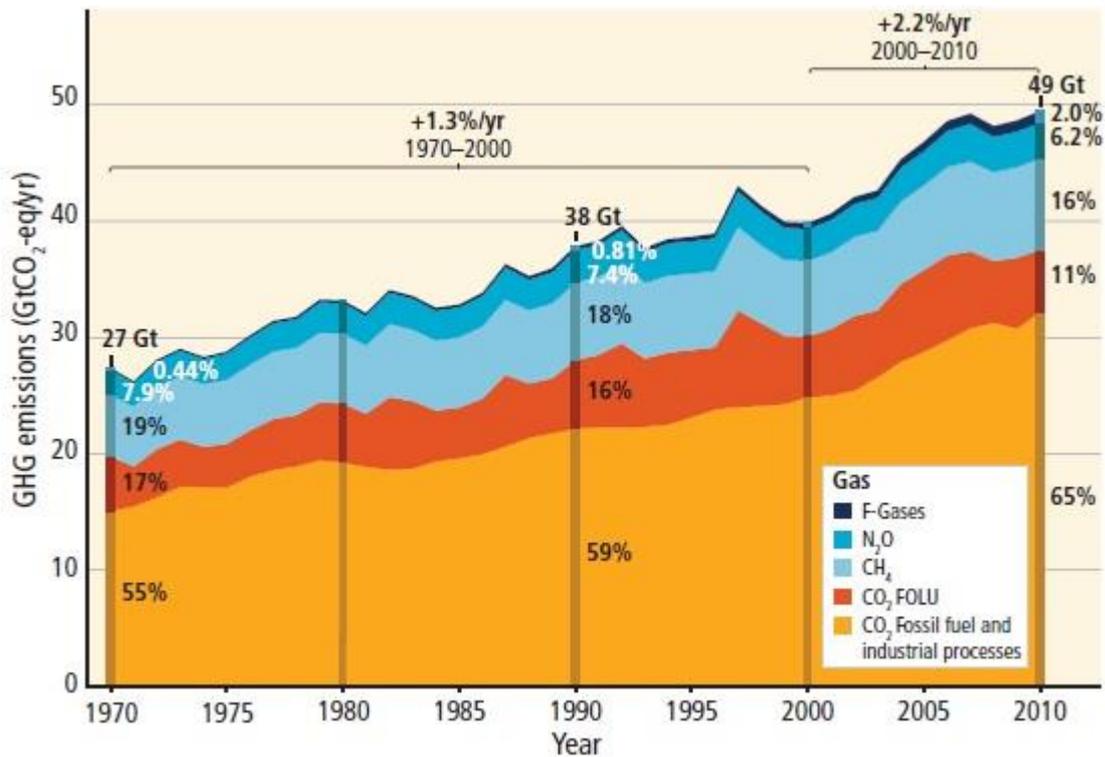


Figure 1.2: Total annual anthropogenic GHG emission by gases 1970-2010 (Anderson, 2016).

As mentioned above, the main source of worldwide CO₂ emission is the combustion of coal in coal-fired power plants. To prevent greenhouse effect, it is necessary to seek suitable technologies to reduce the CO₂ gas emission produced from coal-fired power plants. Carbon capture and sequestration (CCS) is explored as a unique and important technique for the sustainability of coal-fired power plant, because of its efficiency and effectiveness in reducing CO₂ emission (Metz and Intergovernmental Panel on Climate Change. Working Group III., 2005). As can be seen from figure 1.3, CCS is generally consisted of three components: capture, transportation and storage. Firstly, capture technologies separate CO₂ from gases produced in industrial processes by one of three methods: post-combustion, pre-combustion and oxyfuel combustion. Then CO₂ is transported by pipeline or ship for safe storage. A large amount of CO₂ is transported annually for commercial purposes by pipeline,

ship and road tanker. Lastly, CO₂ is stored carefully several kilometres below the earth surface. Recently, CCS has been making significant gains in the world. In Norway, CCS technology has been applied to reduce CO₂ for 20 years and it permanently and safely stores 17 million tons of CO₂ deep under the North Sea. In Australia, the first geosequestration project was launched by an industrial plant in April 2008, owned by a non-profit research collaboration. It has stored 65000 tonnes of carbon dioxide approximately two kilometres below the surface. According to the International Energy Agency (IEA), approximately 570 GW of global coal-fired power plants will be integrated with CCS in 2050 and 40GtCO₂ will be captured in the period to 2050 (Naceur and McCulloch, 2016).

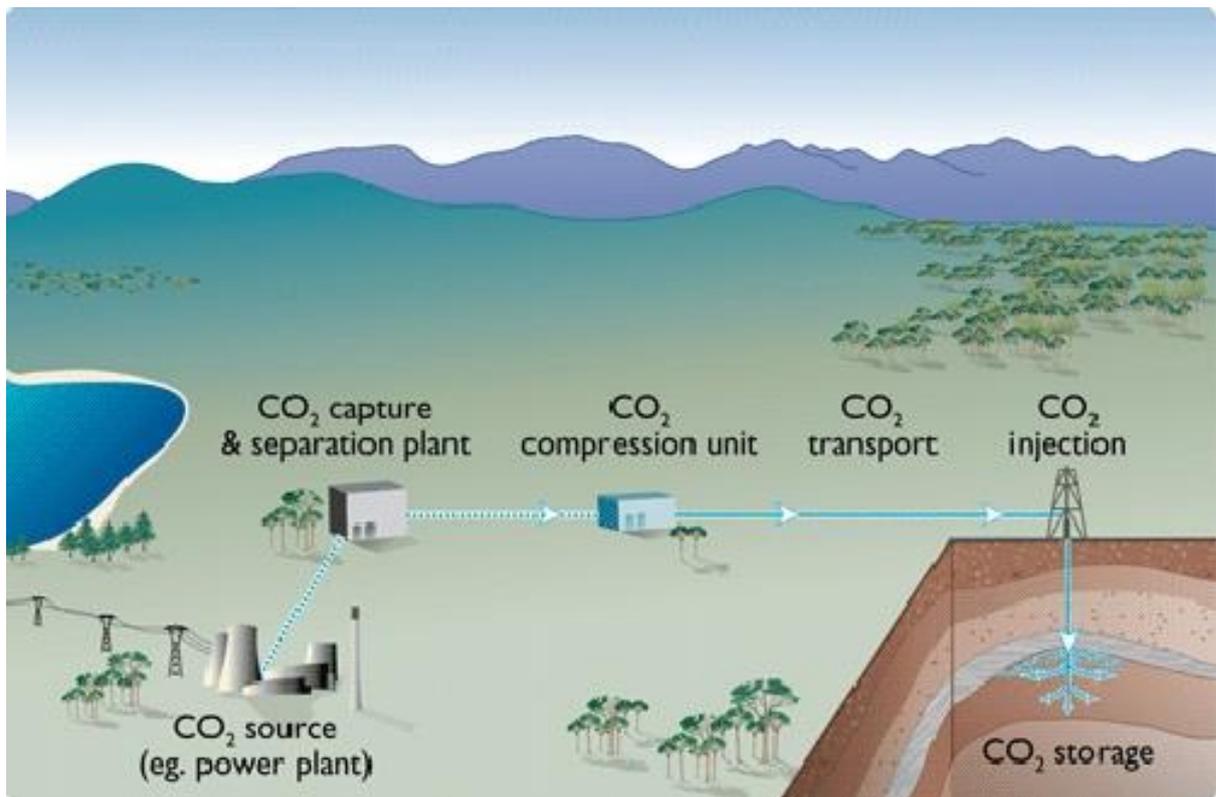


Figure 1.3: A simple schematic diagram of CCS.

Amongst the capture technologies, amine-based post-combustion CO₂ capture process is now treated as the first choice for large scale CO₂ capture. This is because this technology can retrofit the existing power generation plants easily and capture low partial pressure of CO₂ in flue gas (Biliyok *et al.*, 2012a). However, it still has some disadvantages, one of which is the large energy requirement for absorbent regeneration. In industrial scale, the thermal energy for regeneration usually comes from extracted steam from the low pressure steam turbines of upstream power plant, which will reduce the efficiency of the coal-fired power plant. As a result, a slight adjustment of capture efficiency will affect the thermal consumption in the process and thus the profitability of upstream power plants. Tock and Marechal (2014) have

attempted to maximise the overall capture efficiency and concurrently minimise the electricity production costs. They indicated the trade-off between the capture efficiency (from 75% to 90%) and electricity production costs.

The energy requirement in the regeneration unit is strongly influenced by CO₂ capture target, operation conditions and equipment dimensions. As a result, the exploration of relationships between the process variables appears much significant for the selection of control structures. The developed models and selected control structures can then be used in the optimisation of CO₂ capture process integrated with coal-fired power plants.

1.2 Motivation

Although the carbon capture process has been extensively researched in last few decades, the issue of process optimisation is still the focus of most concern. Finding the trade-off between energy consumption and CO₂ capture efficiency is significant in process optimisation. This is because, if the energy consumption is high, the electricity production costs will be high as well. The low energy consumption will possibly reduce the capture efficiency. Therefore, the basic suggested question is what and how process variables affect carbon capture efficiency? What is the relationship between energy consumption and capture efficiency? Consider the selection of control structures, what is the best one to control?

The development of modelling techniques is significant to explore the intrinsic features of process and determine the control structures, thereby optimizing the carbon capture process. A number of attempts to CO₂ capture process modelling have been conducted by three different ways: mechanistic, statistic and neural networks. However, the techniques still need further improvements, regarding computational time and model generalization capability.

This thesis concentrates on the development of novel methods to model and optimize CO₂ capture process integrated with coal-fired power plant by applying advanced modelling and control techniques. Development of reliable and accurate data driven models is the focus in this research.

1.3 Aims and Objectives

This project is aimed to develop efficient operating techniques for CO₂ capture process integrated with coal-fired power plant.

The main objectives are:

- Developing rigorous data-driven models of post-combustion carbon capture with chemical absorption.
- Process analysis, optimal design and operation of post-combustion carbon capture with chemical absorption.
- Control structure selection, system identification and control system design of post-combustion with chemical absorption.

1.4 Contribution

This thesis contributes to developing different modelling techniques and control strategies to improve capture efficiency of post-combustion carbon capture process. The post-combustion carbon capture process is targeted because of its unique effective application to current coal-fired power plant and its excellence use to improve climate change.

In last few decades, the study of carbon capture process focused on traditional modelling techniques, such as mechanistic, statistical and single-hidden layer neural network models. This study goes a further step to use advanced data-driven modelling methodologies, such as bootstrap aggregated neural networks (BA-NNs), bootstrap aggregated extreme learning machine (BA-ELM) and deep belief networks (DBNs), to improve the capture efficiency of carbon capture process.

This study also demonstrates the comparison of decentralized control strategy (PID based control) and centralized control strategy (MPC-based control) in designing post-combustion carbon capture process. MPC is indicated as more advanced control methodology to improve the efficiency of the carbon capture process.

Furthermore, three first-author papers have been published on the relevant journals, which is detailed in the following section.

1.5 Structure of the thesis

In chapter 2, the general literature review of separation technologies, the solvents used for chemical absorption, the modelling technologies, and the controllability and optimisation analysis on post-combustion carbon capture process is presented. The knowledge gap related to the research is also highlighted

Chapter 3 presents a study of modelling post-combustion CO₂ capture process using bootstrap aggregated neural networks. The multiple feedforward neural networks models are developed from bootstrap re-sampling replications of the original training data and are combined, thereby enhancing model accuracy and reliability. Both static and dynamic models are developed and they offer accurate predictions on unseen validation data. In addition, the optimisation problem of the CO₂ capture process, which is based on the neural network static model, is solved.

Chapter 4 is focused on modelling of post-combustion CO₂ capture process using bootstrap aggregated extreme learning machine (ELM). In this chapter, the weights between input and hidden layers are randomly assigned and the weights between the hidden layer and output layer are obtained using principal component regression (PCR). ELM can provide fast learning speed and good generalisation performance.

In Chapter 5, a neural network with multiple hidden layers, called deep belief network (DBN), is explained. The hidden layer of restricted Boltzmann Machine (RBM) is able to extract a deep hierarchical representation of training data. Two stages are included in DBN technique: an unsupervised pre-training phase and a supervised back-propagation phase. A greedy layer-wise unsupervised learning algorithm is introduced to optimize DBN, which can bring better generalization than a single hidden layer neural network.

In chapter 6, the efficient decentralized control strategy based on general relative disturbance gain (GRDG) and dynamic relative gain array (DRGA) is discussed. Then, the centralized control scheme, model predictive control (MPC), is presented and compared to decentralized control structure for their performance.

Chapter 7 is focused on highlighting the conclusions from the research and recommendations for future work.

1.6 Publications

Book chapter

1. Bai Z, Li F, Zhang J, Oko E, Wang M, Xiong Z, Huang D. ‘Modelling of a Post-combustion CO₂ Capture Process Using Bootstrap Aggregated Extreme Learning Machine’. *Computer Aided Chemical Engineering*, vol 38, 2007-2012.

Published journal papers:

1. Li F, Zhang J, Oko E, Wang M. 'Modelling of a Post-combustion CO₂ Capture Process Using Neural Networks'. *Fuel* 2015,151, 156-163.
2. Li F, Zhang J, Oko E, Wang M. 'Modelling of a Post-combustion CO₂ Capture Process Using Extreme Learning Machine'. *International Journal of Coal Science & Technology*, 2017, 4(1), 33-40.
3. Li F, Zhang J, Shang C, Huang D, Oko E, Wang M. 'Modelling of a Post-combustion CO₂ Capture Process Using Deep Belief Network'. *Applied Thermal Engineering*, 130,997-1003.

Peer reviewed conference paper

1. Li F, Zhang J, Oko E, Wang M. 'Modelling of a Post-combustion CO₂ Capture Process Using Extreme Learning Machine'. In: 2016 21st International Conference on Methods and Models in Automation and Robotics (MMAR). 29 August – 1 September, 2016, Miedzyzdroje, Poland: IEEE, 1252-1257

Conference Presentations:

1. Fei Li and Jie Zhang (2014) 'Modelling of a Post-combustion CO₂ Capture Process Using Neural Networks', 10th European Conference on Coal Research and its Applications, 15th-17th September, 2014, Hull, UK.
2. Fei Li and Jie Zhang (2015) 'Steady-state and dynamic models of post-combustion CO₂ chemical absorption process for coal-fired power plants'. 29th May, 2015, Newcastle University, UK.
3. Fei Li and Jie Zhang (2016) 'Controllability analysis and control system design for a post-combustion CO₂ capture plant based on centralized and decentralized control techniques'. 3rd June, 2016, Newcastle University, UK.

Chapter 2. Literature review

2.1 Separation technologies of Post-combustion CO₂ capture

The flue gas exhausting from coal-fired power plants are approximately consisted of 76-77% N₂, 12.5-12.8% CO₂, 6.2% H₂O, 4.4% O₂ and a small amount of CO, NO_x and SO₂ gases. Post-combustion CO₂ capture process is able to remove CO₂ emission after the combustion of the fossil fuel in combustor. The advantage comes at easy retrofitting of the most existing coal-fired plants for NO_x, SO_x and CO₂ capture. The large energy consumption in the regenerator unit accounts for 75% - 80% of the total cost of CCS (Davison, 2007). A number of separation technologies can be employed for post-combustion CO₂ capture process. In details, they are adsorption, physical absorption, chemical absorption, cryogenics separation and membranes respectively. A brief introduction of these technologies will come as follows.

2.1.1 Adsorption

Adsorption is a physical process that adopts adsorbents to attach CO₂ to its surface. The physical conditions, including temperature, vacuum and pressure swing operations, will impact the regeneration of adhered gas. Figure 2.1 indicates the simple CO₂ adsorption process, which is consisted of adsorption and desorption steps. The flue gas is pumped through the column, meanwhile, the sorbents will adsorb CO₂ in flue gas. After that, CO₂ desorption from the sorbents can be achieved with a pressure swing or temperature swing cycle. The affinity between the sorbents and CO₂ is the key factor to determine adsorption performance. However, if the affinity is strong, it is difficult to desorb CO₂ due to the requirement of large energy consumption. Therefore, the selection of suitable sorbents which can carefully balance adsorption and desorption steps, seem more important. Recently, solid adsorbents, such as Metal Organic Frameworks (MOF), mesoporous silicates, zeolites, alumina and activated carbons have been extensively applied for gas mixer separation. New adsorbents such as monolithic carbon fibre adsorbents (Thiruvengkatachari *et al.*, 2009), MgO material ball-milling treated for 2.5 hours (MgO-BM2.5h) (Elvira *et al.*, 2016), have been researched with post-combustion CO₂ capture process. On the other side, a suitable replacement strategy of sorbents also appears rather significant to capture efficiency, because it is able to maintain the plant performance. Under the circumstance, a novel removal strategy, based on density separation, was suggested (Colantuono and Cockerill, 2017). Specifically, the removed sorbents was sorted by their density, in which the lower density represents the lower capture capacity of older particles. Then the older fractions of separated sorbents are replaced by the fresh materials. However, when integrate carbon capture plant with large

scale power plant, adsorption is not viable because of its limited adsorption capacity of most available sorbents.

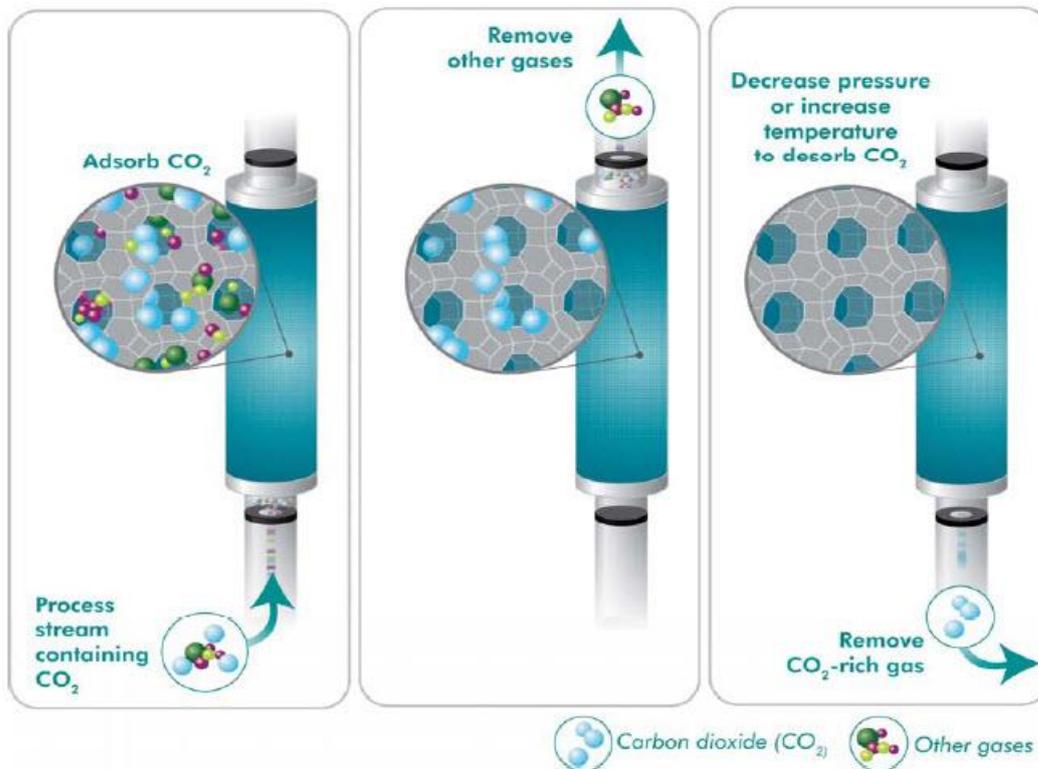


Figure 2.1: Post-combustion carbon capture by physical adsorption.

2.1.2 Physical absorption

Based on Henry's Law, CO_2 is physically absorbed by some solvents, such as dimethyl ethers of polyethylene glycol and methanol. Absorption occurs with increasing pressure and with decreasing temperature, whilst regeneration of solvents takes place with heat, pressure reduction or both. The typical physical solvents are including cold methanol (Rectisol process), dimethyl ether of polyethylene glycol (Selexol process), N-methyl-2-pyrrolidone (Purisol process) and propylene carbonate (Fluor Solvent process). The absorption capacity of absorbents is related to CO_2 partial pressure. The advantage of physical absorption is the relative little energy consumption, but CO_2 partial pressure needs to be high. Therefore, the main challenge of physical absorption in post-combustion carbon capture process is the high cost of treating low CO_2 partial pressure (<15 vol %) in flue gas (Wang *et al.*, 2011).

2.1.3 Cryogenics separation

The cryogenics separation of CO_2 from exhausting flue gas is processed in ultra-low temperature. As well known, the critical temperature and triple point of CO_2 are 31.1°C and -

56.6°C. In other words, the physical state of CO₂ may change to liquid or solid under the according operations. Figure 2.2 provides a detailed schematic diagram of post-combustion CO₂ capture process with cryogenics separation. The flue gas is cooled in condensing heat exchanger by exchanging heat and dried before compression. Following that, the treated gas is further cooled in heat recovery heat exchanger to separate different contaminants. Lastly, the remaining light gas is expanded to further cool to extract the solid CO₂. Some researchers have focused on the development of cryogenics separation of CO₂ in carbon capture process, by application of multi-compression stages along with the intercooler (Meisen and Shuai, 1997; Zanganeh *et al.*, 2009). To further improve this technology, Hart and Gnanendran (2009) has proposed Cryocell technique, based on a clear understanding of the vapour-liquid-solid thermodynamic equilibrium (VLSE) across a Joule-Thomson valve, to remove CO₂ from natural gas. The pre-cooling temperature and the isenthalpic flash pressure are significant control variables in the process. The gas feeding into the Cryocell separator is supposed to be pre-treated to have certain CO₂ and ethane plus composition specifications. The advantage of cryogenics separation is its disposal ability with high CO₂ concentration and high CO₂ recovery, while the disadvantage is inherent energy intensive. Besides, cryogenics separation is further used in combination with membrane technology, which will be introduced next.

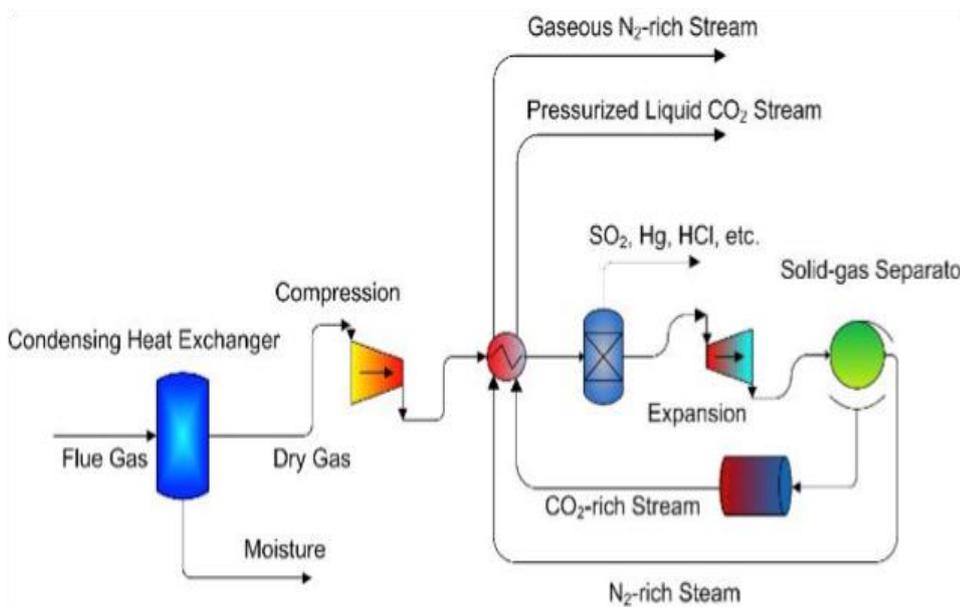


Figure 2.2: Post-combustion carbon capture with cryogenics separation.

2.1.4 Membrane absorption

Membrane absorption applies membrane only as a contact device between gas and liquid, while the membrane cannot provide supererogatory selectivity. As shown in figure 2.3, flue gas will be feed through a bundle of membrane tubes, while the solvents such as amine pass through the shell side of the bundle. The gas-liquid contactors are mainly porous hydrophobic membranes. CO₂ was physically absorbed by solvent according to its diffusion through the membranes and desorbed from solvent by heat. Therefore, the parameters affecting mass transfer between two phases seems particularly important. As mass transfer is taking place at the pores of membranes, the properties of membranes, such as pore size, pressure, temperature, porosity and flow rates of absorbents and flue gas, draws attention of researchers. Recently, the studies about membrane absorption is mainly focused on mass transfer, membrane wetting, membrane development and absorbent selection (Favre and Svendsen, 2012; Mosadegh-Sedghi *et al.*, 2014; Sreenivasulu *et al.*, 2015; Zaidiza *et al.*, 2016; Abdulhameed *et al.*, 2017). The outstanding point of this method is the membranes are more compact so that they will not be influenced by flooding, entrainment, channelling or foaming. In addition, membrane absorption has high operation flexibility and extremely high interfacial area, which result in low costs and high efficiency. Nevertheless, extra mass transfer resistance from membrane and membrane wetting will increase mass transfer resistance. In this case, the requirement of high CO₂ partial pressure is of great importance. In other words, membrane absorption is more suitable for high concentration of CO₂ in flue gas, such as the flue gas coming from oxyfuel process.



Figure 2.3: The schematic diagram of membrane gas absorption.

2.1.5 Membrane-based gas separation

Gas separation membrane allows preferential selective permeation of gas mixture. The structure of membranes for gas separation is usually formed as tube-and-shell configuration (see figure 2.4). The main types of membrane materials are polymer. In details, the membrane applied in this method can determine the selectivity, which means the permeation exists in the gas mixture due to partial pressure of different constituent species. The reason why this method is superior is there is no requirement of a separation agent because of non-regeneration stages. In addition, it is suitable for retrofitting applications due to its compact and lightweight characters, as well as low maintenance requirement. However, the separation capacity is too low as stated in International Energy Agency. and Organisation for Economic Co-operation and Development. (2004). This is because the membrane selectivity depending on materials for separation of CO₂ and N₂ is larger than any other existing applications, such as O₂ and N₂, CO₂ and CH₄. Therefore, balancing between selectivity and permeability is becoming the main concern as to membrane-based gas separation.

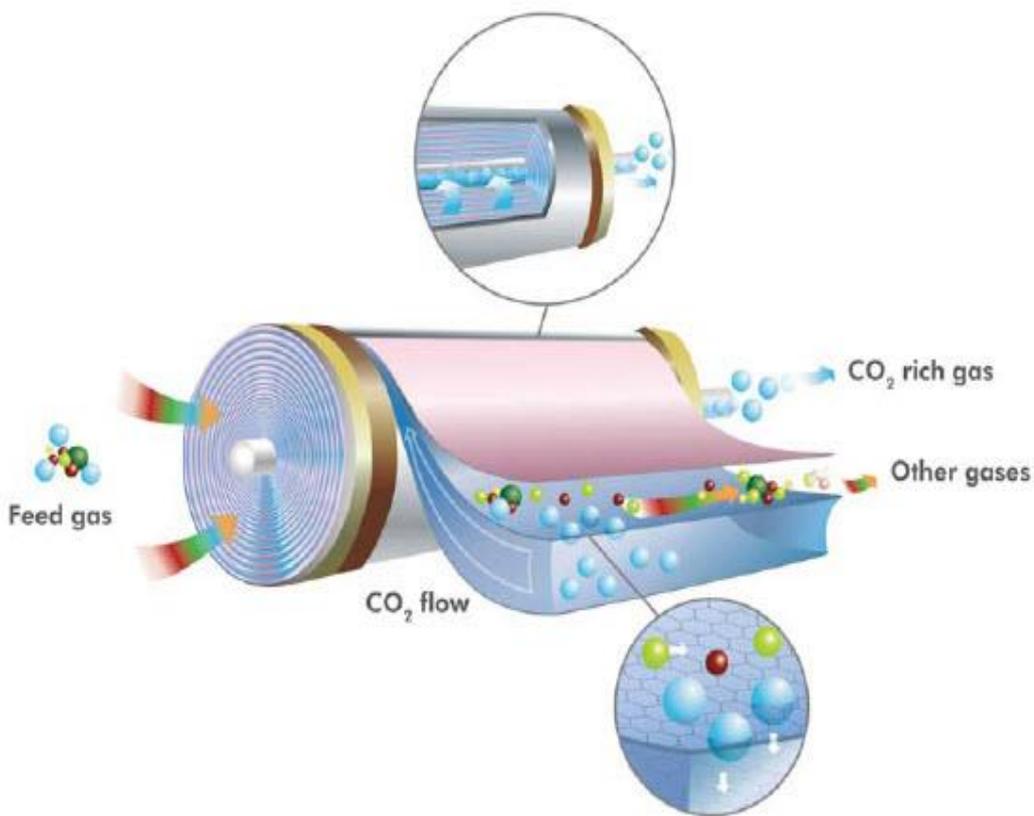


Figure 2.4: The schematic diagram of membrane absorption.

2.1.6 Chemical absorption

The post-combustion carbon capture with chemical absorption is a widely used technology to remove CO₂ from exhausting flue gas. It is a chemical reaction process in which the chemical solvent is used to react with CO₂ to form a new intermediate compound with weak bond, and then CO₂ is regenerated in the circumstance of high temperature. The high selectivity and final pure CO₂ stream make chemical absorption being widely used for CO₂ capture of industrial flue gas.

The flue gas from the power plant should be pre-treated before sending to the system units (Wang *et al.*, 2011). The reason is that SO₂ and NO₂ involved in flue gas can react with amine solution and form heat stable salts, which decrease the regeneration capacity of lean amine solution. Thus, SO₂ is removed by Flue Gas Desulphurization (FGD) unit. NO_x is removed by applying Selective Catalytic Reduction (SCR), Selective Noncatalytic Reduction (SCNR) or low NO_x burner. In addition, the particulate matter such as fly ash could cause foaming in the absorber and regenerator, which affects the performance of the CO₂ capture system. Fly ash is removed by either electrostatic precipitators (ESP) or filters. The oxygen, which could cause corrosion and degradation of alkanolamines, is removed by Fluor Daniel ECONAMINE™ process. Lastly, the temperature of flue gas should be maintained between 45°C and 50°C, as it is helpful for improving CO₂ absorption and reducing solvent loss.

As described in figure 2.5, one of the widely used technologies for CO₂ capture with chemical absorption is mainly consisted of two parts: the scrubbing column (absorber) and regeneration column (regenerator), which are both packed columns. The flue gas from power plant is fed into the bottom of absorber and contacted counter-currently with lean amine solution from the top side. The lean amine solution chemically reacts and absorbs CO₂ in flue gas. Then the treated gas stream containing much lower CO₂ contents is generated and leaves from the top of absorber. The amine solution of much more CO₂ (now rich amine), coming from the bottom of absorber, is pumped to the stripper unit after preheating in cross heat exchanger. In the stripper, the absorber amine solution is regenerated by heating rich amine in a reboiler. The low-pressure steam from power plant is used in reboiler to maintain the operating condition, resulting in large energy consumption. In details, the heat supplied in the reboiler is used for increasing the rich solution coming from the absorber, desorption heat required for separating CO₂ in rich amine, and vaporization of gas in stripper (Mores *et al.*, 2012b). After that, the vapour is cooled in condenser and returned to the regenerator, while CO₂ leaves the condenser and is compressed for storage. In addition, the amine solution

coming from the regenerator (now lean amine) is cooled in cross heat exchanger by exchanging heat with rich amine and pumped back to absorber for absorption. Therefore, from the view point of the process system, the performance is strongly influenced by the operating conditions, such as temperature and pressure in absorber, stripper, reboiler and condenser. On the other hand, the conditions of flue gas and amine solution (temperature, pressure, composition and flow rate) are necessary to be investigated for their impacts on the process performance.

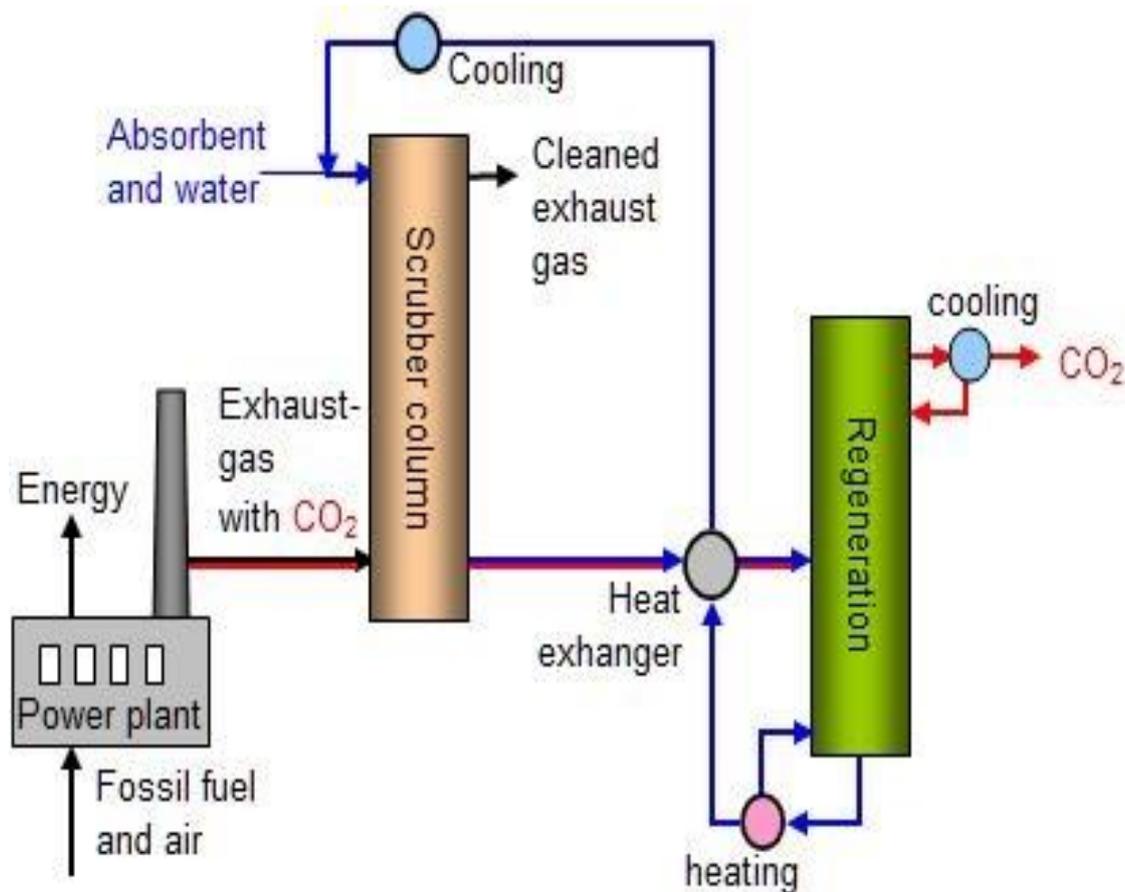


Figure 2.5: Simplified process flow diagram of chemical absorption process for post-combustion capture.

2.1.7 Summary

Chemical absorption is selected as the most efficient and economical way to capture CO₂ in the stream. The reason why this technology is superior is it can easily retrofit current existing power plants and capture low partial pressure CO₂ in flue gas stream. Besides, this system can be operated at ordinary temperature and pressure. However, there are several disadvantages for using chemical absorption technology. In details, the scrubbing solvents will be degraded at high temperature, resulting in a corrosion of the regeneration system. As well, the

regeneration of the scrubbing solvents requires a large amount of energy, which leads to a costly operation. The application of various solvents to post-combustion carbon capture system with chemical absorption will be discussed in the next section.

2.2 Solvents of post-combustion CO₂ capture with chemical absorption

The scrubbing solvents react with CO₂ chemically in absorber and a new bond between the two components is formed, while during regeneration, the bond is broken by heat and high concentration of CO₂ is yielded. The main challenge of post-combustion carbon capture system with chemical absorption is the selection of suitable solvents, which are regenerated with minimal energy.

2.2.1 Amines

Amines, especially alkanolamines, are widely used as absorbents to treat the flue gas stream in post-combustion CO₂ capture process. The commonly used amines are classified into three categories: Primary amine (RNH₂), secondary amine (R₂NH) and tertiary amine (R₃N) (Kenarsari *et al.*, 2013). Figure 2.6 presents the molecular structures of these three categories of amines. In details, primary amines arise when one of three hydrogen atoms in ammonia is replaced by an alkyl or aromatic. Whereas secondary amines have two organic substitutes (alkyl, aryl or both) and one hydrogen atom connected with nitrogen. Lastly, tertiary amines have three organic substitutes bound to the nitrogen atom.

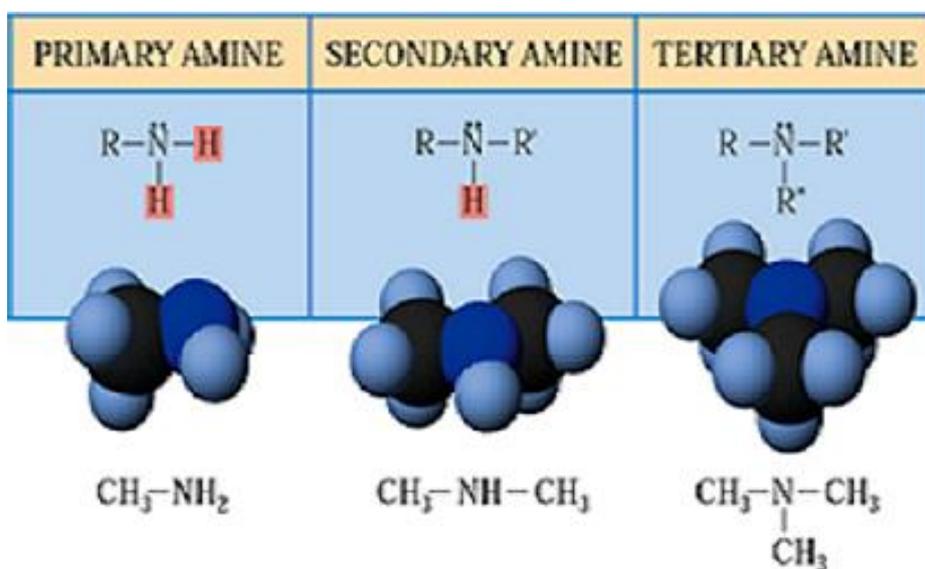
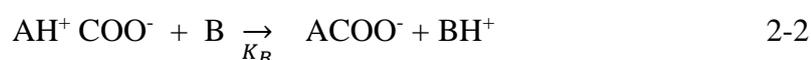


Figure 2.6: Molecular structures of primary amine, secondary amine and tertiary amine.

As to CO₂ reactions with amines, there are three different mechanisms to describe the pathways according to the distinct molecular structures of amines. It is theoretically summarized as follows:

(1) Zwitterion Mechanism

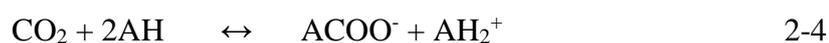
Zwitterion mechanism usually describes the CO₂ reaction with primary, secondary and sterically hindered amines. It was proposed originally by Caplow (1968), which suggested that zwitterion is formed as an intermediate by the reaction between CO₂ and amine (AH) and the reaction is separated into two steps.



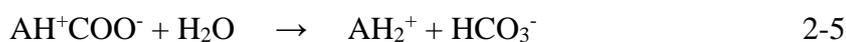
The reaction 2-2 describes the intermediate AH⁺COO⁻ is deprotonated by a base (B), thereby forming the carbamate ACOO⁻. If the base B is the amine itself, the reaction 2-2 can be described instead as:



Therefore, the overall reaction of CO₂ with primary or secondary amines is the sum of reactions 2-1 and 2-2:



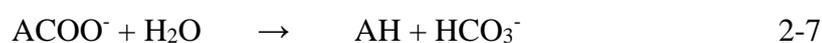
However, if the amine is sterically hindered, the reaction of intermediate and water is easier to arise, rather than reaction 2-3.



Under the circumstances, the reaction 2-6 occurs for bicarbonate formation.



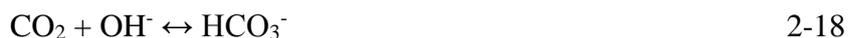
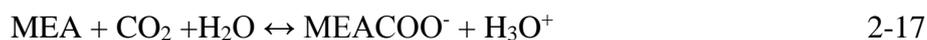
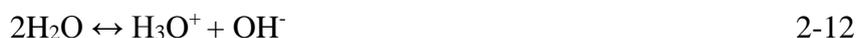
This is because the steric effects will result in the reduced stability of carbonate, which is as stable as zwitterion. In this case, the carbonates in reaction 2-3 may also perform a hydrolysis reaction, forming bicarbonates and releasing free amine molecules.



The reproduced free amine molecules will react with CO₂ again. As a result, the amount of bicarbonates will be much more than carbonates. In addition, the following reactions may also appear simultaneously in the solution.



Based on zwitterion mechanism, MEA, DGA, DEA and DIPA are applied as absorbents to capture CO₂. The chemical reactions with solvent MEA in absorber unit are considered as following:



The reactions 2-17 and 2-18 are assumed to appear in liquid film, which are kinetically controlled reactions. The reactions 2-12 to 2-16 are assumed to be in chemical equilibrium. A few problems will appear during the chemical absorption process, such as emulsion, foaming, unloading and flooding when the two fluid phases contact (Gabelman and Hwang, 1999), solvent degradation, the energy consumption of regeneration and corrosion (Davidson, 2007). With consideration to degradation problem specifically, two reaction types were identified: oxidative degradation and thermal degradation (S. B. Fredriksen¹, 2013). The oxidative degradation including two distinct ways such as autoxidation pathways and oxidation in the presence of metal ions, shown in figure 2.7. The primary products are ammonia, organic acids and aldehydes. The organic acid will form heat-stable salts, leading to a reduction in CO₂ absorption capacity. The amides are produced by reaction of acids and the amine function of

MEA or amines together with water elimination or by reaction of the amine function with formaldehyde in the presence of oxygen. Figure 2.8 shows the thermal degradation in the presence of CO₂ at high temperature in cross heat exchanger and regenerator, thereby forming large polymeric compounds during the reaction (Davis and Rochelle, 2009).

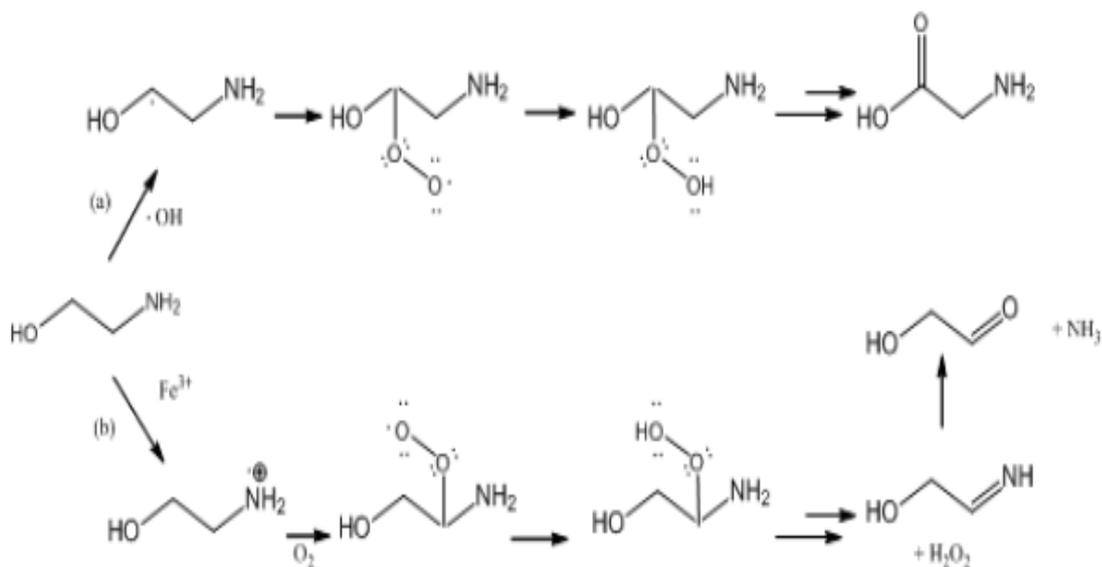


Figure 2.7: Primary oxidation pathways for MEA (S. B. Fredriksen1, 2013). (a) Autoxidation pathways; (b) Oxidation in the presence of metal ions.

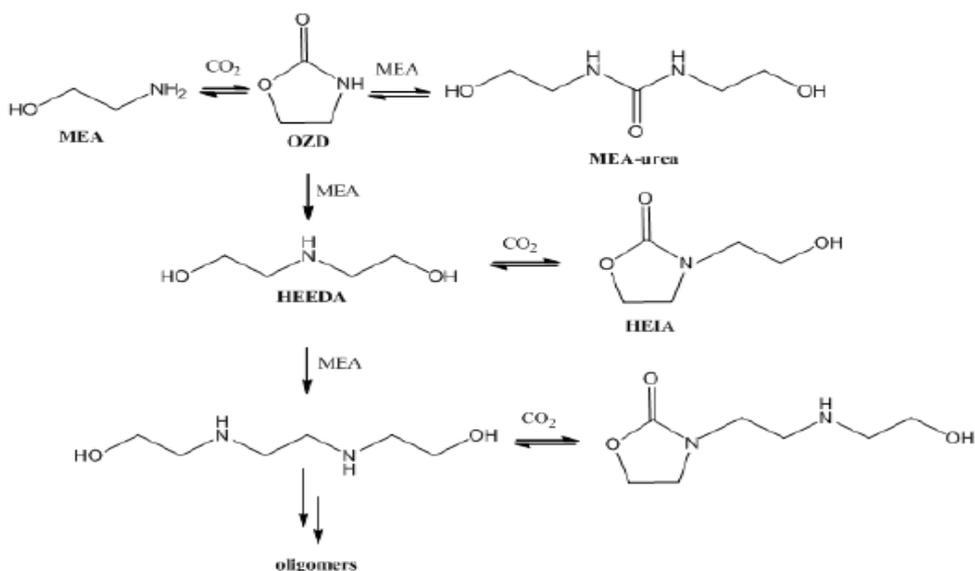


Figure 2.8 Thermal degradation pathways for MEA (S. B. Fredriksen1, 2013).

Sexton and Rochelle (2009) carried out an experiment at 55°C and discovered the dissolved metals catalyzed MEA degradation in the order copper > chromium/nickel > iron > vanadium. In addition to the catalyst, a 100:1 ratio of ethylenediaminetetraacetic acid (EDTA) to Fe is able to inhibit the oxidation of MEA. That is to say, MEA loss and oxidative degradation products cannot be observed in this condition. They also found that adding formaldehyde, formate or sodium sulfite would increase the degradation rate of MEA.

Zoannou *et al.* (2013) tested the effect of CO₂ concentration on thermal degradation at 160 °C, ranging from 0.19-0.37 mol of CO₂/mol of MEA. They found that when the initial molar CO₂-loading was 0.37, MEA loss was 20% greater than others. (Davis and Rochelle, 2009) also pointed that the amine concentration had a higher effect than CO₂ loading and MEA degradation was reduced obviously when the temperatures in stripper units were kept below 110 °C.

The post-combustion carbon capture process using DEA has been identified to have gain in power output and reduction in capital cost, due to its lower reboiler heat duty (Lee *et al.*, 2013). In other words, less steam is required to regenerate DEA solution, resulting in more steam available for power generation. The result showed that DEA has the large potential to be the best absorbent compared with MEA. (von Harbou *et al.*, 2013) have explored a new solvent of secondary amine purchased from Sigma-Aldrich (≥98%) and compare its reboiler heat duty with MEA. They found that the optimal reboiler heat duty was reduced by 16% and the numbers for optimum L/G (liquid to gas) ratio was lower than for MEA by 37%, while the overall mass transfer kinetic was similar with that of MEA.

(2) Termolecular mechanism

Termolecular mechanism was firstly suggested by Crooks and Donnellan (1989), which is considered as the limiting case of Zwitterion mechanism for $k_{-1} \geq k_B$. This is a single-step reaction, producing a loosely-bound encounter complexes instead of a zwitterion. The reaction can be represented as follow:



A large fraction of complexes are broken up to form reagent molecules again, while the remaining is reacting with second molecule of amine or water molecule to form ionic products.

(3) Base-Catalysed Hydration Mechanism

The tertiary amine, excluding from hydrogen atoms, cannot react with CO₂ directly (Donaldson and Nguyen, 1980). However, amines are usually dissolved in aqueous solution as weak bases, which is protonated to combine with hydrogen ion. Simultaneously, CO₂ is combined with hydroxyl ion. The reaction is believed as follows:

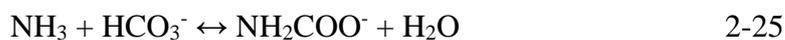
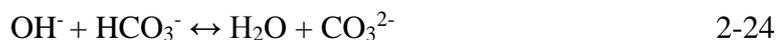


If the pH value is higher than 12, a direct reaction between tertiary amine and CO₂ may still take place and then monoalkylcarbonate will be formed. The typical tertiary amines, such as MDEA and DEMEA, are recently studied. Compared to MEA, the typical tertiary amines have low reactivity due to their own characteristics (Davidson, 2007). However, they need lower energy for regeneration and the capacity of absorbing CO₂ is higher. Furthermore, the tertiary amines have low level of degradation and corrosion (Kenarsari *et al.*, 2013).

Different mechanisms lead to the discrepancies of the reaction rate coefficients for CO₂ absorption. Aroonwilas and Veawab (2007) have used blended MEA-MDEA with appropriate mixing ratio in the simulation of the integration of CO₂ capture unit. It was found that the energy penalty was reduced ranging from 6% - 12% and more CO₂ captured per energy penalty. However, the application of MEA-MDEA should require much more capital cost, as a result of taller absorber and regenerator. Vaidya and Kenig (2007) has investigated the effects of amine-blend solvents on CO₂ reaction, and found that, the reaction of CO₂ with tertiary and sterically hindered amines (MDEA, AMP) was promoted by the addition of MEA, DEA and PZ, because MEA, DEA and PZ can facilitate zwitterion deprotonation. The amine-blend solvents have been further studied by Adewale Adeosuna (2013). They have tested DEA-AMP, DEA-MDEA, MEA-AMP and MEA-MDEA at different mixing ratio, while all the total amine blends concentration was kept at the value of 30wt%. The result shows that MEA require more energy than DEA when they are considered as activating agents. DEA/AMP proved to be the best blending mixture, due to its lower energy penalty. Besides, 5wt% DEA/25% AMP mixture performed better than any other mixtures, in the respects of reboiler duty, solvent rate and cycling load.

2.2.2 Aqueous ammonia

The reaction of aqueous ammonia and CO₂ is also apparently a ternary system. A number of ionic species are formed during the reaction CO₂ – NH₃ – H₂O as a result of hydrolysis, such as H⁺, OH⁻, NH₄⁺, NH₂COO⁻, HCO₃⁻ and CO₃²⁻.



The significantly lower heat is generating during the reaction than amine-based systems, as well as higher CO₂ capture capacity. The degradation problem can be neglected, so that the oxygen is allowed in the flue gas. However, it still exists some problems, for instance, it is necessary to cool the flue gas to 60 – 80 °F to enhance the CO₂ absorptivity and minimize ammonia vapour emission during the absorption stage. Besides, during regeneration step, the ammonia will be vaporized to loss at the elevated temperature.

2.2.3 Ionic liquids

Ionic liquids (ILs) have attracted much attention as absorbents for CO₂ capture, due to their favourable properties such as easy regeneration, low vapour pressure, moderate viscosity, high thermal stability, high CO₂ solubility and selectivity (de Riva *et al.*, 2017). In addition, they have suggested other criteria to select ILs, such as CO₂ mass transfer kinetics for physical absorption and influence of ILs nature in ILs regeneration. Figure 2.9 shows the commonly used anions and cations of ion liquids to chemically capture CO₂ in carbon capture process. Recently, imidazolium-based ionic liquids, consisted of anions such as BF₄⁻, PF₆⁻ and TF₂^{N-}, have been employed extensively as CO₂ is very soluble in these liquids. This is because imidazolium-based ionic liquids have their unique properties such as higher reaction velocity and selectivity, negligible vapour pressure, higher chemical and thermal stability, less expensive, high efficiency and non-inflammability. The absorption performance of an ionic liquid, 1-Butyl-3-methylimidazolium, was compared with that of MEA and it was concluded that the energy loss and cost of IL process was reduced by 16% and 11% than MEA, respectively (Shiflett *et al.*, 2010). However, as the future price of ILs most suited for CO₂ separation is increased to 10-30 €/kg (Meindersma and de Haan, 2008), the extension of the ILs lifetime has been a challenge.

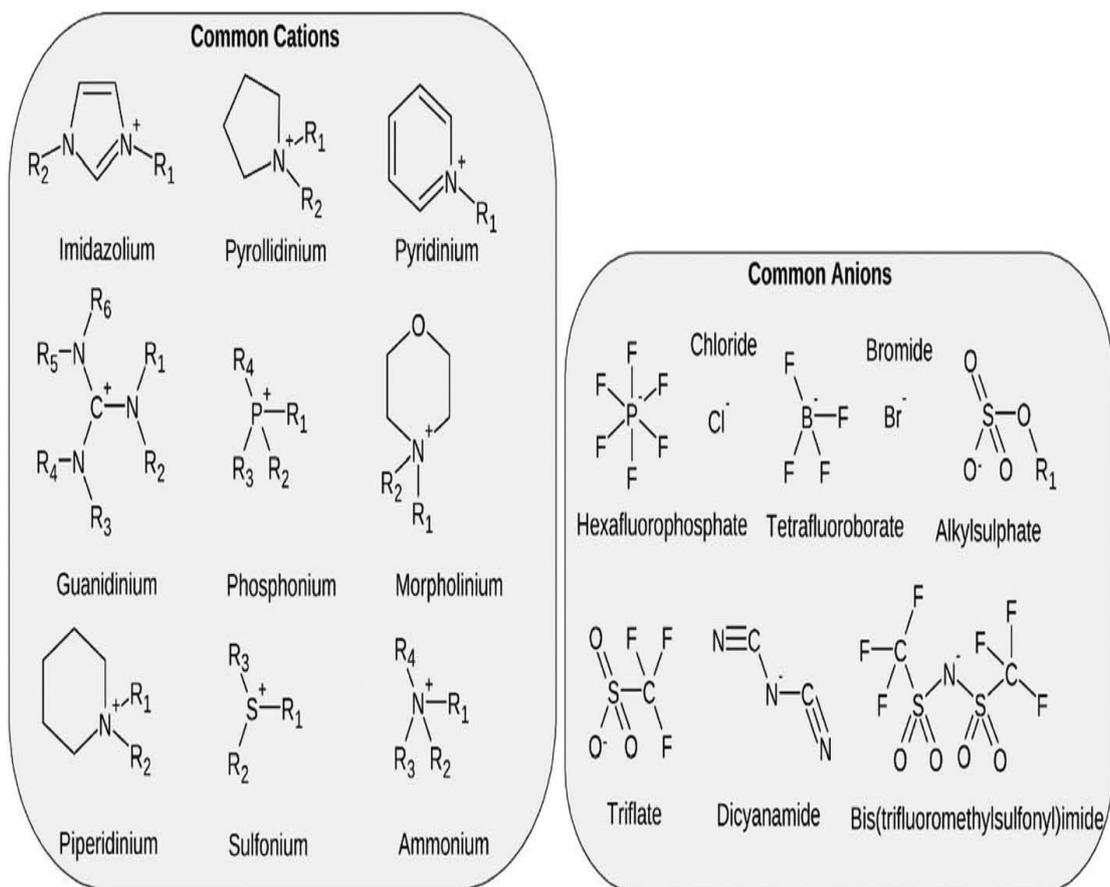


Figure 2.9: Commonly used anions and cations of ionic liquids.

2.2.4 Potassium-based solvents

The potassium-based solvents was also studied as absorbents due to their low cost, low toxicity, low solvent losses, low enthalpy requirements and high degradation resistance (Sreedhar *et al.*, 2017). With comparison to MEA, the heat of absorption by K_2CO_3 solution is only one third of MEA, so that the required regeneration energy in K_2CO_3 is less than amine based system. However, the main challenge is the low rate of mass transfer with pure K_2CO_3 . To remedy this shortcoming, the promoters such as inorganic salts, biological enzymes organics and alkaline amino acids are employed in the carbon capture process. Lee *et al.* (2006) have compared the performance of activated carbon (AC), Al_2O_3 , MgO and TiO_2 as promoters for 30 wt. % K_2CO_3 . Amongst the promoters, MgO supporting on K_2CO_3 has the highest capture capacity of 119 mg CO_2/g K_2CO_3 and capture 99.4% CO_2 . Table 2.1 shows the CO_2 capture capacity of various promoters supporting on K_2CO_3 .

Promoter	Concentration (wt %)	K ₂ CO ₃ concentration (wt %)	Temperature (K)	Acceleration*
MEA	0.5	1.8	291	0.2
	5	30	336	15
	10	30	336	45
	5	25	294	6
DEA	5	25	294	2.6
	-	-	363	4-5
	2	20	353	1.6
	2	25	323-363	-3
	5	25	323-363	-6
PZ	5	20	333	10
MDEA	5	25	294	1
Arginine	0.077 M	35	322	0.44
	0.387 M	35	322	1.35
Histidine	0.104 M	35	322	1.54
Glycine	1 M	30	333	22
Sarcosine	1 M	30	333	45
Proline	1 M	30	333	14
Carbonic Anhydrase	300 mg/L	20	298	8.8-11.3
	300 mg/L	20	313	5.2-6.4
	300 mg/L	20	323	3.4-4.0
	300 mg/L	20	313-333	2-6
	55 mg/L	30	313	0.3
	300 mg/L	20-30	298	6-20
	300 mg/L	20-30	323	2-8

*Acceleration = (Absorption rate promoted K₂CO₃ solution) / (Absorption rate in unprompted K₂CO₃ solution at same operating conditions).

Table 2.1: Summary of various promoters.

2.2.5 Summary

The ideal scrubbing solvents for post-combustion carbon capture with chemical absorption should satisfy the following requirements: minimum energy demand, a high level of CO₂ capture and minimum liquid and gas contaminant. (Sharma and Azzi, 2014). Aaron and Tsouris (2005) have concluded that the current most preferable solvents to capture CO₂ is 30 wt% aqueous MEA. This is because MEA is nontoxic and biodegradable (Shao R, 2009), as well as its cheap price and regeneration with commercial availability (Rao, 2004). In addition, the high enthalpy of MEA solution with CO₂ results in a high rates of dissolution process.

2.3 Modelling of post-combustion CO₂ capture with chemical absorption: a state-of-the-art review.

The thermal energy for MEA regeneration usually comes from extracted steam from the low pressure steam turbines, whereas it may reduce the efficiency of the coal-fired power plant. In order to optimize the process performance, the construction of an accurate model is as a priority. Since the post-combustion CO₂ capture process with chemical absorption is non-linear and cannot be solved with an easy analytical formulation, the model establishment of process performance is appeared to be of great significance. Generally, the models explored by a tremendous amount of researches can be categorized into three groups: mechanistic, regression and artificial intelligence based models. The specific details of those models will be given as below.

2.3.1 Mechanistic models

(1) Model complexity

Post-combustion CO₂ capture with chemical solvent is a reactive absorption, including two simultaneous phenomena in the process. One is mass transfer of CO₂ from the bulk vapour to the liquid solvent and the other one is chemical reaction between CO₂ and the solvent. To design an appropriate mechanistic process model, Kenig *et al.* (2001) have proposed a level of process complexity, as well as feasibility for simulation. Figure 2.10 displays the details about the different levels of modelling complexity. Specifically, model 4 is an equilibrium stage model, based on an assumption of no chemical reaction in the packed column and fast mass transfer in single stage. Model 5, which bulk phase reaction kinetics or both the bulk and film reaction kinetics are taken into account, is more accurate than model 4. Moving upwards, models 1, 2 and 3 are called rate-based models, which consider the mass transfer kinetics. In details, model 1, with the lowest level of complexity, assumes that chemical reactions are at equilibrium. It can be accurate only when CO₂ and solution reacts very fast. An enhancement factor, representing the effects of chemical reaction on mass transfer rates, is employed in model 2, to increase the model accuracy. The chemical reactions are assumed to be in liquid film while the bulk fluid remains chemical equilibrium. However, the enhancement factor can only describe the effect of a single irreversible chemical reaction on mass transfer, while is not accurate in case of several parallel and consecutive reversible reactions (Kucka *et al.*, 2003). To overcome this limitation, model 3, with a consideration of mass transfer resistances, electrolyte thermodynamics, the reaction system and column configuration, is

employed. As the process hydrodynamic is involved in model 3, it is able to relate the column outputs with operation conditions, thus making most accurate and reliable predictions.

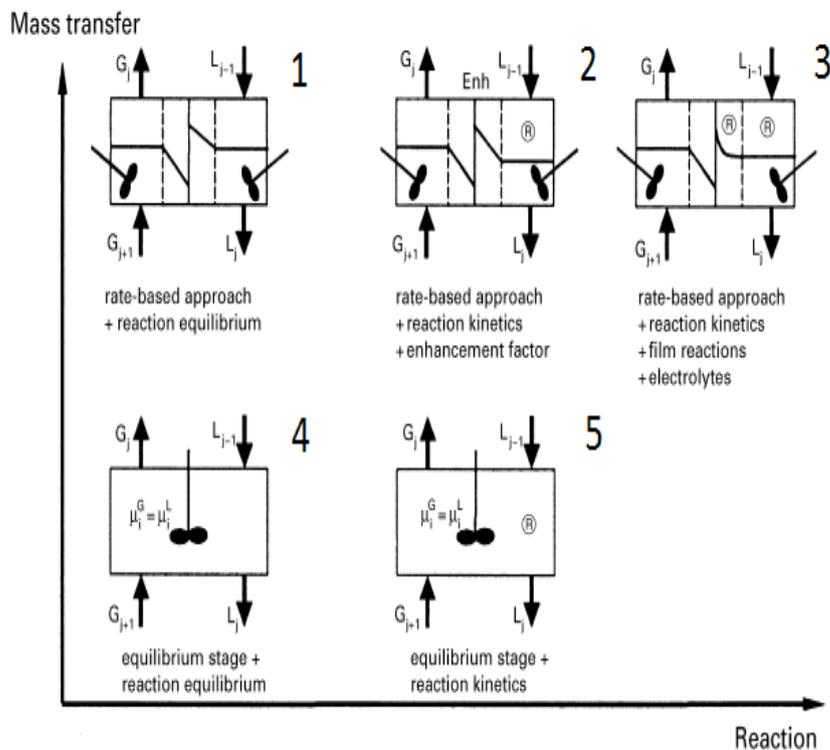


Figure 2.10: Level of complexity. Source: Kenig *et al.* (2001).

On the whole, figure 2.10 shows the models are categorized into two groups, such as equilibrium stage models and rate-based models. The significant difference between them is the consideration of interfacial mass transport by using rate-based model. Two-film theory and penetration theory are employed as using rate-based models (Wang *et al.*, 2011). As shown in figure 2.11 for two-film theory, there is an assumption that the liquid and vapour phases are both consisted of two regions: bulk and film. The effects of heat and mass transfer resistances are taken into account only in the laminar film regions. The mass transfer rates in the liquid and vapour film is estimated using Maxwell-Stefan theory. As stated by Kenig *et al.* (2001), mass transfer rate contributes a lot to reactive absorption design. The relationship between transport and reaction rate will determine where the species can react, such as in the bulk phase, or in the bulk and interfacial regions, or purely in the interfacial layers. The penetration theory assumes the exposure time between every element on surface of liquid and the vapour phase is same. The exposure time affects mass transfer coefficient significantly, because it can imply the effects of hydrodynamic properties of the system.

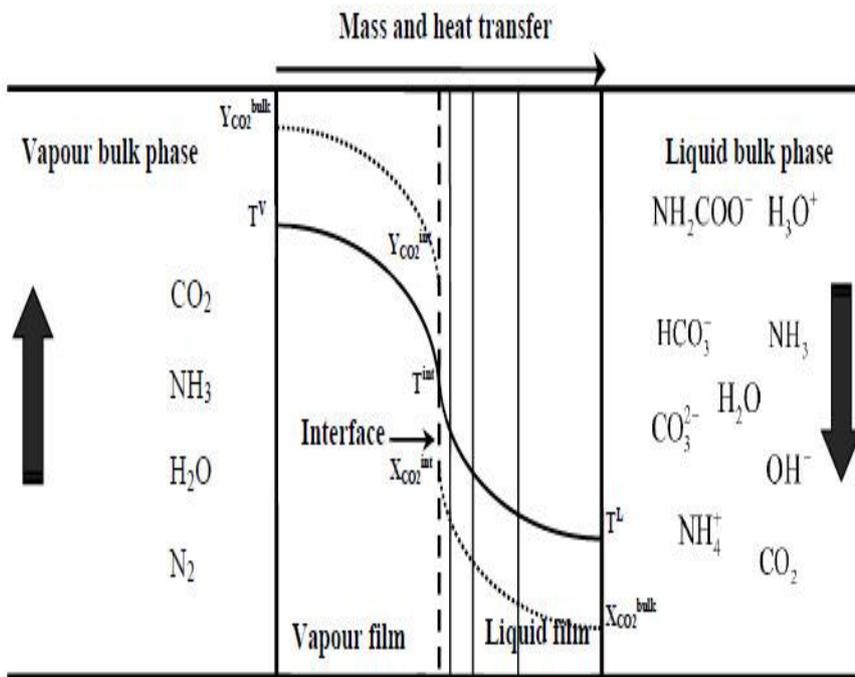


Figure 2.11: A diagram of two-film theory.

Firstly, the steady-state performance which is most often encountered in absorption units of post-combustion carbon capture plant is considered. Abu-Zahra *et al.* (2007) has developed an equilibrium-stage mathematical model (Model 5) for steady-state complete process performance in Aspen Plus. An enhancement factor was brought in the mathematic model to represent the effects of chemical reaction. The results showed that, without the consideration of corrosion and solvent degradation, the minimum thermal energy requirement (3.0 GJ/ton CO₂) is under the condition of 0.3 lean loading, 40 wt% MEA concentration and stripper operating pressure of 210 kpa. Pintola (1993) proposed a rate-based model (Model 2) for steady-state absorption process performance, thereby predicting the profiles of components concentrations and absorber temperature. The enhancement factor was calculated according to the expression by WeRek *et al.* (1978). It is worth mentioning that the variation of enhancement along the absorber column affects the liquid mass transfer coefficient significantly. The evaporation and condensation of water, the variations in physical properties and heat of chemical reaction all play a vital role to build a reliable model. A further study on modelling complete steady-state recycling process (Model 2) was implemented by Alatiqi *et al.* (1994). The enhancement factor of absorption reaction is from Decoursey and Thring (1989), while the desorption enhancement factor derives from Astarita and Savage (1980). The reason to adopt different enhancement factor expressions is that the elevated temperature in regenerator makes the reversible reactions instantaneous. Both enhancement factors were introduced to mass transfer rates through the interface to build rate-based mathematical

models. The effects of operating conditions on process performance, such as type and concentration of amines, flow rate and composition of inlet flue gas on solution circulation and reboiler duty was investigated. They found using AMP instead of MEA resulted in high partial pressure of CO₂ in outlet gas of regenerator. Besides, the increase in CO₂ loading capacity of solvent leads to a decrease in the solvent circulation rate, thereby reducing required energy. A rate-based model (Model 3) for steady-state process was further proposed with consideration of relevant reaction, diffusion kinetics and specific features of electrolyte solutions (Kucka *et al.*, 2003). Instead of enhancement concept, this model considers the precise description of the accelerating effects from chemical reactions. It is found out, the simulated data has a good agreement with the experimental data.

However, steady-state models are not sufficient to understand the impacts of post-combustion capture on the operability of the power plant. For instance, what is the response of post-combustion capture plant when the power plant is operating with a varying load? Will modifications (flooding and higher pressure drop) occur during transient conditions, such as start-up and shutdown procedures? What is the effect of heat integration between power plant and capture plant on their operation (Kvamsdal *et al.*, 2009)? Therefore, the dynamic modelling is considered as next step towards the deep process analysis.

The dynamic equilibrium-based (Model 4) and rate-based models (Model 1) were developed with an assumption of chemical reaction equilibrium (Lawal *et al.*, 2009a). With respect to testing the dynamic performance, the scenario of reducing power plant load and increasing lean MEA solution loading is regarded as disturbances, and the parameters of liquid to gas ratio is adjusted. From their results, it was found that the rate-based model gave better performance than the equilibrium-based approach. However, at the final section of research, they also suggested that the rate-based mass transfer model would be improved by considering kinetic and equilibrium reactions in liquid film. Following that, Kvamsdal *et al.* (2009) has developed a dynamic rate-based model (Model 2) for absorber by taking an enhancement factor into account to represent liquid film reactions. The prior to adopt the enhancement factor in the expression is that the physical mass transfer coefficient is large enough. Two different transient operation scenarios were performed to demonstrate the dynamic model, such as start-up and load change in an upstream power plant. The dynamic analysis of regeneration process is also implemented by Ziaii *et al.* (2009). The dynamic rate-based model (Model 1) of regeneration with reaction equilibrium was created in ACM. The dynamic behaviours, such as reducing reboiler steam rate with and without adjusting the rich solvent rate. By adjusting the ratio of rich solvent rate to steam rate, the lean loading and

temperature remained constant, as well as less response time for the system. They suggested that the residence time in the reboiler at the final steady state condition could be a dominant factor in the response time of the regeneration section. The dynamic rate-based model (Model 2) of the amine regeneration unit with an enhancement factor to represent the influence of the reactions on the CO₂ mass transfer was also developed by Mores *et al.* (2012b). The proposed mathematical model in this study was formulated as Neuro Linguistic Programming (NLP) model, which was associated with continuous variables and highly non-linear and non-convex restrictions. In this study, different correlations were used in the mathematical model to compute the specific area for mass transfer in different section of the stripper, which were Bravo's correlation at the bottom of the regenerator and Onda correlation from the middle to the top of the unit. The results showed that the combination of two correlations would give better predictions for temperature profiles and CO₂ loading of the stripper. It is not enough if only look at the individual unit. As absorber and regenerator are linked together with a recycle loop in capture plants, it is necessary to analyse the dynamic complete process. In this case, Lawal *et al.* (2010) has dynamic rate-based model (Model 1) for integrated columns in gPROMS, with the assumption of reaction equilibrium. With the comparison of stand-alone model, the results showed that the dynamic integrated model predicted the temperature profile better than stand-alone model. Besides, the parameter water balance was found to affect the performance of the system significantly. For instance, if water was lost by evaporation in the absorber, the CO₂ loading would increase followed by a maximum value. Simultaneously, the capture level would be initially reduced and then increased. It was also found out that the performances of absorber and regenerator were affected by molar L/G ratio, reboiler duty and CO₂ concentration of flue gas to the absorber column. In the study by Gaspar and Cormos (2011), a rate-based dynamic model (Model 2) of the complete absorber/desorber was developed to evaluate the operational challenges, with an enhancement factor involved, and model simulation was carried out using Matlab-Simulink. Two simulations of different scenarios have been performed, including changing the power plant load and decreasing the temperature of rich amine stream. The results showed that the decreased power plant load lead to an increase of CO₂ capture rate, while the decreased rich amine temperature resulted in an reduction of CO₂ capture rate. The important finding in this research is that the capture performance has a higher sensitivity to changes in rich amine temperature compared to the L/G ratio. In another study by Lawal *et al.* (2012), a rate-based model (Model 1) was developed to analyse two dynamic cases, including reducing power plant loading and increasing capture level set point to 95%. They summarized that the CO₂ capture plant had a slower response than power plant. It was further explored how capture level affects the power

plant loading and difficulties to achieve a steady power plant output quickly. Three further cases were considered by Biliyok *et al.* (2012a): a conventional capture process with a step decrease in lean amine flow rate into the absorber, an intercooled process with falling flue gas flow rate into the absorber and an intercooled process with a step decrease in the return temperature of the intercooled solvent. The rate-based dynamic model was developed, assuming all chemical reactions were at equilibrium (Model 1). It was observed that a higher moisture content affected capture level a little, while influenced the temperature profile significantly. When the intercooler is located close to the temperature bulge, it will improve the absorber performance a lot. This means the mass transfer, rather than chemical kinetics, is an important factor for chemical absorption CO₂ capture with MEA.

All these mechanistic models, based on chemical-, fluid mechanic- and thermodynamic laws, require extensive knowledge and underlying physics of the process. Even though they can provide advanced features such as customizing component models for the application in hand, there are still limitations to carry out complicated simulations. On the one hand, it is difficult to identify which underlying theory and assumption result in the rising uncertainties of the simulation model. On the other hand, the solution of these simulators is very complex and time consuming. Therefore, the other two types of model techniques such as multiple regression models and artificial intelligence based models are taken into account.

2.3.2 Regression model

Recently, Zhou *et al.* (2009) have proposed a multiple linear regression (MLR) model of carbon dioxide capture process, in which the outcome variable is predicted from the combination of all the input variables multiplied by their respective coefficients.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon_i \quad 2-26$$

where Y is outcome variable, β_n is the coefficient of n th input variable X_n , and ε_i is the difference between predicted and actual values of Y .

Prior to statistical analysis, they suggested four assumptions: (1) randomly distributed residuals, (2) normal distribution of residuals, (3) respective linear relation between each input variable and predicted variable, and (4) non-multicollinearity between the input variables. The reasons of non-multicollinearity between the input variables are explained as follows: (1) if the input variables are highly correlated, it is difficult to identify the importance of an input variable; (2) the variance of regression coefficients will be increased within multicollinearity. Under this circumstance, the removal of correlated input variables appears

to be important before the embellishment of regression models. In their study, the establishment of regression model is divided into three steps: correlation analysis, regression analysis and model assessment. Firstly, the correlation between each input variable and predicted variable is identified as Eq.2-27.

$$r_{X_i/Y_j} = \frac{n \sum_{k=1}^n (X_i)_k (Y_j)_k - \sum_{k=1}^n (X_i)_k \sum_{k=1}^n (Y_j)_k}{\sqrt{n \sum_{k=1}^n (X_i)_k^2 - (\sum_{k=1}^n (X_i)_k)^2} \sqrt{n \sum_{k=1}^n (Y_j)_k^2 - (\sum_{k=1}^n (Y_j)_k)^2}} \quad 2-27$$

where r_{X_i/Y_j} is correlation coefficient between i th input variable X_i and j th outcome variable Y_j , which in the range of 0 and 1. If the value is close to 1, the input variable and outcome variable are highly correlated. However, if the correlation coefficient value is small, the input variable should be removed for more accuracy.

Following the first step, the regression model is conducted according to Eq. 2-26. The indicators of R , R^2 and adjusted R^2 are significant to analyse the regression model. R represent how much the combination of input variables correlates with the outcome variables, while R^2 indicates the proportion of variance in outcome variables that is explained by combined input variables. The adjusted R^2 represents the accuracy of model across different samples. In addition, F -ratio is an important indicator to identify whether the regression model is good or not. The expression is seem as follows:

$$F = \frac{R^2 / k^2}{(1-R^2)/(n-k-1)} \quad 2-28$$

where R is multiple correlation coefficient, k is the number of input variables, n is the number of samples. F can explicitly indicates what extent of accuracy the model has improved. The large values of F means a good established model. t test is also adopted to analyse the regression model, as it can demonstrate the contribution of each input variable to the predicted variable.

$$t = \frac{\beta_t}{S_{\beta_t}} \quad 2-29$$

where β_t is the regression coefficient and S_{β_t} is the standard error of the respective coefficients. The large value of t means that input variable contributes a lot to predicted variable.

In the study by Zhou *et al.* (2009), CO₂ production rate, heat duty, CO₂ absorption efficiency and CO₂ lean loading were selected as consequent variables, while the reboiler pressure, steam pressure to reboiler, steam rate to reboiler, amine circulation rate, amine concentration,

absorber off-gas were considered as predictor parameters. The correlation between consequent variables of CO₂ absorption efficiency and CO₂ production rate and the two predictors of reboiler pressure and amine concentration was not showed in the result, which was different from the operators' knowledge. The advantage of the statistical model is analysing and modelling the relationships among parameters without understanding the theoretical relationships. However, it is unable to represent the non-linear relationships among the parameters and the selection of input variables strongly relies on the experts' knowledge.

2.3.3 Artificial intelligence based models

To overcome the shortcomings of mechanistic and statistical models, artificial intelligence (AI) based model techniques are employed to assess the process performance. In this section, various AI based models, such as artificial neural networks, neuro-fuzzy and fuzzy logic, are briefly introduced with their important mathematical aspects and previous applications to post-combustion carbon capture process.

(1) Artificial neural network (ANN)

Unlike regression model, ANN model requires no pre-assumption and relationship between predicted and response variables. In other words, the built-up process of ANN model appears to be more convenient and efficient. The foremost character is to map non-linear systems with high interpolation capacity. Among various types of ANNs, one of the simplest and most widely used ANN models is single hidden-layer feedforward neural networks (SHLFNNs), in which data moves forward through all networks in only one direction. Figure 2.12 demonstrates the structure of SHLFNNs. It is consisted of three layers: input layer, hidden layer and output layer. The input layer is responsible for accepting input signals, while the output layer is exporting the outcomes. The hidden layer contains sufficient hidden nodes with activation functions, which converts a neuron's weighted input to its output activation. In details, for N arbitrary distinct samples (x_j, t_j) , $j = 1, \dots, N$, where $x_j = [x_{j1}, x_{j2}, \dots, x_{jn}]^T \in R^n$ is a vector of network inputs and $t_j = [t_{j1}, t_{j2}, \dots, t_{jm}]^T \in R^m$ is a vector of the target values of network outputs. The output of a standard SLFNs, $o_j = [o_{j1}, o_{j2}, \dots, o_{jm}]^T \in R^m$ with \tilde{N} hidden nodes and activation function $g(x)$ is shown in the following equation:

$$o_j = \sum_{i=1}^{\tilde{N}} \beta_i g_i(w_i \cdot x_j + b_i), \quad j = 1, \dots, N \quad 2-30$$

where $w_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ is a vector of the weights between the i th hidden node and the input nodes, b_i is the bias of the i th hidden nodes, x_j is the j th input sample, $\beta_i \in R^m$ is the

weight linking the i th hidden node and the output node. The output node is chosen to have linear activation function in this paper.

In theory, the standard SLFNs can approximate any continuous nonlinear functions with small error, which means $\sum_{j=1}^{\tilde{N}} ||o_j - t_j|| = \varepsilon_j$. Specifically, there exists β_i , w_i and b_i to make:

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i(w_i \cdot x_j + b_i) - t_j = \varepsilon_j, \quad 2-31$$

Since the error is pretty small, it can be assumed as zero when train the parameters. Therefore,

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i(w_i \cdot x_j + b_i) = t_j, \quad j = 1, \dots, N \quad 2-32$$

The above equation can be written as $\mathbf{H}\boldsymbol{\beta}=\mathbf{T}$, where:

$$\mathbf{H}(w_1, \dots, w_{\tilde{N}}, b_1, \dots, b_{\tilde{N}}, x_1, \dots, x_N) = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_{\tilde{N}} \cdot x_1 + b_{\tilde{N}}) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_{\tilde{N}} \cdot x_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}} \quad 2-33$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m} \quad \text{and} \quad \mathbf{T} = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m} \quad 2-34$$

In the above equations, \mathbf{H} is called hidden layer output matrix of the neural network and the i th column of \mathbf{H} is the i th hidden node output with respect to inputs x_1, x_2, \dots, x_N . Training of SLFNs can be done through finding the minimum value of $E = \min \|\mathbf{H}_{N \times \tilde{N}} \boldsymbol{\beta}_{\tilde{N} \times m} - \mathbf{T}_{N \times m}\|$.

SLFNs are usually trained based on a cost or error function, which is normally the least mean square error (MSE) function of $\|\mathbf{H}_{N \times \tilde{N}} \boldsymbol{\beta}_{\tilde{N} \times m} - \mathbf{T}_{N \times m}\|$. There are two advantages for MSE: It is the most appropriate cost function to be incorporated with the learning algorithm and the learning rate is very fast. The MSE expression is as follows:

$$\text{MSE} = \frac{1}{N \times m} \sum \sum \|\mathbf{H}_{N \times \tilde{N}} \boldsymbol{\beta}_{\tilde{N} \times m} - \mathbf{T}_{N \times m}\|^2 \quad 2-35$$

The update process of weights and bias is generally based on back propagation (BP) algorithms, which typically need many iterations and typically slow. BP algorithm repeats two phase cycle: propagation of error signal and parameter update. It is used in conjunction with various optimized methods, such as gradient decent, Gaussian-Newton method and Levenberg-Marquardt (LM) algorithm. From the structure perspective, ANN is particularly sensitive to the number of hidden neurons. Too few hidden neurons could possibly result in

under-fitting, while too many neurons may lead to overfitting. In the situation of overfitting, the model is unable to predict unknown data accurately, even though the training error is small. To avoid this, three methods, for instance training with Bayesian regulation, a modification of Levenberg-Marquardt (LM) algorithm and early stopping are normally employed in ANN training routines.

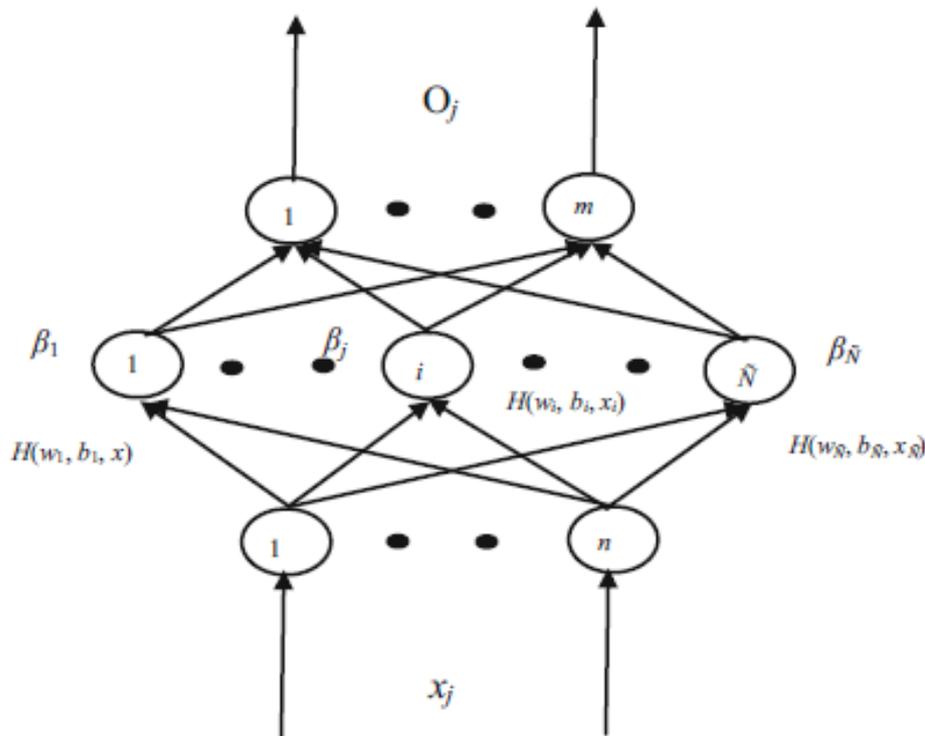


Figure 2.12: The schematic structure of SHLFNNs.

As to exploration of ANN models on post-combustion CO₂ capture process, Wu *et al.* (2010) has proposed a feedforward ANN model and compare it with regression model. The ANN model was based on gradient decent algorithm. To compare ANN model with regression model, an indicator of R-value was introduced, which represents the amount of variation in the consequent variable that is accounted for by the model. It has shown that ANN model has higher accuracy than regression model. However, the standard back-propagation with gradient decent method still exits some limitations, such as slow convergence and possibility to fall into local minima. Sipocz *et al.* (2011) has developed an ANN model in conjunction with two other training algorithms of Scaled Conjugate Gradient (SGC) and the Levenberg-Marquardt (LM) on the complete recycle process. The data was collected from CO2SIM simulator and divided into three groups: training data, validation data

and testing data. The results have shown that ANN trained with LM algorithm has better prediction accuracy than SGC algorithm. Furthermore, compared to rigorous rate-based model, it was also proved that ANN model is able to predict the outcomes faster with high accuracy.

(2) Neuro-fuzzy technology

As ANN model has no ability to interpret the relationship between input variables and response variables, Zhou *et al.* (2010a) has further proposed a new technology on post-combustion CO₂ capture process, called adaptive-network-based fuzzy inference system (ANFIS). Like feedforward back propagation ANN model, the consequent variables in ANFIS is also calculated forward and the parameters are updated backward. However, it has more layers than back propagation ANN model. Figure 2.13 indicates the structure of ANFIS contains 5 layers, namely, a fuzzy layer, a product layer, a normalized layer, a defuzzy layer and a total output layer. Specifically, Layer 1 is the fuzzification layer, with the function as follows:

$$O_{1,i} = \mu A_i(m), \quad i = 1, 2 \quad 2-36$$

$$O_{1,j} = \mu B_j(n), \quad j = 1, 2 \quad 2-37$$

where m, n are inputs to Layer 1, A_i and B_i are the linguistic labels associated to inputs m and n , $O_{1,i}$ and $O_{1,j}$ represent the degree to which the inputs m and n are related to A_i and B_j , respectively. The node functions used in this layer are normally determined by the type of the membership function, such as triangular, trapezoid and Gaussian functions. In this study, Gaussian function was selected, as it can demonstrate response of output variables sensitively when input variables varies. The expression of Gaussian function is as follows:

$$\mu A_i(m) = \exp[-(m - c_i)^2/a_i^2] \quad 2-38$$

$$\mu B_j(n) = \exp[-(n - c_j)^2/a_j^2] \quad 2-39$$

where c_i and a_i denote the centre and width of the Gaussian function, respectively. The values of parameter set (c_i, a_i) will be tuned during the learning process.

Layer 2 is a production layer, which multiplies the outputs of previous layer and send them out to layer 3. The node function is expressed as follows:

$$O_{2,i} = w_i = \mu A_i(m) \times \mu B_j(n) \quad 2-40$$

Layer 3 is the normalization layer, which indicates the ratio of i th rule's firing strength to the sum of all rule's firing strength. The output of Layer 3 is coming up with follow expression.

$$O_{3, i} = \hat{w}_i = \frac{w_i}{w_1 + w_2} \quad 2-41$$

Layer 4 is a defuzzification layer, in which two fuzzy rules are applied to fuzzy sets to obtain output variables.

$$\text{Rule 1: if } m \text{ is } A_1 \text{ and } n \text{ is } B_1, \text{ then } f_1 = p_1 m + q_1 n + r_1 \quad 2-42$$

$$\text{Rule 2: if } m \text{ is } A_2 \text{ and } n \text{ is } B_2, \text{ then } f_2 = p_2 m + q_2 n + r_2 \quad 2-43$$

where p_1, p_2, q_1 and q_2 are linear parameters, and A_1, A_2, B_1 and B_2 are nonlinear parameters.

Layer 5 is the total output layer, which sums all the inputs and computes the overall outputs.

$$O_{5, i} = \sum_i \hat{w}_i f_i = \frac{\sum_i \hat{w}_i f_i}{\sum_i \hat{w}_i} \quad i=1, 2 \quad 2-44$$

The learning rule adopted in this research is a hybrid learning algorithm, which can decrease the time consumption of learning process with comparison of gradient decent algorithm.

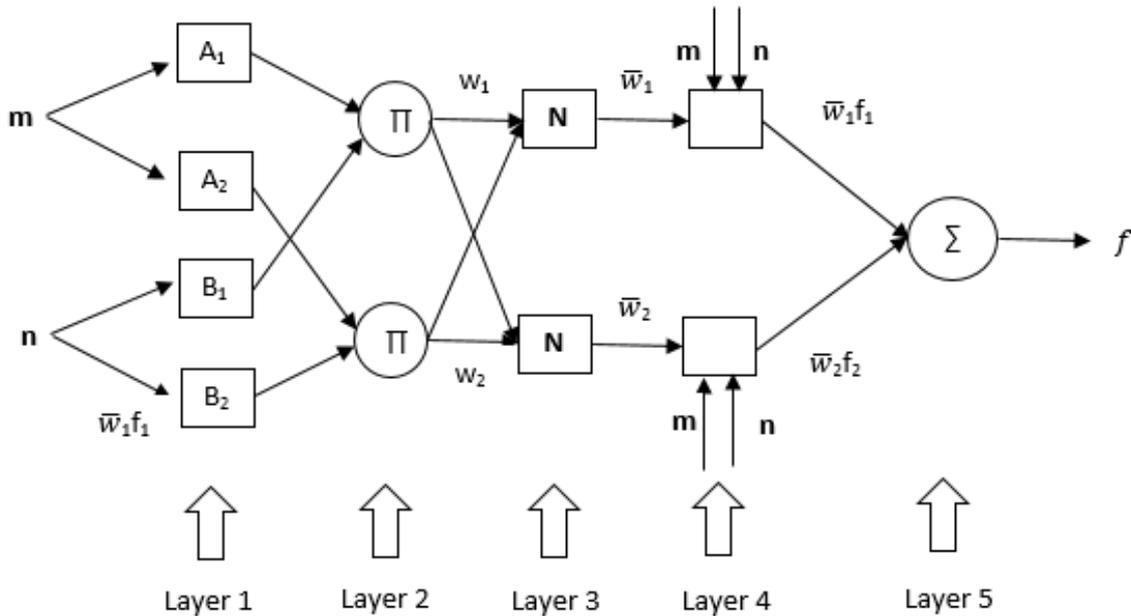


Figure 2.13: Architecture of ANFIS.

From the results of predicting output variables, such as CO_2 production rate, heat duty, absorption efficiency and lean loading, it appears high accuracy. Specifically, in predicting CO_2 production rate, 120 out of 150 tuples of data have more than 95% accuracy. 112 out of

150 tuples of data are accurate over 95% in predicting heat duty. In the prediction of absorption efficiency, 111 out of 150 tuples of data have accuracies more than 95%. Lastly, in modelling lean loading, 69 out of 150 tuples of data have accuracies over 95%. Later, they compared the performance of ANFIS with that of feedforward back propagation ANN model, and concluded that ANFIS had average higher accuracy for predicting dependent variables (Zhou *et al.*, 2010b).

2.4 Process Control

As the upstream power plant is usually operating with a varying load or transient conditions, the flexible operation of PCC plant plays a vital role to cope with the dynamic loading. Hence, the strategies of control systems design of PCC plant is of great importance. In recent years, a number of studies have focused on control system design of MEA-based PCC process. The process controllability analysis is mainly consisted of several steps explained in figure 2.14. Firstly, the process control goals are specified. Secondly, sensitivity analysis is used to identify the manipulated variables (MVs) and controlled variables (CVs). Then a suitable control scheme is selected. The control schemes are mainly categorized into two groups: decentralized configuration (PI, PID controller) and centralized configuration (MPC controller). MPC-based control structure can achieve the goal of online tracking and estimation. Next, as to centralized control scheme, a model between CVs and MVs is established, while the pairs selection between CVs and MVs are determined regarding to decentralized configuration. Further, the parameters for each control scheme are adjusted. Lastly, the accomplished control scheme is evaluated with its performance.

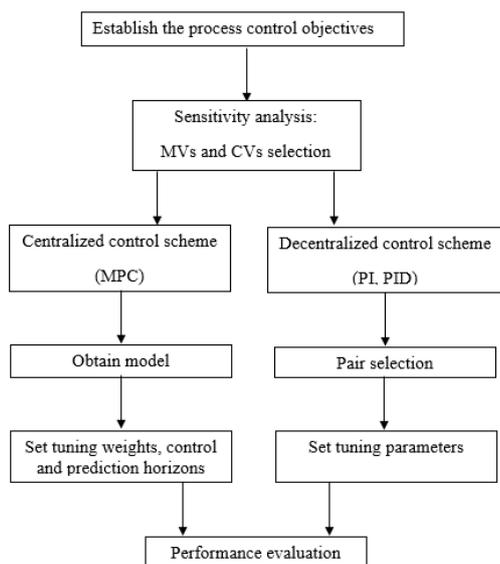


Figure 2.14: Flow chart of controllability analysis.

2.4.1 Decentralized Control strategy

It is estimated that only 5%-10% of control loops cannot be controlled by single-input single-output (SISO) controllers (Koivo and Tantt, 1991).

2.4.1.1 Pair selection between MVs and CVs

As to the decentralized control strategy, the main challenge is the design of single input single output (SISO) control loop that accounts for the interaction among the other control loops. Therefore, the first step is to determine SISO control loops through the manipulated and controlled variable pair selection, thereby minimising the interaction between SISO loops and achieving an acceptable control performance.

The relative gain array (RGA), introduced by Bristol (1966), has been widely used for over 50 years, especially after the issue of closed loop stability was resolved by using Niederlinski Index (NI) as a stability criteria. To measure the interaction between each control loop, the process steady state gain is used. For a 2×2 process gain matrix with elements K_{ij} , the RGA is calculated as follows:

$$\text{RGA} = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{22} & \lambda_{21} \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 1 - \lambda_{11} \\ 1 - \lambda_{11} & \lambda_{11} \end{bmatrix} \quad 2-45$$

where

$$\lambda_{11} = \frac{1}{1 - \frac{K_{12}K_{21}}{K_{11}K_{22}}} \quad 2-46$$

In practice, there may be some uncertainties in system, due to the process model mismatch, operating condition variation, and drift of physical conditions or parameters. The uncertainty bounds, including lower and upper bounds of λ_{ij} , should be considered on RGA analysis. To calculate them, Chen and Seborg (2002) have introduced an analytical expression for uncertainty bounds of steady state gains as below:

$$A.X \leq b \quad 2-47$$

where A is an appropriate matrix of size $(2n^2) \times (n^2)$ satisfying Eq. 2.47, b is a vector of size $(2n^2) \times 1$ containing the lower and upper bounds of K, and X is a vector of size $n^2 \times 1$ containing all elements of K as its elements:

$$X = [K_{11} \ \dots \ K_{nn}]^T \quad 2-48$$

Therefore, the lower and upper bounds of λ_{ij} can be expressed as:

$$\text{Lower bounds: } \min_X \lambda_{ij} = f(X) \quad 2-49$$

$$\text{Upper bounds: } \max_X \lambda_{ij} = f(X) \quad 2-50$$

However, RGA is considered to have some deficiencies as it does not take process disturbance and dynamic into account. Stanley *et al.* (1985) has proposed a method called relative disturbance gain (RDG), which considers the process disturbance. Specifically, the multivariate process is expressed as Eq. 2.51

$$y = Gu + G_d d \quad 2-51$$

where y is a vector of controlled variables, u is a vector of manipulated variables, d is a disturbance, G is the process gain matrix, and G_d is the disturbance gain matrix. Hence, the i th element of RDG is denoted as:

$$\beta_i = \frac{\left[\frac{\partial u_i}{\partial d} \right] y_j}{\left[\frac{\partial u_i}{\partial d} \right] y_j, u_j, j \neq i} \quad 2-52$$

The vector of RDG can be arranged as:

$$\text{RDG}\{G, G_{diag}, G_d\} = (G^{-1}G_d) \div (G_{diag})^{-1}G_d \quad 2-53$$

where \div denotes element by element and G_{diag} defines a diagonal matrix of G .

According to the above formulas, relative disturbance gain array (RDGA) is expressed as follows:

$$B = [G^{-1}diagG_d]^{-1}[diag(G^{-1}G_d)] \quad 2-54$$

Agustriyanto and Zhang (2007) have also come up with lower and upper bounds of RDGA elements for uncertainty process models. In details, assume an $n \times n$ system containing steady state gain and disturbance gain:

$$K = [K_{ij}]_{n \times n} \quad 2-55$$

$$K_d = [K_{dj}]_{n \times 1} \quad 2-56$$

The relationship between RDGA and RGA is:

$$\beta_{ij} = \lambda_{ij} + \sum_{k=1, k \neq i}^n \frac{K_{ij} \dot{K}_{jk} K_{dk}}{K_{di}} \quad 2-57$$

where \dot{K}_{jk} is the jk th element of K^{-1} , β_{ij} is a function of K and K_d .

According to Eq. 2.57, for RDGA, A is an appropriate matrix of size $(2(n^2 + n)) \times (n^2 + n)$ satisfying the inequalities, b is a vector of size $(2(n^2 + n)) \times 1$ consisted of lower and upper bounds of X , and X is a vector of size $(n^2 + n) \times 1$ containing elements of K and K_d as

$$X = [K_{11} \ \dots \ K_{nn} \ K_{d1} \ \dots \ K_{dn}]^T \quad 2-58$$

Hence, the expression of lower and upper bounds of β_{ij} is:

$$\text{Lower bounds: } \min_X \beta_{ij} = f(X) \quad 2-59$$

$$\text{Upper bounds: } \max_X \beta_{ij} = f(X) \quad 2-60$$

To overcome the limitation of dynamic by RGA rule, a new approach called dynamic relative gain array (DRGA) was firstly proposed, in which the demonstrator of DRGA achieve perfect control at all frequencies while the numerator was only the open loop transfer function (Witcher and Mcavoy, 1977). Mc Avoy *et al.* (2003) also published an approach to calculate DRGA, which was based on the proportional output optimal controller gain matrix. In this research, they assumed a linear state space process model as:

$$dx/dt = Ax + Bu \quad 2-61$$

$$y = Cx \quad 2-62$$

where y , u is denoted as the measurements and manipulated variable respectively. The first step is to scale y and u to \tilde{y} and \tilde{u} by their operating ranges or steady state values. The controller gain K is calculated based on dynamic model of the process, which means the dynamic information is involved. Therefore, the i,j th element of the DRGA is defined as:

$$\lambda_{Dij} = \frac{\partial u_i / \partial y_i \big|_{u_j \neq 0, k \neq i}}{\partial u_i / \partial y_i \big|_{u_j = 0, k \neq i}} \quad 2-63$$

The numerator gives the change in manipulated variable to the change in controlled variable, in the case where the optimal controller is bringing the system back to the origin starting from a random initial state on the unit sphere. The demonstrator is obtained by optimal controller gain matrix. Eq. 2.63 can be transformed into as follows

$$\lambda_{Dij} = \frac{-K_{ij}}{-1/K_{ji}} = K_{ij}K_{ji} \quad 2-64$$

The rule to select best pairs of MVs and CVs by using RGA, RDGA and DRGA methods is the value of elements in these matrix is positive and close to 1.

2.4.1.2 Controller tuning

PID controller is the most widely used controller for SISO control loop today. Theoretically, the PID controller is continuously calculate error $e(t)$ as the difference between a desired set point and a measured process variable and applies a correction based on proportional (P), integral (I) and derivatives terms (D) (Eq. 2.65). This error signal $e(t)$ will be sent to the PID controller, and the controller computes both the derivative and the integral of the error signal. The control signal $u(t)$ to the plant is equal to the proportional gain K_p times the magnitudes of the error plus the integral gain K_i times the integral of the error plus the derivatives gain K_d times the derivatives of the error.

$$u(t) = K_p e(t) + K_i \int_0^t e(t) dt + K_d \frac{de(t)}{dt} \quad 2-65$$

This error is simply multiplied by one, two or all of the calculated P, I and D actions. Hence, there are three modes used in different combinations, such as proportional (P) controller, proportional integral (PI) controller and proportional integral derivative (PID) controller. Eq.2.65 can be transformed to an expression of Laplace domain for P, PI, and PID controllers as follows:

$$U(s) = G_c(s) E(s) \quad 2-66$$

where $G_c(s)$ is PID controller transfer function

A simplified block diagram of SISO feedback control system is shown in figure 2.15. y_{sp} is the desired output, while y is the controlled output. u is the controller output/manipulated input, and e is the difference between y_{sp} and y . the load difference is entered into the process input, and the feedback signal is corrupted by random measurement noise n at the process output. $G_p(s)$ is representing the process dynamic function. This is usually a first-order plus dead time (FOPDT) function model (Eq. 2.70), as high order and even slightly nonlinear behavior in experimental data are often represented by FOPDT model to facilitate the controller design.

$$G_p(s) = \frac{K e^{-\theta s}}{\tau s + 1} \quad 2-67$$

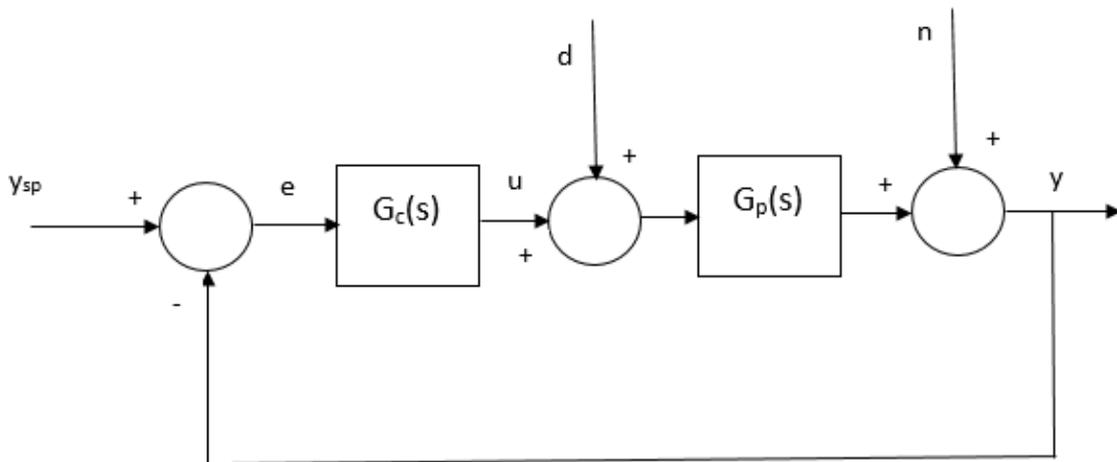


Figure 2.15: A simplified block diagram of SISO feedback control system.

The PID-based control loop is tuned by trial and error of changing one or more of proportional, integral and derivatives. Hence, controller tuning process is to find the optimal values of three parameters: K_c , τ_i , and τ_D . According to the statistics, 293 out of 408 sources of tuning rules have been reported since 1992, reflecting many attentions to the PID controller in past decades (O'Dwyer, 2009). Amongst them, the PID controller tuning rules may be classified as follows(O'Dwyer, 2003):

- Tuning rules based on measured step response.
- Tuning rules based on minimising an appropriate performance criterion.
- Tuning rules that gives a specified closed loop response.
- Robust tuning rules, with an explicit robust stability and robust performance criterion built into the design process.
- Tuning rules based on recording appropriate parameters at the ultimate frequency.

The first four tuning rules require the process model parameters while the last one does not. Several tuning methods have been applied extensively in last decades, such as process reaction curve and ultimate cycle tuning rule (Ziegler and Nichols, 1942). The outstanding point of process reaction curve tuning strategy is its requirement of only one single experimental test, while the load changes may occur to affect the process model accuracy (O'Dwyer, 2006). However, as to applying the ultimate cycle tuning rule, there are still some disadvantages. For example, the system must be generally unstable under proportional control. An attempt of several trials should be made to determine the ultimate gain. Besides,

the disturbance may result in some negative effects on process quality. The possibility of misunderstanding the limit cycle is detrimental to obtaining the parameters. Therefore, it is usually impossible to get a precise optimum settings of controllers. Some actions are taken to address these disadvantages, such as a modification of the rule and the combination with biggest modulus tuning (BLT). The flow chart for the ultimate cycle tuning procedure is shown in figure 2.16.

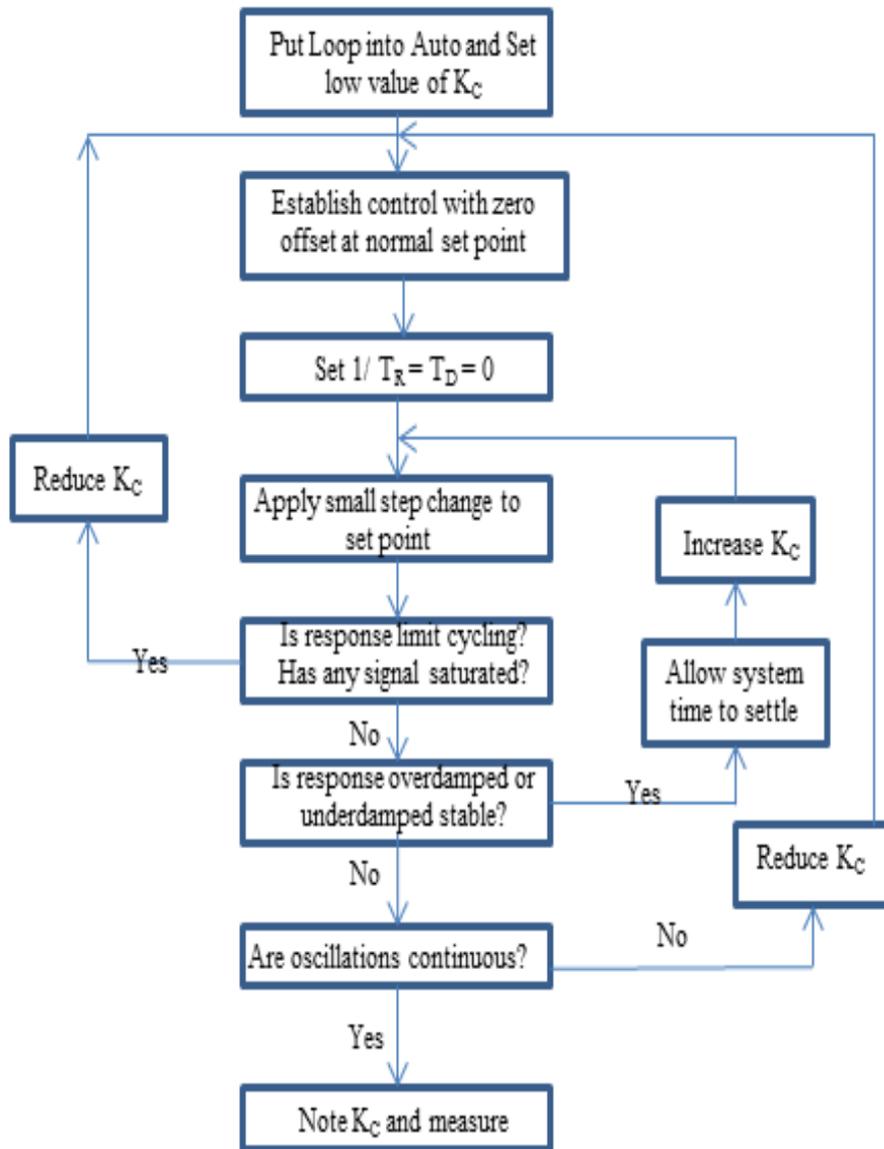


Figure 2.16: Ultimate cycle tuning procedure. Source: (Love, 2007).

2.4.1.3 An overview of decentralized control strategy on post-combustion CO₂ capture plant

Lawal *et al.* (2010) has presented a heuristic decentralized control structure of post-combustion CO₂ capture plant using PID controllers. It was revealed that CO₂ capture performance was more sensitive to L/G ratio than individual flow rate. Besides, the appropriate water balance in the absorber was shown to be of great importance. Panahi and Skogestad (2011) developed three different control structures using self-optimising method to find best CVs for three active constraints regions of the flue gas flow rate. With application of self-optimising control structure, the process will be not re-optimized when the disturbance occur. Later, Nittaya *et al.* (2014) compared the performance of three control structures based on RGA (control structure A) and heuristic approaches (control structure B and C), by using PI controllers. 6 manipulated variables and 6 controlled variables are selected for building SISO control loops (table 2.2). The details of control loops between MVs and CVs for three control structures are shown in table 2.3. Through sensitivity analysis, as V1 has faster effect on %CC than Q_{reb}, control structure B is determined. With respect to control structure C, V2 is used to control the reboiler temperature by adjusting the rich amine solution entering into the regenerator. In this case, the reboiler temperature could decrease slowly to heat the rich amine solution to set point of reboiler temperature. The response of control loop V2 - T_{reb} is faster than control loop Q_{reb} - T_{reb}. The parameters of PI controllers in control structure A and C are adjusted using Internal Model Control (IMC), while those in control structure B are tuned by process insights. It was concluded that control structure B and C have better performance than control structure A, as RGA did not consider process dynamics. Recently, Manaf *et al.* (2016) have proposed two decentralized control structures using PID controllers, which are based on RGA and Morari index of integral controllability (MIC) approaches. RGA analysis suggested capture efficiency (%CC) was controlled by lean solvent flow rate (V1) and energy performance (EP) was controlled by reboiler heat duty (Q_{reb}), which is similar to the findings by Nittaya *et al.* (2014). However, by using MIC approach, the opposite control loop, %CC- Q_{reb} and EP-V1, was determined. This control structure will not hurt the robustness and stability of existing closed loop system.

	Variable		Variable
MV1	Reboiler heat duty (Q _{reb})	CV1	Condenser temperature (T _{cond})
MV2	Condenser heat duty (Q _{cond})	CV2	lean amine stream temperature (T _{tank})

MV3	Buffer tank heat duty (Q_{tank})	CV3	Reboiler temperature (T_{reb})
MV4	Outlet valve position of the buffer tank (V1)	CV4	Percentage of CO ₂ removal (%CC)
MV5	Outlet valve position of the absorber sump tank (V2)	CV5	Liquid level in absorber sump tank (L1)
MV6	Outlet valve position of the reboiler surge tank (V3)	CV6	Liquid level in reboiler surge tank (L2)

Table 2.2: List of manipulated variables and controlled variables. Source: (Nittaya *et al.*, 2014).

	Control structure A	Control structure B	Control structure C
T_{cond}	Q_{cond}	Q_{cond}	Q_{cond}
T_{tank}	Q_{tank}	Q_{tank}	Q_{tank}
T_{reb}	V1	Q_{reb}	V2
%CC	Q_{reb}	V1	Q_{reb}
L1	V2	V2	V1
L2	V3	V3	V3

Table 2.3: Control loops of three control structures. Source: (Nittaya *et al.*, 2014).

2.4.2 Centralized control strategies

Multivariate controller such as MPC is widely used in chemical process as an advanced centralized control technique. Two attractive features such as operation of MIMO systems consistently and explicit consideration of constraints actions on controllers, were indicated by Prolss *et al.* (2011). In specific details, the future process behaviour over several future finite time intervals is optimized by MPC algorithm known as the prediction horizons. The process dynamics is expressed by linear or non-linear process models. In this case, the present and future M control actions is anticipated by process models. After taking M th control actions, the constraints are assumed to be 0. The process only accepts the first of optimal input sequence and the entire process will be repeated at each time interval. With the application of MPC control strategy, only one MPC controller is adopted. The main issue of MPC control

strategy is the appropriate choice of process model representation. According to the intrinsic nature of process dynamics, MPC controllers are categorized into two groups: linear MPC controllers (LMPC) and non-linear MPC (NMPC) controllers.

LMPC controllers have been extensively used in MPC control schemes to date. All the process linear model forms can be derived from a state space model, which linearize about an operating point:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k + \mathbf{C}\mathbf{v}_k + \mathbf{w}_k \quad 2-68$$

$$\mathbf{z}_k = \mathbf{D}\mathbf{x}_k \quad 2-69$$

$$\mathbf{y}_k = \mathbf{D}\mathbf{x}_k + \boldsymbol{\varepsilon}_k \quad 2-70$$

where \mathbf{x}_k , \mathbf{u}_k , \mathbf{v}_k and \mathbf{w}_k denote the vector of state variable, MVs, measured disturbance variables (DVs) and unmeasured DVs, respectively. $\boldsymbol{\varepsilon}_k$ is a vector of measurements noise.

According to the indication by Kailath (1980), the above discrete-time transfer function model can be written equivalently in a form of matrix fraction expression.

$$\mathbf{y}_k = [1 - \boldsymbol{\Phi}_y(q^{-1})]^{-1}[\boldsymbol{\Phi}_u(q^{-1})\mathbf{u}_k + \boldsymbol{\Phi}_v(q^{-1})\mathbf{v}_k + \boldsymbol{\Phi}_w(q^{-1})\mathbf{w}_k] + \mathfrak{z}_k \quad 2-71$$

$$\mathbf{y}_k = \boldsymbol{\Phi}_y(q^{-1})\mathbf{y}_k + \boldsymbol{\Phi}_u(q^{-1})\mathbf{u}_k + \boldsymbol{\Phi}_v(q^{-1})\mathbf{v}_k + \boldsymbol{\Phi}_w(q^{-1})\mathbf{w}_k + \delta_k \quad 2-72$$

$$\delta_k = [1 - \boldsymbol{\Phi}_y(q^{-1})] \mathfrak{z}_k \quad 2-73$$

where q^{-1} is a backward shift operator.

Besides, the Box-Jenkins model form combines the error terms together into one term $\boldsymbol{\varepsilon}_k$:

$$\mathbf{y}_k = [1 - \boldsymbol{\Phi}_y(q^{-1})]^{-1}[\boldsymbol{\Phi}_u(q^{-1})\mathbf{u}_k + \boldsymbol{\Phi}_v(q^{-1})\mathbf{v}_k] + [\boldsymbol{\theta}_\varepsilon(q^{-1})]^{-1}\boldsymbol{\Phi}_\varepsilon(q^{-1}) + \boldsymbol{\varepsilon}_k \quad 2-74$$

If the system is stable, the finite impulse response (FIR) model is expressed as an approximation to Eq. 2-75:

$$\mathbf{y}_k = \sum_{i=1}^{N_u} H_i^u \mathbf{u}_{k-i} + \sum_{i=1}^{N_v} H_i^v \mathbf{v}_{k-i} + \sum_{i=1}^{N_w} H_i^w \mathbf{w}_{k-i} + \mathfrak{z}_k \quad 2-75$$

However, in most cases, the process dynamic of chemical engineering systems is appeared to be nonlinear. A number of nonlinear modelling techniques, such as differential equations, differential-algebraic equations, discrete time algebraic descriptions, Wiener models and neural networks, etc., have been proposed with specific details in the open literature (Morari and Lee, 1999; Qin and Badgwell, 2003).

The parameters of linear and nonlinear models are estimated by minimizing the following least-squares criterion:

$$J = \sum_{k=1}^L \|y_k - y_k^m\|^2 \quad 2-76$$

where y^m is the predicted outputs.

There are two approaches to estimate the model parameters, one is called equation error approach, and the other one is called output error approach. As to the former, the past output measurements are fed back to calculate current output, while regarding to the latter, the past model output estimates are fed back to calculate current output. In other words, the former is called one-step ahead prediction and the latter is called long range prediction.

A multi-level control scheme of post-combustion CO₂ capture process based on LMPC approach was indicated by Arce *et al.* (2012), which included high-level and low-level control loops. The high-level control loop has reduced the operating cost associated with regeneration system by as much as 10%, while the low-level control loop has shown the good performance of LMPC control, with comparison to PID based control. Sahraei and Ricardez-Sandoval (2014) have introduced a multivariable LMPC scheme on post-combustion carbon capture process, with both energy and environments constraints. In their study, the linear transfer functions were firstly obtained from sensitivity analysis and then transformed into discrete linear state-space model. The prediction horizon and control horizon were valuable parameters, which were also determined by sensitivity analysis. LMPC controller tuning parameters were set by several tests. With the comparison of the performance between LMPC controller and PI controller, they have concluded that, when deal with load changes and set-point tracking, LMPC control strategy was better to maintain MVs with their feasible limits and perform faster response. Zhang *et al.* (2016) proposed a LMPC controller for CO₂ capture rate and reboiler temperature in the case of disturbances in flue gas flow rate and CO₂ composition of flue gas. In addition, they also compared the performance between LMPC and PID controller and found LMPC controller was able to avoid overshoot scenario and achieve settling time very shortly. However, a few contributions in the literature present the performance of NMPC control architecture for post-combustion CO₂ capture process. To compensate the limitations of LMPC control scheme, Akesson *et al.* (2012) have suggested a NMPC controller for online optimisation of post-combustion CO₂ capture under dynamic load change conditions. The nonlinear dynamic model applied in their study was derived from first principles and determined as differential-algebraic expressions. The results showed that

NMPC control system had fast response to changes in operation conditions of upstream power plant.

2.5 Process optimisation

To maximize profitability within a set of given constraints, process optimisation and control is required. The typical objective function is denoted as the economic model of the process, which is shown as below (Darby and White, 1988):

$$\begin{aligned} \text{Objective} = & \text{Productive value} - \text{Feed costs} - \text{Utility costs} \\ & + \text{Other variable economic effects} \end{aligned} \quad 2-77$$

The constrained optimisation problem is usually expressed as the following form of mathematical equations.

$$\min_x F(x) \quad 2-78$$

$$\text{subject to: } h(x) = 0, g(x) \leq 0 \quad 2-79$$

where $F(x)$ is linear or nonlinear objective function, h and g are defined constraints.

The optimisation problem solvation is to estimate the derivatives of objective function regarding to operating variables. In this case, the optimisation should be continuous.

Two technologies of real-time dynamic optimisation are proposed: dynamic real-time optimisation (D-RTO) working together with MPC and economic model predictive control (EMPC). The former is to send the target trajectory calculated by RTO to MPC controller, acting as two-layer architecture. The optimized operation point is obtained by a rigorous plant model in the upper layer, while the MPC controller maintains the manipulated variables as close as their set point (optimized operation point) in the lower layer. The details is specifically shown in figure 2.17. RTO system is consisted of steady-state detection, data reconciliation, process model updating, optimisation calculation and command conditioning to advanced controller (Sequeira *et al.*, 2002). Specifically, the process should be firstly detected when the steady state is reached. The data under steady state process is collected and validated to implement corrective actions to fix the gross errors found in data, as well as to ensure the consistency for model updating. Then the measurements are used to establish the model which represent the plant dynamic correctly at the current operation point. Lastly, the optimum set point of controller is calculated by optimisation algorithms and transferred to the control system. However, there are several main drawbacks with the application of RTO

system. In details, the models between RTO and MPC controller are usually mismatched, resulting in a situation that the desired operation point calculated by RTO is unreachable by feedback control layer. Wang *et al.* (2017) have presented a global RTO method without mismatch models between optimisation and control layers. The dynamic model equations were transformed into a nonlinear algebraic model by using trapezoidal formula, which was applied in both two layers. The modified normalized multi-parametric disaggregation technique (NMDT) was used to as a global optimization algorithm to solve dynamic RTO layer. The result has shown that the problems of unreachability and infeasibility of set point in control layer was not appeared. Besides, since the optimisation is complete after the process reach steady state, the computation of new optimized condition will be delayed which affect the process performance. A solution was proposed to solve the problem, which optimise the problem with a high frequency (Sequeira *et al.*, 2002). However, it will lead to a unstable closed-loop system (Engell, 2007). The two-layer strategy has been applied successfully in optimisation of distillation column system (Zhu *et al.*, 2004) and polymerization process (Pontes *et al.*, 2015).

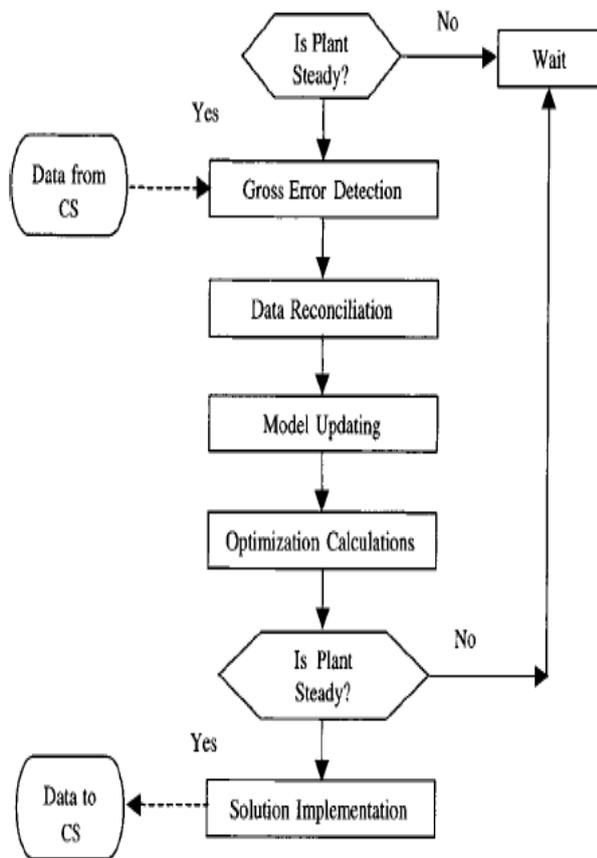


Figure 2.17: Schematic representation of the closed loop system of RTO and MPC controller. Source: Sequeira *et al.* (2002).

As explained by Ellis *et al.* (2014), EMPC approach is to integrate the economic process optimisation and MPC control into one layer, which allows detecting process improvement and operation to be consistent. The cost function for direct or indirect flexion of process economics will be incorporated into MPC formulation. It can be expressed as below:

$$\min_{u \in S(\Delta)} \int l_e(\tilde{x}(t), u(t)) dt \quad 2-80$$

$$\text{subject to } \dot{\tilde{x}} = f(\tilde{x}(t), u(t), 0) \quad 2-81$$

$$\tilde{x}(0) = x(\tau_k) \quad 2-82$$

$$g(x(t), u(t)) \leq 0, \forall t \in [0, \tau_N) \quad 2-83$$

$$g_e(x(t), u(t)) \leq 0 \quad 2-84$$

where l_e is the economic cost function, $\tilde{x}(t)$ is the open-loop predicted state trajectory, $u(t)$ is manipulated variable. Eq. 2-85 is a continuous-time, time-invariant nonlinear dynamic process. Eq. 2-86 is the initialization of state measurements, where τ_k is the sequence equivalent to $\tau_0 + k\Delta$. k is the time step of the discrete model and Δ is the sampling period. g denotes the process constraints including state and manipulated variable. Eq. 2-88 is economic-based constraints. Firstly, the current state measurement of process system $x(\tau_k)$ is used to initialize EMPC. The optimal input trajectory according to optimisation problem (Eq.2-82) over the prediction horizon $[\tau_k, \tau_{k+N}]$ is calculated by Eq.2-85 in real-time. Then the first control action is implement to controller over the period $[\tau_k, \tau_{k+1}]$. EMPC is solved repeatedly following above steps at the next sampling period. Several issues of EMPC approach should be considered, such as the feasibility of the optimisation problem and closed-loop performance using EMPC (Ellis *et al.*, 2014).

Another problem for process optimisation is considered, which in details, the process model is rarely accurate and reliable. Optimisation using an inaccurate model will possibly lead to the infeasible operations. In this case, two main optimisation technologies have been proposed for handling the model uncertainty, which are robust optimisation in the absences of measurements and adaptive optimisation in the presence of measurements respectively (Chachuat *et al.*, 2009).

The most widely used optimisation algorithm to solve the optimisation problem is Sequential Quadratic Programming (SQP) approach. With application of this approach the objective function is approximated by a quadratic function, and the constraints is evaluated by a linear

function. In this case, quadratic programming can be used recursively to find a search direction of minimizing the objective function.

A number of literatures has paid their attentions on optimisation of post-combustion CO₂ capture process. Mac Dowell and Shah (2013) has stated an optimising method for cost-optimal degree problems of CO₂ capture plant integrated with a 660MW sub-critical power plant, which use dynamic non-equilibrium model to describe the capture process. In their study, they take the trade-off between cost of CO₂ emission to atmosphere and energy consumption coming from the power plant into account. The results showed that the cost-optimal CO₂ capture rate is 95% with regard to the integration of a 660MW subcritical coal-fired power plant. Using the optimisation method is able to reduce the specific energy required per tonne of CO₂ recovered by between 10% and 25%. For the goal of minimizing energy consumption of CO₂ capture, Chu *et al.* (2016) have found the optimal height of absorber is 8 m and the optimal operating pressure for columns is the atmospheric pressure. In addition, the surface area per unit volume and the porosity of packing materials is as big as possible. According to the CO₂ capture community, it is commonly accepted that the CO₂ capture level is targeted as 90% or above. The most concerned issues of post-combustion CO₂ capture plant is the large energy consumption for regenerating the scrubbing solution in the stripper column. In terms of capital intensity, this deficiency will result in a costly operation. Hence, several articles have been contributed to optimize the operation conditions, in order to minimum the operation cost and meet the CO₂ capture target as well. For example, Rao and Rubin (2006) performed an integrated modelling framework (IECM-cs) to evaluate the performance and cost of post-combustion carbon capture plant integrated with 1000 MW and 650 MW coal-fired power plant. They examined the cost-effectiveness of PCC plant under varying CO₂ capture efficiency and found there was a nonlinear relationship between them. However, it was shown that the cost-optimal degree of capture target (90%) was not achieved. For 1000 MW and 650 MW power plants, the cost optimal levels of capture was 81% and 87% respectively. The IECM model was also used by other researches, such as Klemes *et al.* (2007) who presented a techno-economic model of CO₂ capture process in coal-fired power plants ranges being 300-2000 MW. The capital expenses, operation cost, sorbent cost, steam cost and electricity cost were taken into account. The interesting finding was the absorber vessel was the largest cost item, which accounted for 30% of the total system cost. The results showed that, as plant size and capture level increases, the cost of CO₂ avoided decreases. For instance, the cost of capturing 95% of CO₂ from 600MW power plant was is approximately 14% cheaper than that of capturing 85% of CO₂. Recently, Mores *et al.* (2012a) have

presented an optimisation study in order to specify the operation conditions of cost-optimal design of amine based post-combustion CO₂ capture plant. Again, they also found that the absorber column is the most costly piece of capital cost. The important finding in their study is that, the annualised cost was increasing linearly with CO₂ capture level to 80%, and then exponentially to 95%.

2.6 Conclusions

The MEA-based post-combustion CO₂ capture process is considered as the most advanced and convenient technology to remove and store CO₂ coming from coal-fired power plants, as it can capture the low partial pressure of CO₂ and retrofit the existing power plants easily. However, there is still a concerned disadvantage, that it consumes a lot of energy for regenerating circulate solvent in regenerator. Therefore, to find the trade-offs between operation costs and capture efficiency appears to be important when apply this capture technology. In this case, it is required to develop an appropriate model for the capture process. As reported in the previous literatures, the computational intelligence based model has a better performance than mechanistic model and statistical model, in terms of calculation speed and generalisation ability. Meanwhile, the controllability analysis of post-combustion CO₂ capture process has also attracted a number of attentions in the past, such as application of PID-based and MPC based control schemes. With respect to the former control scheme, RGA, GRDG and DRGA are usually used to determine the appropriate control loops. On the other hand, the centralized control structure, such as MPC scheme, is based on the linear and nonlinear transfer models. In this study, it is aimed at controlling and optimizing the post-combustion MEA-based CO₂ capture plant integrated with coal-fired power plants using more advanced computational intelligence modelling techniques and control strategies. This is proved as a novel way for maximizing the CO₂ capture efficiency, as well as minimizing the capital cost.

Chapter 3. Modelling of a Post-combustion CO₂ Capture Process Using Bootstrap Aggregated Neural Networks

3.1 Introduction

Due to the limitations in training data and training algorithms, it is generally not possible to obtain a perfect neural network model. For example, neural network training might be trapped in a poor local minimum or the trained network might over fit noise in the training data. Several techniques have been developed to improve neural network generalisation capability, such as regularisation (Bishop, 1991), early stopping (Bishop, 1995), Bayesian learning (MacKay, 1992), training with both dynamic and static process data (Zhang, 2001), and combination of multiple networks (Wolpert, 1992). By training with regularisation, the magnitude of network weight is introduced as a penalty term in the neural network training objective function with the purpose of avoiding unnecessarily large network weights which usually leads to poor generalisation. By training with early stopping, neural network performance on the testing data is continuously monitored during the training process and the training process stops when the neural network prediction errors on the testing data start to increase. Among these techniques, combination of multiple networks has been shown to be a very promising approach to improving model predictions on unseen data.

It is generally considered that a given network architecture cannot represent the inherent nature of the data-generating process. Different neural networks training on different subsets of input space are capable of approximating different classes of functions. Bates and Granger (1969) have initially indicated a point that the model prediction accuracy could be improved by the combination of several individual forecasting models. To improve the accuracy of ANN model, especially with a limited amount of experimental data patterns, the combination of several single neural networks is recommended (Wolpert, 1992). Later, in the study by Sridhar *et al.* (1996), a linear combination of neural networks has been proposed to model chemical process. The individual neural network was trained using different training data set. In addition, they implemented performance comparison between BA-NNs (bootstrap aggregated neural networks) model and ANN model selected using the cross validation scheme, and found the performance of BA-NNs model was better. The results have shown that, the smaller the sample size was, the larger improvement using the BA-NNs was. Following with that, Zhang *et al.* (1998) have presented a BA-NNs model to predict trajectories of polymer quality variables in batch polymerisation reactors from batch recipes with linear combination of multi networks, The BA-NNs model was developed based on

bootstrap re-sampled training data. The prediction confidence bounds were also applied to BA-NNs model for process control and monitor. The results have shown that the generalisation ability is improved by the combination of neural networks. BA-NNs model has already been applied in various aspects, such as estimation of polymer properties (Zhang *et al.*, 1997), fault diagnosis (Zhang, 2002), prediction of chemical reaction yield (Monemian *et al.*, 2010) and word recognition (Ebrahinpour *et al.*, 2011).

The vast majority of the total annualised cost is associated to the operating cost, which is mainly consisted of the solvent regenerating cost. Then, Mac Dowell and Shah (2013) presented an optimisation study to identify the cost-optimal degree of CO₂ capture with 660MW sub-critical coal fired power plant. They have addressed the optimisation problems about trade-offs between cost of CO₂ emission to atmosphere and cost related to electricity output reduction from power plant, by using a dynamic, non-equilibrium model. 95% capture level was identified as the optimum cost-optimal degree for 660 MW sub-critical coal-fired power plant. More interesting, more than 50% energy cost in the system was associated to the cost of solvent regeneration, which accounts for most parts of total annualised cost. Therefore, seeking for the optimal operation conditions with minimum energy consumption and satisfied capture degree appears much more important in designing a PCC plant, while non researches focus on this target before. Lawal *et al.* (2010) have identified that the liquid to gas (L/G) molar flow ratios for the absorber had a large effects on heat requirements for capture. The low L/G ratio means lean solvent is contact with significantly more CO₂ in the flue gas and capture more CO₂ yielding higher rich loading. In the stripper, more CO₂ is vaporized with a constant reboiler duty (387K). Thus, the low L/G ratio will result in a low heat duty requirement for capture (MJ/kg CO₂). As a result, in this chapter, the system operation conditions were designed to keep the lean solvent flow rate to the absorber as low as possible under the assumed capture level and constant conditions of inlet flue.

This chapter is organised as follows. Section 3.2 presents an overview of CO₂ capture processes, including equipment size, parameter selection and data description. Section 3.3 presents the establishment details of BA-NNs and the results of static model and dynamic model of a CO₂ capture process using BA-NNs is detailed. Section 3.4 briefly defined the optimisation problem statement, which lower the energy requirement, as well as meet the designed capture degree. In addition, the results of system operation conditions by using BA-NNs model are also presented in this section. Finally, Section 3.5 describes the conclusions and future works.

3.2 Overview of post-combustion CO₂ capture plant

3.2.1 gPROMS simulator description

The CO₂ capture Process considered here is through chemical absorption. Detailed mechanistic model for this process was developed by Lawal *et al.* (2010) and a simulator based on the mechanistic model was developed in gPROMS at the University of Hull, shown in figure 3.1. The process conditions is described in table 3.1. Simulated static and dynamic process operation data is generated using the simulator. 90 patterns of data is generated from gPROMS simulator when the process reaches steady state after a variable step change. As to dynamic process, an amount of 660 data patterns from 7 runs with different variables step changes were produced by simulator gPROMS with sampling interval of 1 second, while, to reduce the repetition of data information, the samples with sampling interval of 5 seconds were used to develop the model, shown in table 3.2.

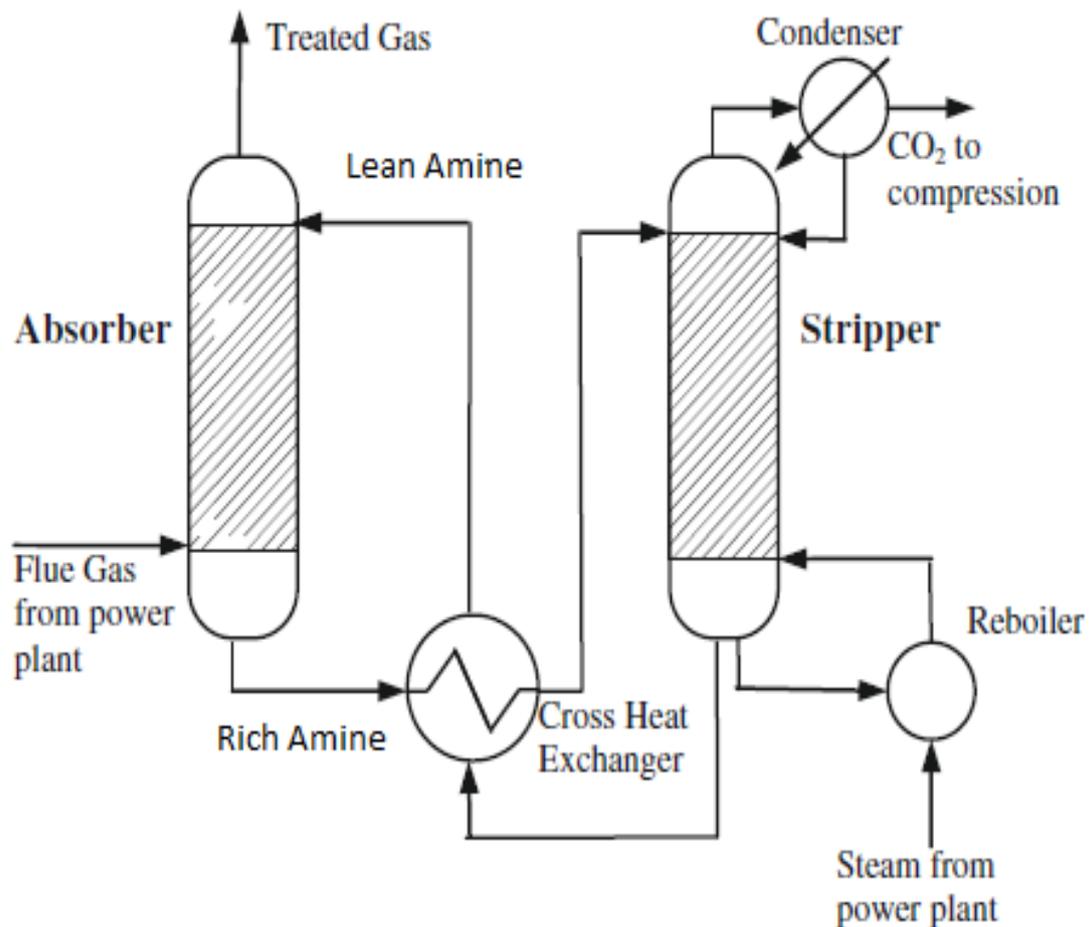


Figure 3.1: Simplified process flow diagram of chemical absorption process for post-combustion capture.

Description	Value	unit
Column internal diameter	0.427	m
Height of packing	6.1	m
Nominal packing size	0.0381	m
Packing specific area (absorber)	145	m ² /m ³
Packing specific area (regenerator)	420	m ² /m ³
Cross sectional area	0.1432	m ²
Reboiler volume	1	m ³
Condenser volume	2	m ³
Solvent	MEA	30wt%

Table 3.1: Equipment Specification.

Run NO.	Samples from gPROMS simulator		Samples to develop the stacked neural network model	
	Sampling rate (second)	Number of samples	Sampling rate (second)	Number of samples
1	1	518	5	104
2	1	512	5	103
3	1	415	5	83
4	1	521	5	105
5	1	311	5	63
6	1	515	5	103
7	1	492	5	99

Table 3.2: The dynamic sampling information from simulator and to develop BANN model.

3.2.2 Variables selection

The objective of this study is to develop a BA-NNs model for assessing the process efficiency and plant performance. Among hundreds of parameters in the post-combustion CO₂ capture process, two parameters were taken into account for reflecting plant efficiency. They are CO₂ production rate and CO₂ capture level, which were defined as dependent variables in this study.

- (1) CO₂ production rate is the amount of CO₂ extracted from the flue gas and amine solvent shown in Equation 3.1. It is measure at the top of regenerator as below:

$$\eta_{CO_2} = m_{CO_2} \times v_{CO_2} \quad 3-1$$

where η_{CO_2} is CO₂ production rate, m_{CO_2} is CO₂ mass fraction, and v_{CO_2} is gas flow rate out of regenerator.

- (2) CO₂ capture level is the amount of CO₂ extracted from the inlet flue gas in absorber column. It is calculated as follows:

$$\delta_{CO_2} = 1 - \frac{\dot{m}_{CO_2} \times \tilde{v}_{CO_2}}{\dot{m}_{CO_2} \times \check{v}_{CO_2}} \quad 3-2$$

where δ_{CO_2} is the CO₂ capture level, \dot{m}_{CO_2} and \check{m}_{CO_2} denote CO₂ mass fraction in gas out of absorber and inlet flow gas of absorber, \tilde{v}_{CO_2} and \check{v}_{CO_2} represent gas flow rate out of absorber and inlet flow gas rate, respectively.

Notably, the CO₂ capture level is a total different parameter comparing with CO₂ production rate. In details, CO₂ capture level is the percentage of CO₂ initially extracted from inlet flue gas. It is measured in the absorber column in the process and is an indicator for performance of the absorber. The extracted CO₂ will be further processed after absorption, which will be regenerated during regeneration process from MEA solution. It is cooled in the condenser and compressed to become the final product. The CO₂ production rate represents the amount of CO₂ product after condenser. This parameter is an indicator for the whole process because it is not affected by a single component of the process. Both CO₂ capture level and CO₂ production rate were selected as model inputs.

The input variables of static neural network model were selected as inlet gas flow rate, CO₂ mass fraction in inlet flow gas, inlet gas flow pressure, inlet gas flow temperature, lean solvent circulation rate, MEA concentration and lean solution temperature (7 variables). However, in dynamic neural network model, there were selected as inlet flow gas rate, CO₂ concentration in inlet flue gas, inlet gas temperature, inlet gas pressure, MEA circulation rate,

lean loading, lean solution temperature and reboiler temperature (8 variables) (Lawal *et al.*, 2009a; Lawal *et al.*, 2010; Biliyok *et al.*, 2012a; Biliyok *et al.*, 2012b; Mac Dowell and Shah, 2013).

3.3 BANN model

3.3.1 Construct of BA-NNs models with sensitivity analysis (SA).

The static BA-NNs model with sensitivity analysis is adopted to develop a model that indicates the relationships between 7 independent variables described in section 3.2.2 and capture level. It is consisted of four steps: (1) construct the BA-NNs model, (2) apply sensitivity analysis to the modelling results, (2) validate the results with experts, (4) reformulating and reapplying the stacked neural network models to the data and generating results. The process steps are shown as a flow diagram in figure 3.2.

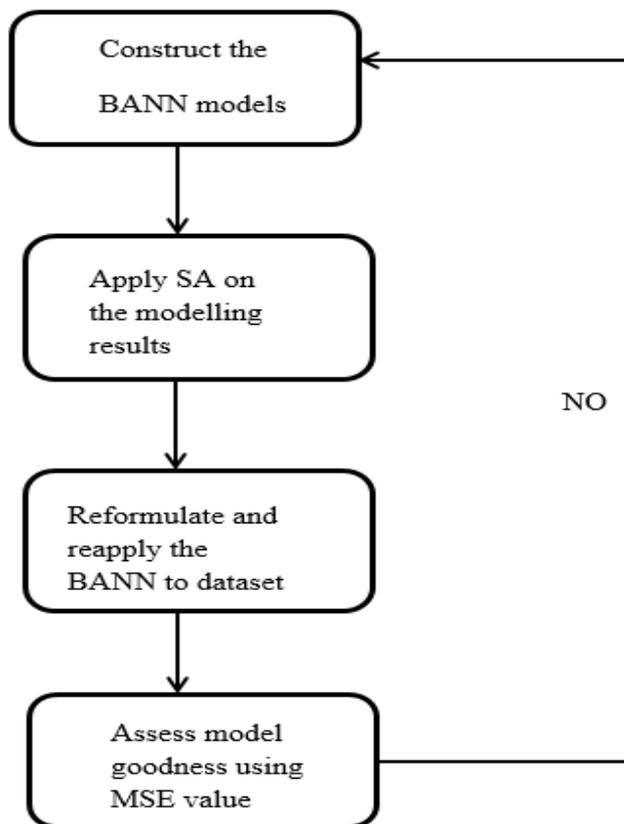


Figure 3.2 BA-NNs model establishment process.

The general form of nonlinear static model was proposed as below:

$$\hat{y}_s = f(u_1, u_2, \dots, u_7) \quad 3-3$$

where \hat{y}_s is the capture level, u is the process inputs, $f(\cdot)$ is the nonlinear function represented by neural network.

The nonlinear dynamic one-step-ahead prediction and multi-step-ahead predictions of capture level and CO₂ production rate by using 8 variables detailed in section 3.2.2 can be expressed by the first-order equations (Eq.3-4 and 3-5), as they are simple to calculate and accurate enough.

$$\hat{y}_{od}(t) = f(y(t-1), u_1(t-1), u_2(t-1), \dots, u_8(t-1)) \quad 3-4$$

$$\hat{y}_{md}(t) = f(\hat{y}_{md}(t-1), u_1(t-1), u_2(t-1), \dots, u_8(t-1)) \quad 3-5$$

where t is the discrete time. $\hat{y}_{od}(t)$ is one-step-ahead prediction, which is predicted by the measured output at time $t-1$, $y(t-1)$. $\hat{y}_{md}(t)$ is the multi-step-ahead prediction, which is predicted by the predicted output at time $t-1$, $\hat{y}(t-1)$.

Figure 3.3 shows a BA-NNs model, where several neural network models are developed to model the same relationship. These individual networks are trained on bootstrap replications of the original training data. Instead of selecting a “best” single neural network model, these individual neural networks are combined together to improve model accuracy and robustness. The overall output of the aggregated neural network is a weighted combination of the individual neural network outputs:

$$f(X) = \sum_{i=1}^n w_i f_i(X) \quad 3-6$$

where $f(X)$ is the aggregated neural network predictor, $f_i(X)$ is the i th neural network, w_i is the aggregating weight for combining the i th neural network, n is the number of neural networks to be combined, and X is a vector of neural network inputs. Since the individual neural networks are highly correlated, appropriate aggregating weights could be obtained through principal component regression (Zhang *et al.*, 1997). Instead of using constant aggregating weights, the aggregating weights can also dynamically change with the model inputs (Ahmad and Zhang, 2005). Another advantage of bootstrap aggregated neural network is that model prediction confidence bounds can be calculated from individual network predictions (Zhang, 1999). The standard error of the i th predicted value is estimated as:

$$\sigma_e = \left\{ \frac{1}{n-1} \sum_{b=1}^n [y(x_i; W^b) - y(x_i; \cdot)]^2 \right\}^{1/2} \quad 3-7$$

where $y(x_i; \cdot) = \sum_{b=1}^n y(x_i; W^b) / n$ and n is the number of neural networks in an aggregated neural network. Assuming that the individual network prediction errors are normally distributed, the 95% prediction confidence bounds can be calculated as $y(x_j; \cdot) \pm 1.96\sigma_e$. A narrower confidence bound, i.e. smaller σ_e , indicates that the associated model prediction is more reliable. Thus, model prediction associated with a narrow prediction confidence bounds is preferred and is considered to be reliable. On the other hand, model prediction with a wide confidence bound is unreliable and should not be trusted.

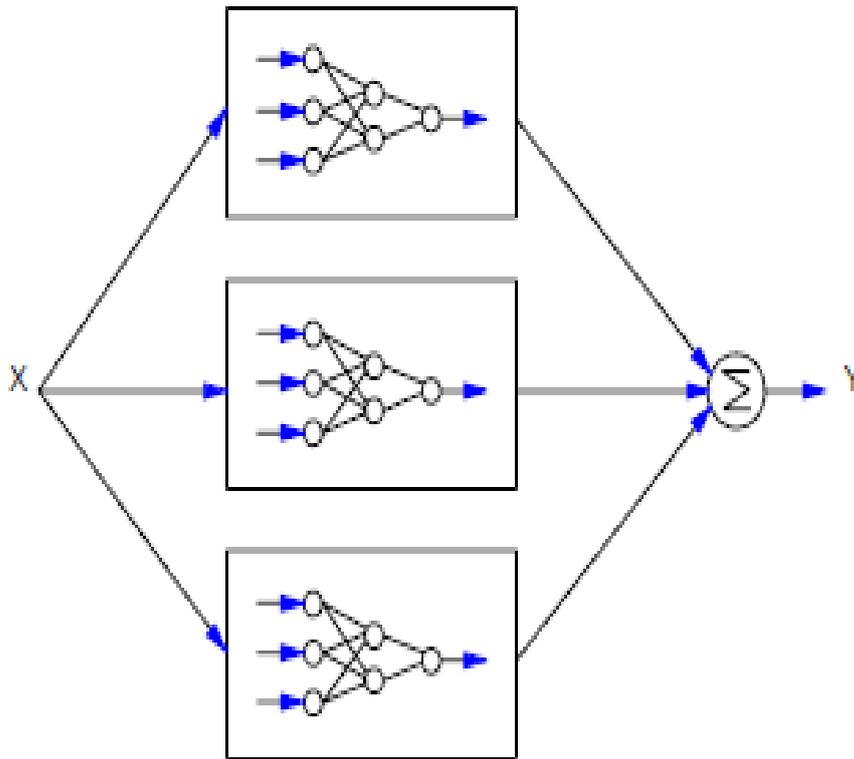


Figure 3.3: An aggregated neural network.

Prior to modelling, the data were pre-screened in case of missing values and outliers. Since the data collected in the process has different physical units, each variable should be rescaled by centring with respect to their means. Next, these pre-processed data has been divided into three sets: training data (56%), test data (24%) and validation data (20%). The activation function used in the output layer of the neural networks is linear activation function while that used in the hidden layer is the sigmoidal function expressed in Eq. 3-8.

$$g(x) = 1/(1 + \exp(-x)) \quad 3-8$$

The number of hidden neurons for individual neural network was determined through cross-validation. In other words, the neural network which gives the lowest mean square error (MSE) on the validation data was considered to have the appropriate number of hidden neurons, as they were trained on different replication of training data. In same way, the number of combined networks is also determined by MSE on validation data. The number of combined networks which has the lowest MSE is used to construct the stacked neural network model. To develop stacked neural network models, bootstrap re-sampling with replacement was used to generate 30 replications of the training data for each combined neural network. Then determined number of neural networks were combined linearly, shown in Equation 2.8. In this study, the aggregating weights were obtained by the simple method, which means the stacked neural network output is an average of individual neural network outputs. Lastly, run 2, 6 and 7 are also used to validate the model developed in this study.

The impacts of the independent parameters on dependent parameters can be identified by sensitivity analysis (SA). The prediction model can be more accurate by cutting one insignificant predictor. In details, there are two ways for implementing the sensitivity analysis in the neural network models, namely equation method and the variable perturbation method (Wu *et al.*, 2010).

With respect to equation method, it is based on Sheriff's theory, which indicates that the effects of input variables can be reflected by the derivative of the ANN's model dependent variables with respected to the independent variables. The formula is shown as follows:

$$\begin{aligned}
 S_i &= \frac{\partial O}{\partial I_i} = \frac{\partial O}{\partial h^{1+1}} \frac{\partial h^{1+1}}{\partial I_i} = O(1-O) \frac{\partial h^{1+1}}{\partial I_i} \\
 &= O(1-O) \sum_k w_k^2 V_k^1 (1-V_k^1) \frac{\partial h_k^1}{\partial I_i} \\
 &= \sum_k O(1-O) w_k^2 V_k^1 (1-V_k^1) w_{ik}^1
 \end{aligned}
 \tag{3-9}$$

where O is the value of output node, h_k^n is the input sum value of the $(n-1)$ th layer to the k th node in the n th layer; V_j^n is the output value of j th node in the n th layer after applying the activation equation to the sum value from previous later; I_i the i th input to the network; w_{jk}^n is the connection weight between the j th node in layer $n-1$ and k th node in layer n ; S_i the sensitivity result to the i th input.

However, the variable perturbation method is based on the perturbation of one input variable at a time and calculating the variations of outputs. There are two ways of adding perturbation, shown as Eq.3-10 and 3-11.

$$I_n = I_n + \sigma \quad 3-10$$

$$I_n = I_n \times \sigma \quad 3-11$$

where I_n represents the n th input and σ is the perturbation introduced to the input variable. The output changes are collected from each run and an averaged value is calculated. The average value reflects the impacts of inputs on the outputs. This step will repeat on each input variables.

In this study, for sensitivity analysis, the MSE values of the actual and predicted outputs before and after removing an assumed input variable are calculated. Then, compare both MSE values, if the latter one is less than the former one, it means the assumed input variable affects output variable significantly.

3.3.2 Results and discussions of static model

As to steady state model, only the absorber is modelled. The trajectories for simulated static process operation data using first principal model developed in Lawal *et al.* (2009b) are shown in figure 3.4. The process variables that are selected as model input variables are: inlet flue gas flow rate, CO₂ concentration in inlet flue gas, pressure of flue gas, temperature of flue gas, lean solvent flow rate, MEA concentration and temperature of lean solvent. They are shown in plots (a) to (g) respectively in figure 3.4. CO₂ capture level, shown in plot (h) in Figure 3, is taken as the model output variable. Considering that static data is usually not abundant in practice as a process is usually operated in just a few steady states, a small number of data samples are produced as shown in figure 3.4. It can be seen clearly that, the step changes occur on each input variable and the output variable (CO₂ capture level) was significantly affected by those input variables. From the correlation assessment of the input variables, it was found that there was no linear relationship between these input variables, which means that they are mutually independent. Therefore, the neural network model is constructed according to Eq. 3-3.

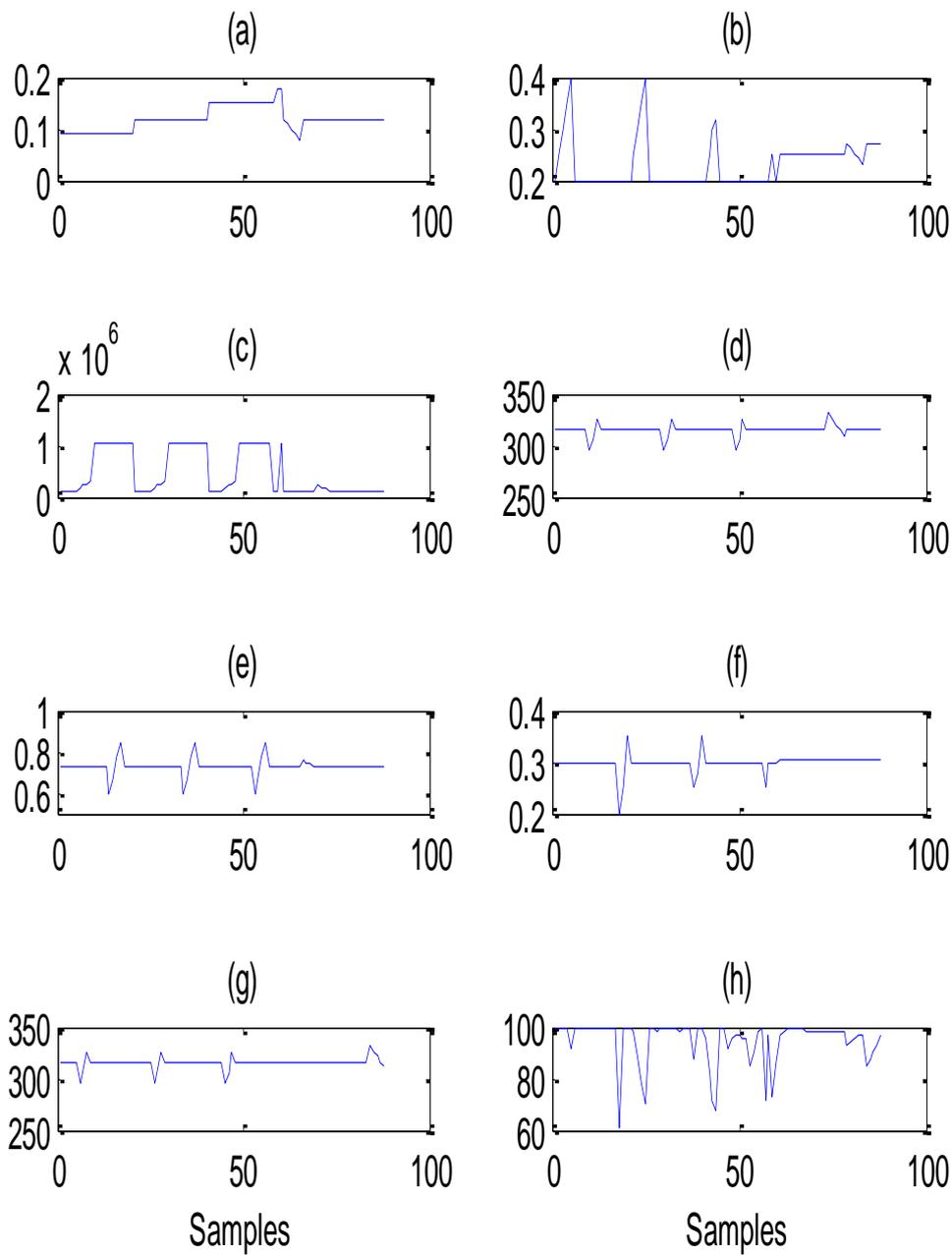


Figure 3.4: Static process operation data (inlet gas flow rate (a), CO₂ mass fraction in inlet flow gas (b), inlet gas flow pressure (c), inlet gas flow temperature (d), lean solvent circulation rate (e), MEA concentration (f), lean solution temperature (g) and capture level (h)).

The generated 90 samples of static data is split into training data (56%), testing data (24%), and unseen validation data (20%). The data is scaled to zero mean and unit variance before they are used for network training. A bootstrap aggregated neural network consists of 30

individual networks is developed. For the development of an individual network, a replication of the training and testing datasets is generated through bootstrap re-sampling with replacement (Efron, 1982) and the network is developed on each bootstrap replication. Each single hidden neural network is a single hidden layer feedforward neural network. The number of hidden neurons in each neural network is determined through cross validation. A number of neural networks with different numbers of hidden neurons (between 3 and 30) are trained on the training data and tested on the testing data. The network with the lowest mean squared errors (MSE) on the testing data is considered to have the appropriate number of hidden neurons. Each network was trained using the Levenberg-Marquardt optimisation algorithm (Marquardt, 1963) with regularisation and cross-validation based “early-stopping”.

Figure 3.5 shows the number of hidden neurons in the individual neural networks. It can be seen that number of hidden neurons vary a lot with different training and testing data sets. This indicates that the “best” neural network structure depends on the model building data and slight variation in the model building data can lead to different neural network structure. The individual networks are then combined through averaging.

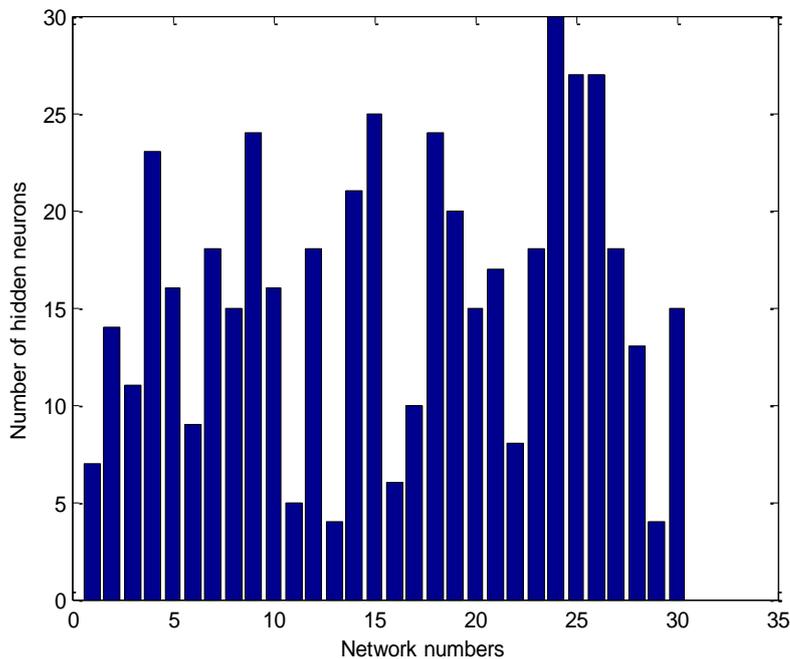


Figure 3.5: Number of hidden neurons in individual neural networks.

Figure 3.6 shows the mean squared errors (MSE) on training and testing data (top) and on unseen validation data (bottom) from the 30 different single neural networks. Figure 3.6 shows these from aggregated neural networks with different numbers of constituent networks.

It is clearly seen that single neural networks give inconsistent performance on the model building data (training and testing data) and the unseen validation data. For instance, the 14th and 17th networks are among the few best networks in terms of performance on the model building data, but their performance on the unseen validation data is not among the best. The non-robustness of single neural networks is clearly indicated by the difference in performance of individual neural networks on model building data and unseen validation data. Figure 3.7 clearly indicates that the bootstrap aggregated neural networks give consistent performance on the model building data and on the unseen validation data. In figure 3.7, the first bar in each plot represents the first single neural network shown in figure 3.6, the second bar represents combining the first two single neural networks, and the last (30th) bar represents combining all the 30 networks. It can be seen from figure 3.7 that as more networks are combined, the MSE values on both model building data and unseen validation data decrease and converge to stable values. Furthermore, bootstrap aggregated neural networks give much more accurate prediction performance than most of the individual networks. This demonstrates that bootstrap aggregated neural networks reliable and accurate prediction performance than single neural networks.

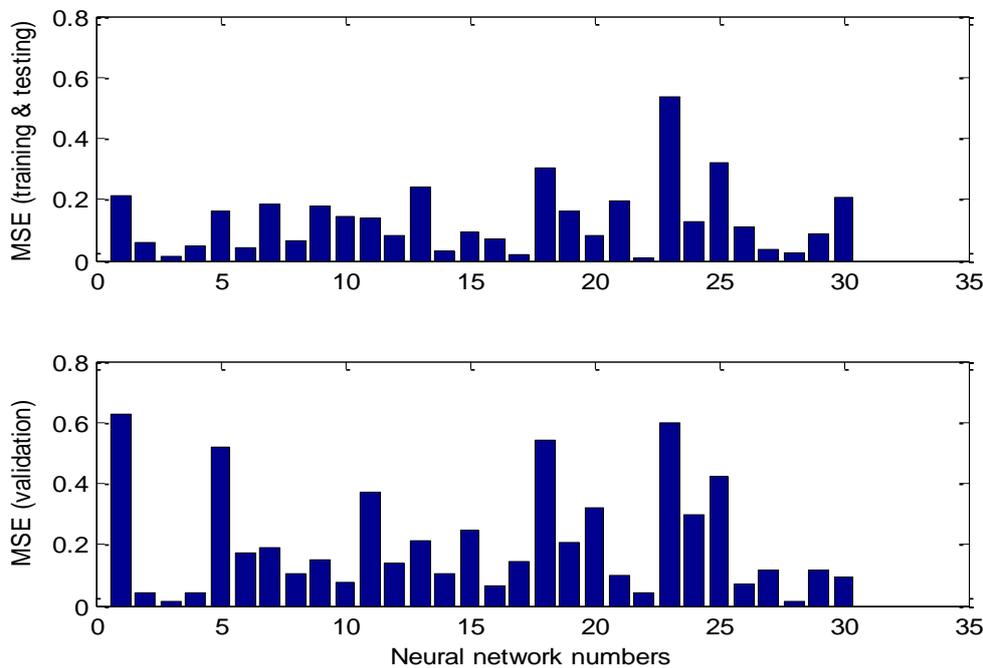


Figure 3.6: MSE of CO₂ capture level for individual neural networks.

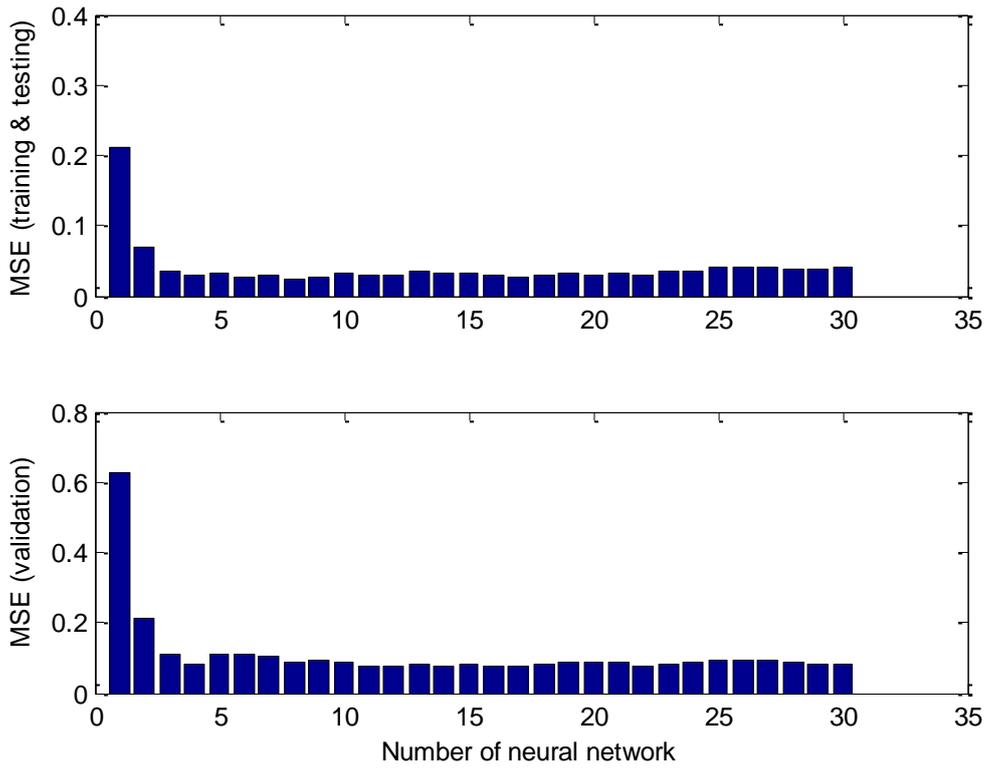


Figure 3.7: MSE of CO₂ capture level for aggregated neural networks.

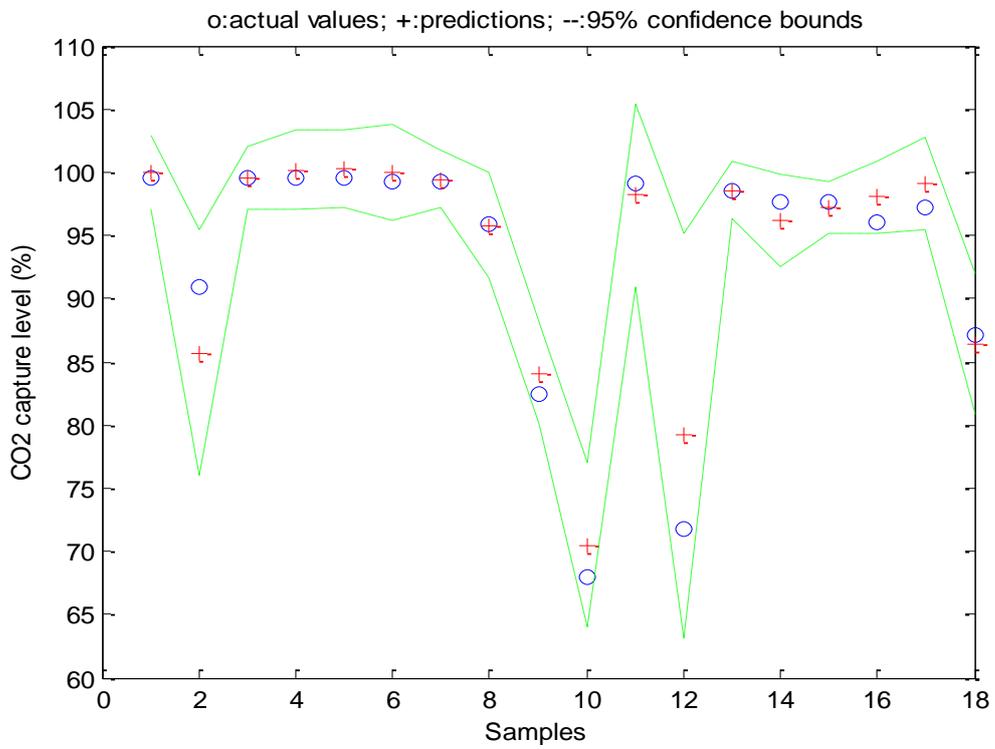
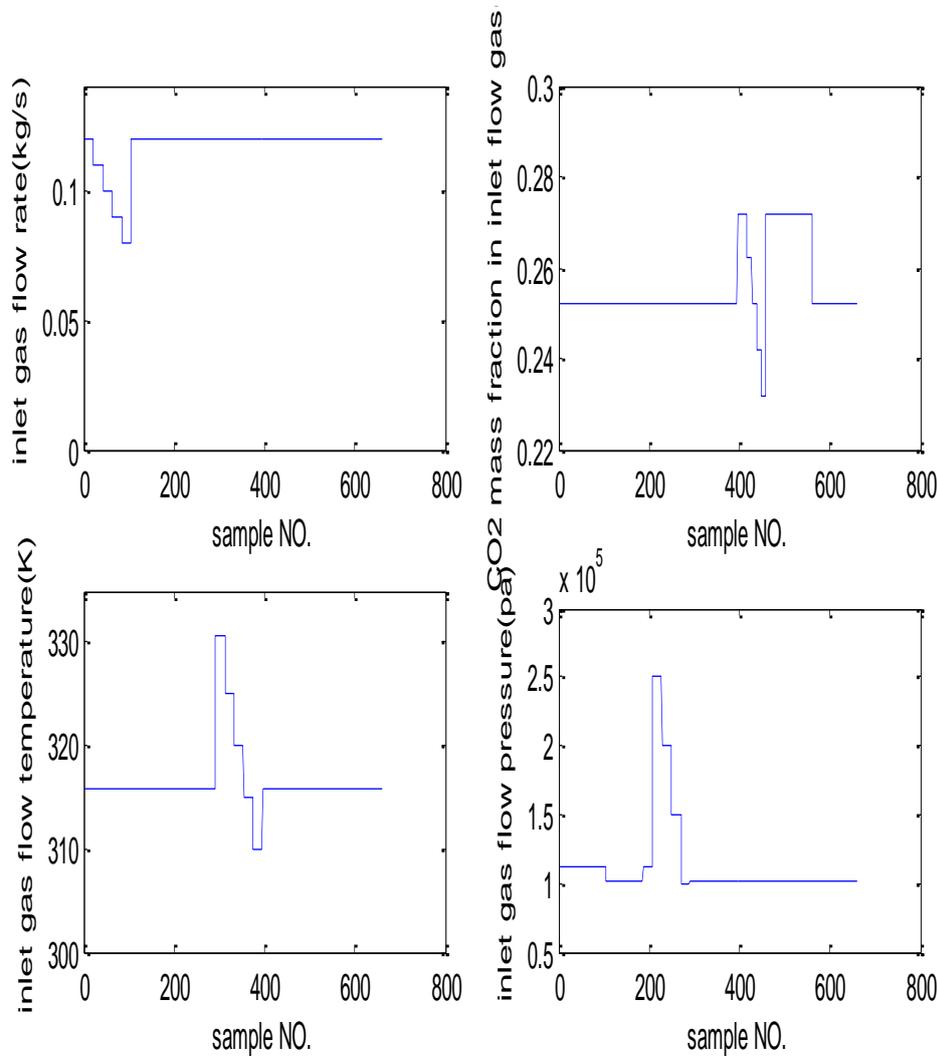


Figure 3.8: Static model predictions for CO₂ capture level on unseen validation data.

Figure 3.8 shows the actual values, predictions, and 95% confidence bounds of CO₂ capture level on the unseen validation data. Clearly, the predictions by using aggregated neural networks are close to the actual values but not extremely accurate. This is because the amount of training data is not enough so that the data feature cannot be learned precisely. The prediction confidence bounds offer extra information to the process operators on the prediction reliability, such as rejection or acceptance of a particular prediction from the stacked neural network model. A prediction with narrow prediction confidence bounds is considered to be reliable while, on the other hand, a prediction with wide prediction confidence bounds is considered to be unreliable. Figure 3.8 shows that the model prediction confidence bounds are quite narrow for almost all samples, except for 2nd, 10th, 11th, and 12th samples. Therefore, extra care needs to be taken when using predictions for these samples.

3.3.3 Results and discussions of dynamic BANN model



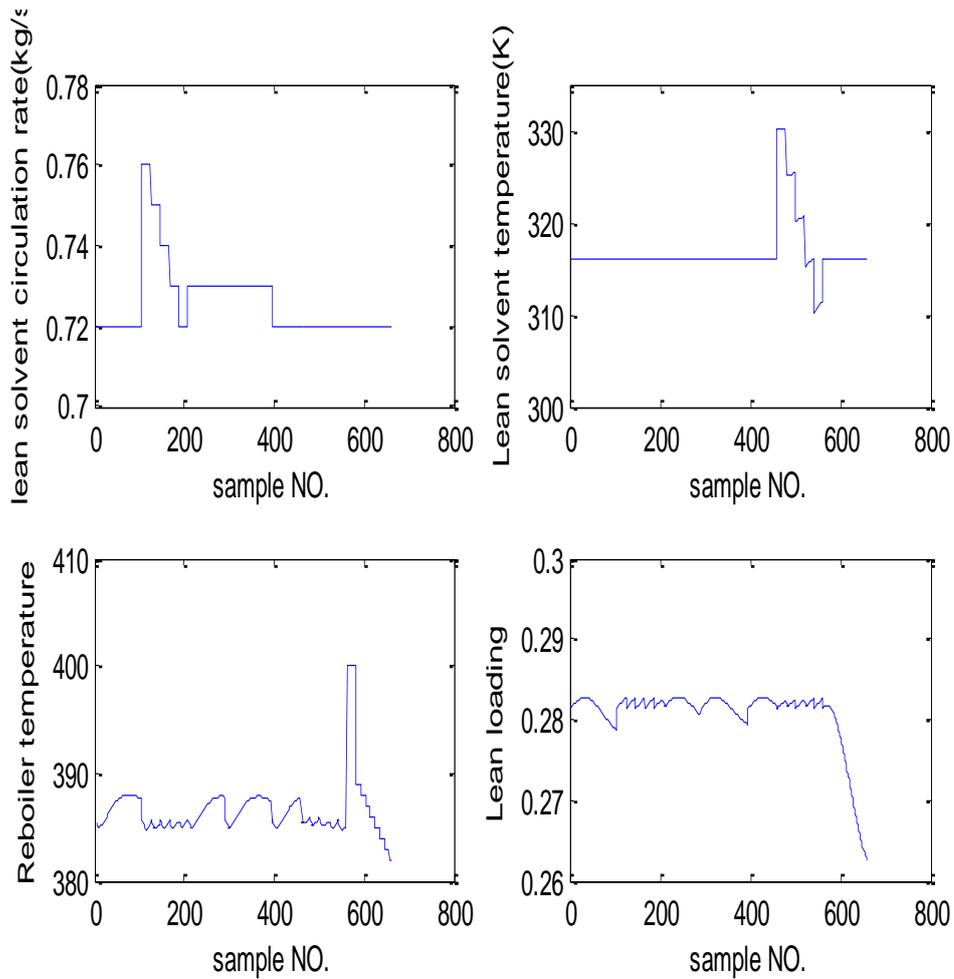
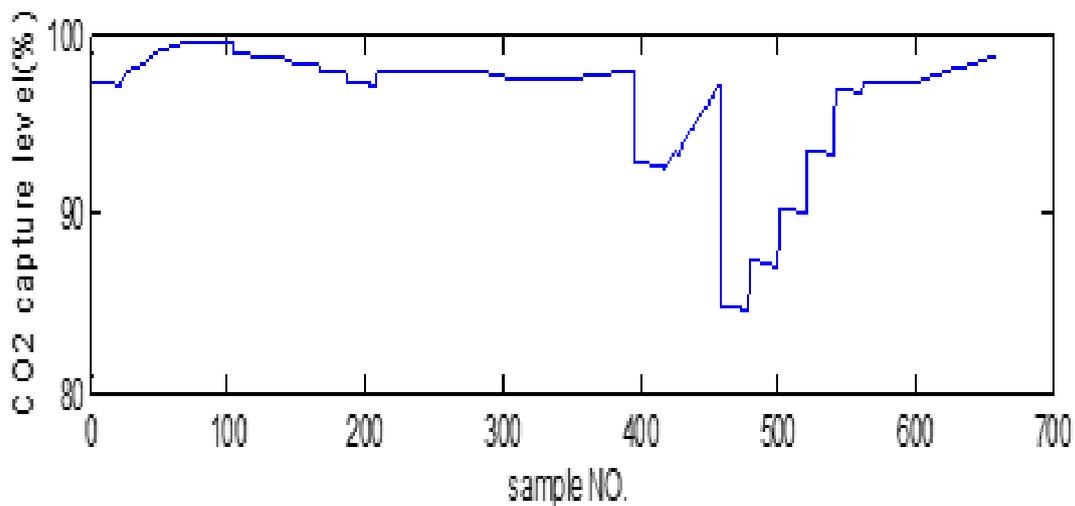


Figure 3.9: The time series plot of input variables (inlet gas flow rate, CO₂ mass fraction in inlet flow gas, inlet gas flow temperature, inlet gas flow pressure, lean solvent circulation rate, lean solvent temperature, reboiler temperature, lean loading).



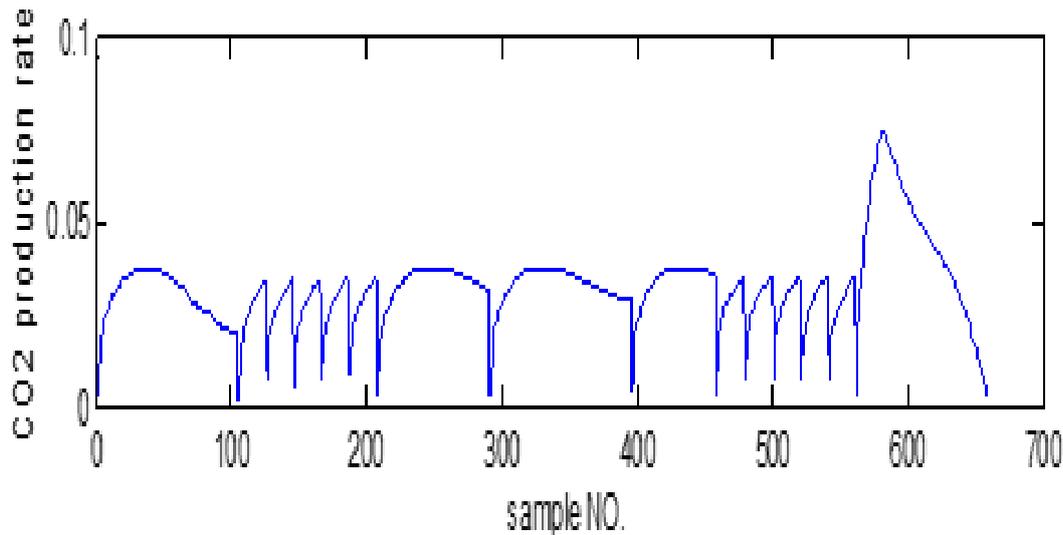


Figure 3.10: The time series plots of output variables (CO₂ capture level and CO₂ production rate).

Figure 3.9 and 3.10 illustrate the trajectories for input variables (inlet flow gas rate, CO₂ concentration in inlet flue gas, inlet gas temperature, inlet gas pressure, MEA circulation rate, lean solution temperature, reboiler temperature and lean loading) and output variables (CO₂ capture level and CO₂ production rate). It is clear to see, when one input parameter was step changed, the other input parameters were kept constant. Furthermore, it should be emphasised here that CO₂ production rate fluctuated significantly during the operation, shown in figure 3.10. This output variable was calculated by gas flow rate out of stripper and CO₂ concentration in outlet gas of stripper, which has been mentioned in Section 3.2. By exploring the original data, it was found that, the values of outlet gas flow rate of stripper were almost ignorable at the start-up of each run. The reason is that, the response time for outlet gas flow rate of stripper existed when implementing the process operation. By calculating the correlation coefficients of each input variables, it was found that there was no linear relationship between the input variables. That is to say, these input variables are mutually independent.

The generated 690 patterns of data were split into training data (56%), testing data (24%), and unseen validation data (20%). The data were scaled to zero mean and unit variance before they were used for neural network training. Two multi-inputs single output (MISO) first order dynamic nonlinear models were developed for CO₂ capture level and CO₂ production rate using bootstrap aggregated neural networks.

Each of the nonlinear dynamic models is developed using a bootstrap aggregated neural network consisting of 30 individual neural networks. These individual neural networks are single hidden layer feedforward neural networks. The number of hidden neurons in each network was determined through cross validation. Each network was trained using the Levenberg-Marquardt optimisation algorithm (Marquardt, 1963) with regularisation and cross-validation based “early-stopping”.

Figure 3.11 shows the MSE values on model building (training and testing) data and unseen validation data from individual neural networks. It can be seen from Figure 8 that the individual networks give various prediction performance. Furthermore, their performance on the training and testing data is not consistent with that on the unseen testing data. For example, network 15 is among the worst performing networks on the training and testing data. However, it offers the best performance on the unseen data. This clearly demonstrates the non-robust nature of single neural networks. Figure 3.12 shows the MSE values on model building data and unseen validation data from different aggregated neural networks. In figure 3.12, the horizontal axes represent the number of individual networks contained in an aggregated neural network. The first bar in figure 3.12 represents the first individual neural network shown in figure 3.11 and second bar in figure 3.12 represents combining the first two individual networks shown in figure 3.11. The last bar in figure 3.12 represents combining all the 30 neural networks. It can be seen from figure 3.12 that bootstrap aggregated neural networks give much more consistent performance on model building data and unseen validation data. The MSE values of aggregated neural networks generally decrease as more networks are combined and converge to a stable level. This occurs in both the model building and unseen data sets. In addition to robustness, Figure 3.12 also indicates that aggregated neural networks give more accurate performance than individual neural networks. Figure 3.13 shows the one-step-ahead predictions and multi-step-ahead predictions of CO₂ production rate on 7th batch (492 samples) using aggregated neural networks. It is clearly seen that the predictions are very close to the actual values, except for a few samples where the CO₂ production rates are very high or very low. The slightly larger prediction errors at these samples are likely due to the fact that training data is scarce at these extreme operating points. The accurate multi-step-ahead predictions are very encouraging indicating that the model has captured the underlying dynamics of the process. The long range predictions are very accurate till about 90 step-ahead predictions. Such accurate long range predictions are more than sufficient for model predictive control and real-time optimisation applications.

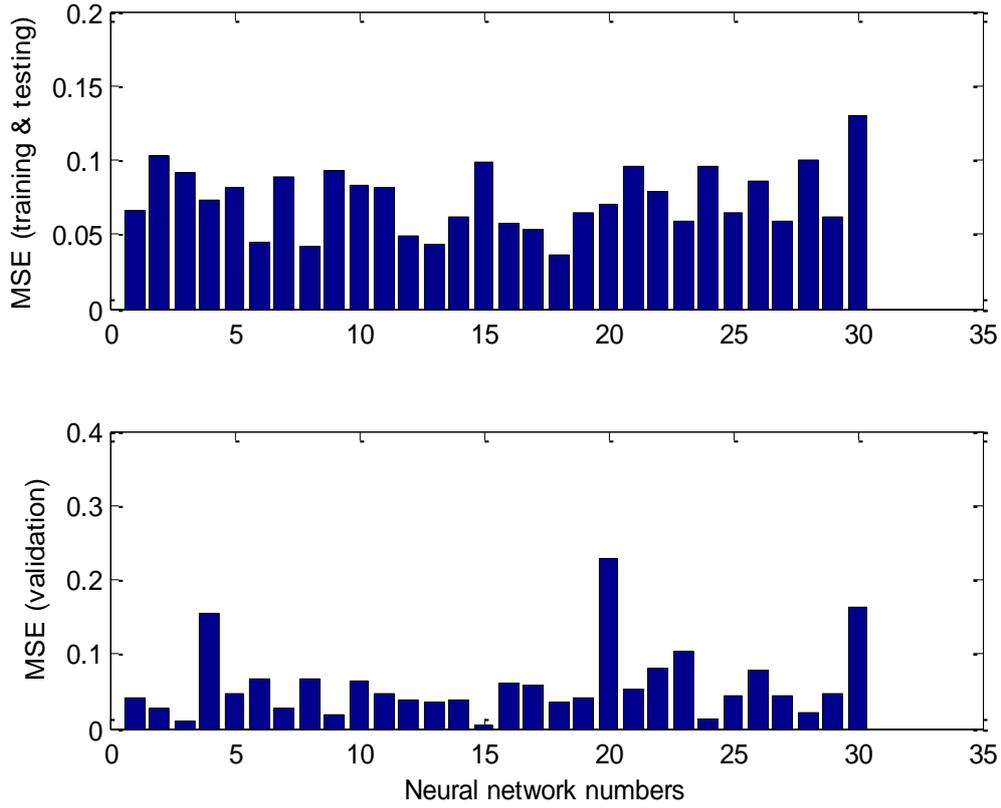


Figure 3.11: MSE of CO₂ production rate for individual neural networks.

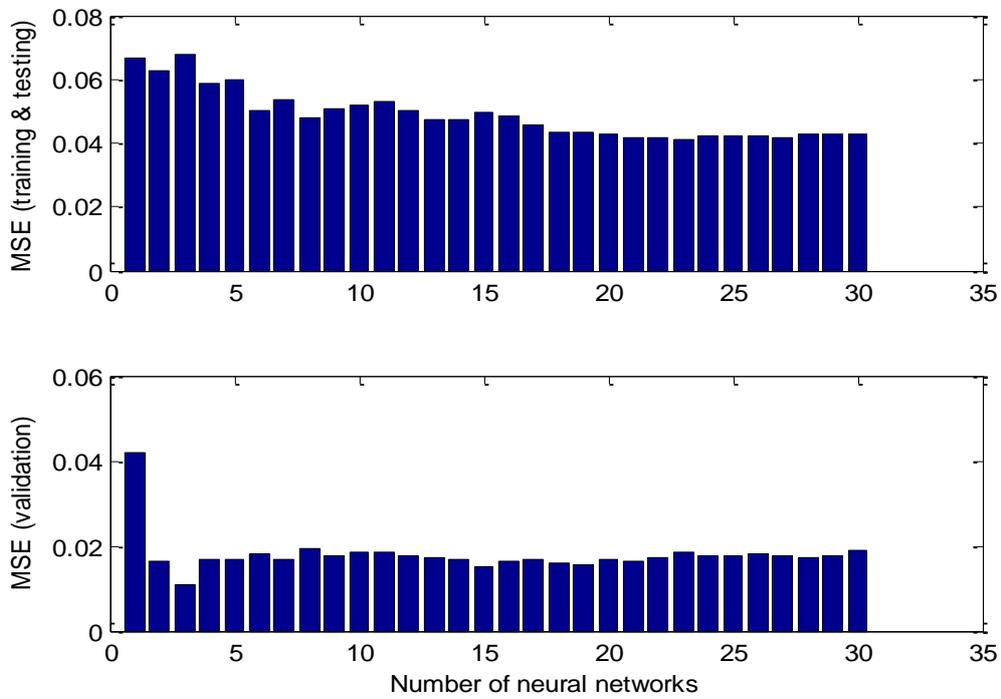


Figure 3.12: MSE of CO₂ production rate for aggregated bootstrap neural networks.

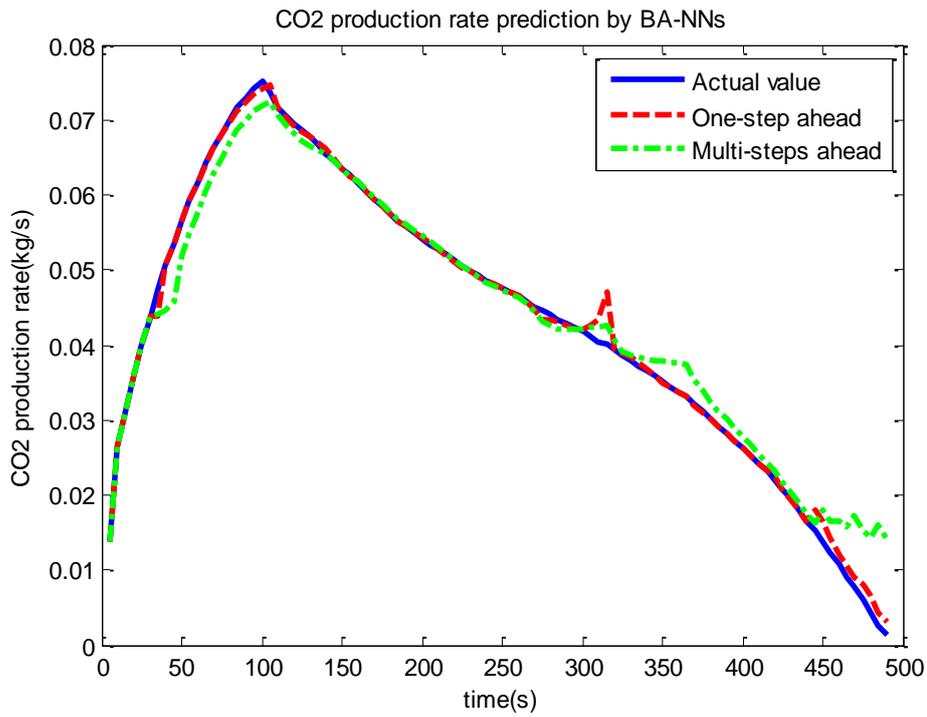


Figure 3.13: Dynamic model prediction of CO₂ production rate by BA-NNs on 7th batch.

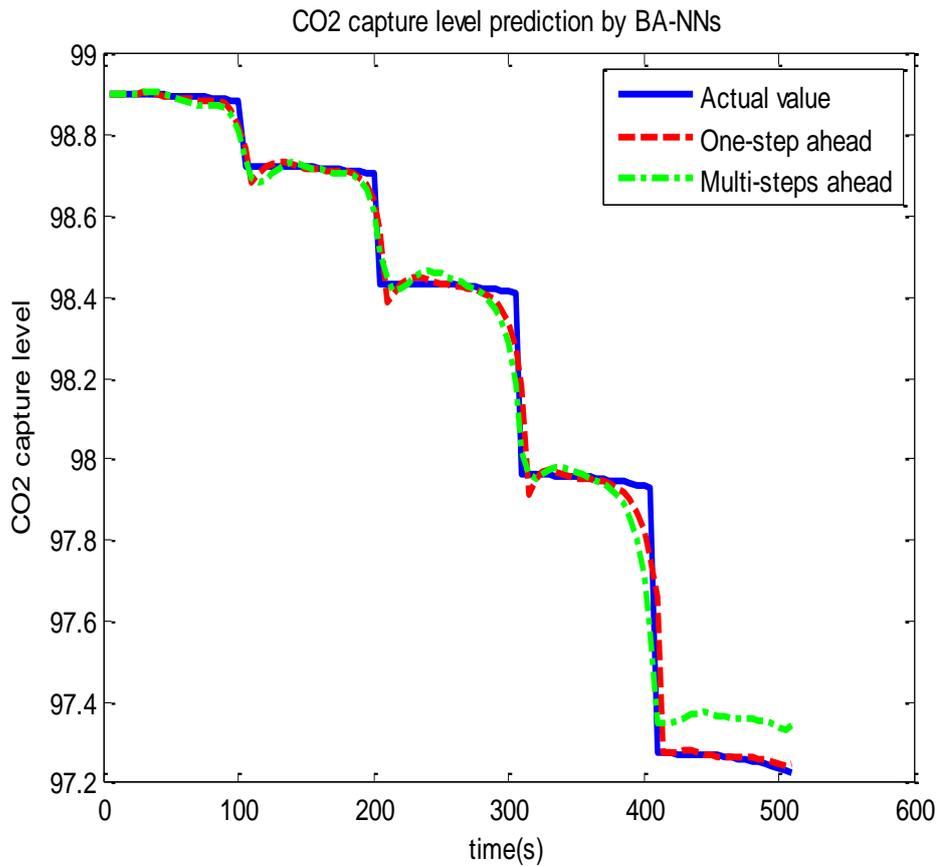


Figure 3.14: Dynamic model prediction of CO₂ capture level by BA-NNs on 2nd batch.

The dynamic BA-NNs predictions of CO₂ capture level on 2nd batch (512 samples) are also accurate as shown in figure 3.14. It can be seen that the long range predictions are accurate until 82-steps-ahead predictions. Again such long prediction horizon is generally adequate for many applications such as model predictive control and real-time optimisations.

3.4 Optimisation of CO₂ capture plant

3.4.1 Optimisation problem statement

Figure 3.15 indicates the typical flow sheet of the CO₂ capture process by chemical absorption constructed in Aspen Hysys, which is similar to the gPROMS model used by Lawal *et al.* (2010). The previous literature has specified the numerical values of the design and operating parameters, shown in table 3.1. One of the important sections in the PCC system is the reboiler unit, which provides heat to the stripper for vaporizing purity CO₂ from rich solution. The heat in the reboiler is generated by the low pressure steam from the coal-fired power generation plant. For the PCC process, several trade-offs are existing such as CO₂ capture level, heating utility required by reboiler, electricity power consumption by compressor, blower and CO₂ pumps. Particularly, the main concern is the large amount of energy consumption for amine regeneration section. As reported by Lawal *et al.* (2010), L/G ratio is the effective indicator to reflect the energy consumption. The capture system is treating the inlet flue gas with constant conditions, such that 0.12 kg/s of flue gas flow rate, 0.25 CO₂ mass fraction, 320 K of temperature and 1.01 bar of pressure. Hence, to minimize lean solvent flow rate v_L as far as possible will minimize the energy consumption in the regenerator.

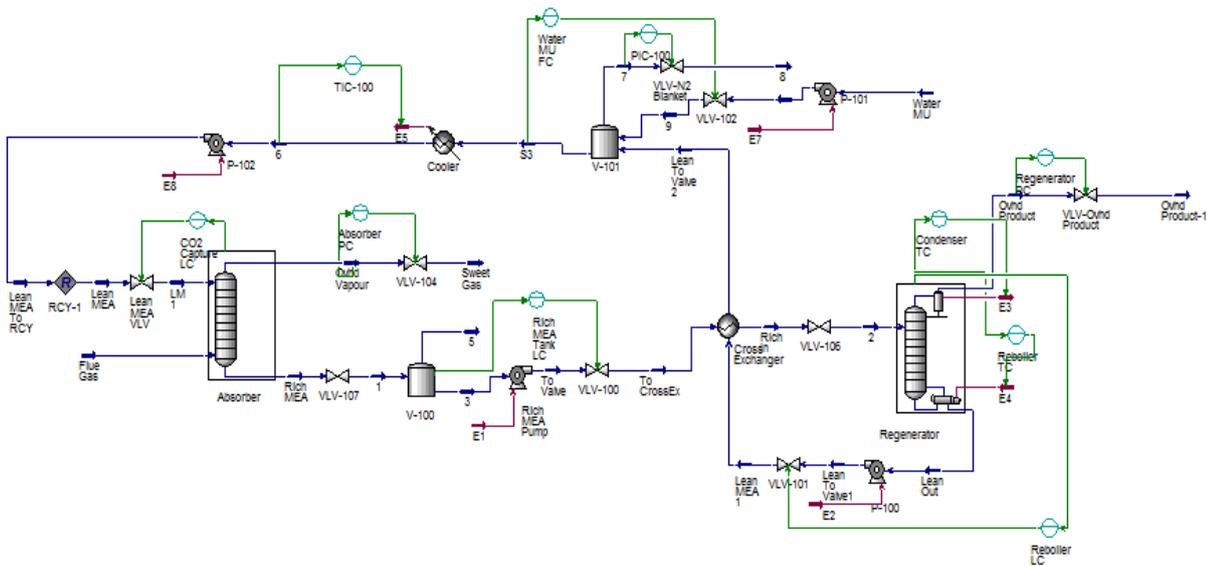


Figure 3.15: Steady-state flow sheet in Aspen Hysys.

In this section, the proposed optimisation problem (OP) consists on the minimization of lean solvent flow rate at the absorber inlet (v_L) with several constraints. It can be formally stated as below:

$$\text{To Minimize: } v_L \quad 3-12$$

subject to the process constraints and operation conditions:

$$0.3 \text{ kg/s} \geq v_L \geq 0.8 \text{ kg/s} \quad 3-13$$

$$25\% \geq C_{MEA} \geq 35\% \quad 3-14$$

$$\delta_{CO_2} \geq 90\% \quad 3-15$$

where C_{MEA} denotes the MEA concentration in lean solvent and δ_{CO_2} represents the capture level. Eqs. 3-13, 3-14 and 3-15 refer to a set of the inequality constraints defined in order to circumscribe a feasible operation region.

The MEA concentration and inlet solvent temperature at the absorber were commonly determined with set points at approximately 30 wt% and 313.15k, respectively (Arias *et al.*, 2016). The BA-NNs model was used in solving the optimisation problem. The sequential quadratic programming (SQP) method implemented in the MATLAB Optimisation Toolbox was used to solve the optimisation problem

3.4.2 Optimal operation

For the optimal operation of an assumed coal-fired power plant integrated with PCC plant it is necessary to explore the optimal values of lean solvent flow rate at a specific optimal capture level. In this case study, the constraint on capture level given by Eq. 3-16 was varied for the following specific values:

$$\delta_{CO_2} = \{90\%, 92\%, 94\%, 96\%, 98\%\} \quad 3-16$$

The optimal values of lean solvent flow rate at different capture level were displayed in figure 3.16. The corresponding L/G ratios are 4.75, 4.93, 5.13, 5.35 and 5.63, respectively. It demonstrates that the more CO₂ it captures, the higher L/G ratio the plant should be controlled. As shown in figure 3.16, the more solvent is required for absorbing more CO₂ at higher capture level. However, it will result in an increasing heat requirement for lean solvent regeneration. To achieve the target which capture level is no less than 90%, the optimal operation of lean solvent flow rate is 0.5698.

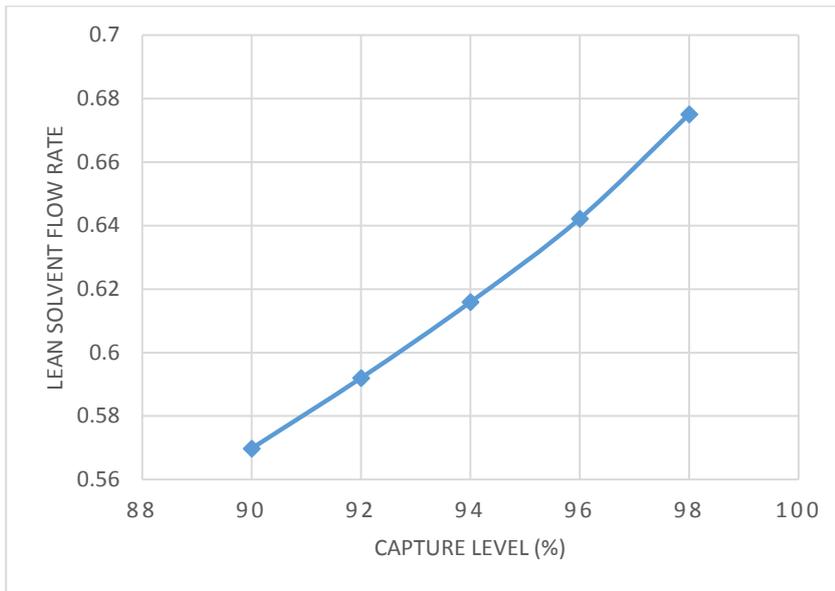


Figure 3.16. Optimal lean solvent flow rate at different capture level.

3.5 Conclusions

The neural network static and dynamic models of CO₂ production rate and CO₂ capture level are developed and they are shown to be able to give accurate predictions. The aggregated neural networks model is found to be the useful tool to predict the post-combustion CO₂ capture process, which is more accuracy and reliable than the traditional neural network models. Bootstrap aggregated neural networks give consistent performance on the model building data and unseen validation data. Furthermore, bootstrap aggregated neural networks can give model prediction confidence bounds, which are a very useful measure on the prediction reliability and can be incorporated in the optimisation framework to give reliable optimisation results (Zhang, 2004). Reliable optimisation of the CO₂ capture process using the developed neural network models will be studied in the future.

In addition, the optimal operation of coal-fired power plant integrated with PCC plant was also investigated. The objective function is to minimize the lean solvent flow rate to low the energy consumption in the process system. The optimal operation studies were carried out for the optimal different carbon capture levels. For the low energy consumption in the regenerator with target ($\delta_{CO_2} \geq 90\%$), the optimal lean solvent flow rate at absorber is set as 0.5698. The future study could be concentrated on the techno-economic evaluation of PCC plant, with a consideration of market price.

Chapter 4. Modelling of a Post-combustion CO₂ Capture Process Using Extreme Learning Machine

4.1 Introduction

As stated in Chapter 2, SLFNNs trained by the back propagation (BP) learning algorithm have experienced some issues: firstly, the most appropriate learning rate is difficult to determine; secondly, the presence of local minima affects the modelling results; then, networks would possibly be over trained because of too many hidden neurons, leading to poor generalization performance; lastly, it is also time-consuming when applying gradient based learning. As all the parameters of SLFNNs are required to be adjusted randomly, the training procedures of SLFNNs with traditional method may takes several hours, several days or even more time.

To address the issue of slow training in traditional SLFNNs, Huang *et al.* (2004) has firstly proposed a new method called extreme learning machine (ELM). The structure of ELM is similar to SLFNNs, while their ways of parameters updating are different. Specifically, the weights between the input and hidden layers are randomly assigned instead of tuned, while the weights between the hidden and output layers are determined in a one-step regression type approach using Moore-Penrose (MP) generalised inverse. In this case, an ELM can be built very quickly and the generalisation performance is better. As the weights between the input and hidden layers are randomly assigned, correlations can exist among the hidden neuron outputs and variations in model performance. Later, they extended ELM from SLFNNs to radial basis function (RBF) case, which arbitrarily assigns the kernels instead of adjusting them (Huang and Siew, 2004). It was compared to support vector machine (SVM) and the regression results showed that the learning speed of ELM was faster and the generalisation performance was as good as SVM. However, as the inputs weights and hidden biases are randomly assigned, there may exist dissatisfied and unnecessary selections. As a result, ELM requires more hidden neurons than traditional tuning-based SLFNs, which may make ELM response slowly to testing data. To make up for the deficiency, Zhu *et al.* (2005) have used the modified differential evolution (DE) to determine optimal input weights and hidden biases, in which both validation fitness and the norm of output weights are used as selection criteria. In details, when the difference of the RMSE between different sets of inputs weights and hidden biases is small, the one resulting in smaller norm of output weights is selected. This type of ELM, called E-ELM, has a faster response speed to unknown testing data and better generalisation performance than original ELM. It is noted that E-ELM is not suitable

for data sets with a large amount of features, because DE algorithm takes much time to search for optimal input weights and hidden biases. Huynh *et al.* (2008) used a fast regularized least-squares scheme to replace DE. It can achieve higher generalisation performance and faster learning speed in both small and large number of inputs features. ELM has been applied in various aspects, such as fault diagnosis (Hu *et al.*, 2008), forecasting (Sun *et al.*, 2008), regression (Frenay and Verleysen, 2016), and classification (Iosifidis *et al.*, 2015).

In this chapter, principal component regression (PCR) is used to obtain the output layer weights, instead least square algorithms. It is able to overcome the correlation issues among hidden neuron outputs. Besides, the multiple ELMs are built on bootstrap re-sampling replications of the original training data and then combining these ELMs in order to enhance model accuracy and reliability. The proposed method is applied to the dynamic model development of the whole post-combustion process plant.

This chapter is structured as follows: Section 4.2 briefly presents the BA-ELM and a method for calculating output layer weights in BA-ELM using PCR, as well as aggregating multiple ELM. Application results and discussions are presented in section 4.3. Section 4.4 draws some concluded remarks.

4.2 Development of BA-ELM

4.2.1 Single hidden neural networks

Figure 4.1 shows the structure of a single hidden layer feedforward neural network (SLFN). For N arbitrary distinct samples (x_j, t_j) , $j = 1, \dots, N$, where $x_j = [x_{j1}, x_{j2}, \dots, x_{jn}]^T \in R^n$ is a vector of network inputs and $t = [t_{j1}, t_{j2}, \dots, t_{jm}]^T \in R^m$ is a vector of the target values of network outputs. The output of a standard SLFNs, $o_j = [o_{j1}, o_{j2}, \dots, o_{jm}]^T \in R^m$ with \tilde{N} hidden nodes and activation function $g(x)$ is shown in the following equation:

$$o_j = \sum_{i=1}^{\tilde{N}} \beta_i g_i(w_i \cdot x_j + b_i), \quad j = 1, \dots, N \quad 4-1$$

where $w_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ is a vector of the weights between the i^{th} hidden node and the input nodes, b_i is the bias of the i^{th} hidden nodes, x_j is the j^{th} input sample, $\beta_i \in R^m$ is the weight linking the i^{th} hidden node and the output node. The output node is chosen to have linear activation function in this paper.

In theory, the standard SLFNs can approximate any continuous nonlinear functions with small error, which means $\sum_{j=1}^{\tilde{N}} \|o_j - t_j\| = \varepsilon_j$. Specifically, there exists β_i , w_i and b_i to make:

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i(w_i \cdot x_j + b_i) - t_j = \varepsilon_j, \quad 4-2$$

To obtain the values of network parameters, the small error was assumed to be ignored. Hence,

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i(w_i \cdot x_j + b_i) = t_j, \quad j = 1, \dots, N \quad 4-3$$

The above equation can be written as $\mathbf{H}\boldsymbol{\beta}=\mathbf{T}$, where:

$$\mathbf{H}(w_1, \dots, w_{\tilde{N}}, b_1, \dots, b_{\tilde{N}}, x_1, \dots, x_N) = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_{\tilde{N}} \cdot x_1 + b_{\tilde{N}}) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_{\tilde{N}} \cdot x_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}} \quad 4-4$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m} \quad \text{and} \quad \mathbf{T} = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m} \quad 4-5$$

In the above equations, \mathbf{H} is called hidden layer output matrix of the neural network and the i^{th} column of \mathbf{H} is the i^{th} hidden node output with respect to inputs x_1, x_2, \dots, x_N . Training of SLFNs can be done through finding the minimum value of $E = \min \|\mathbf{H}_{N \times \tilde{N}} \boldsymbol{\beta}_{\tilde{N} \times m} - \mathbf{T}_{N \times m}\|$.

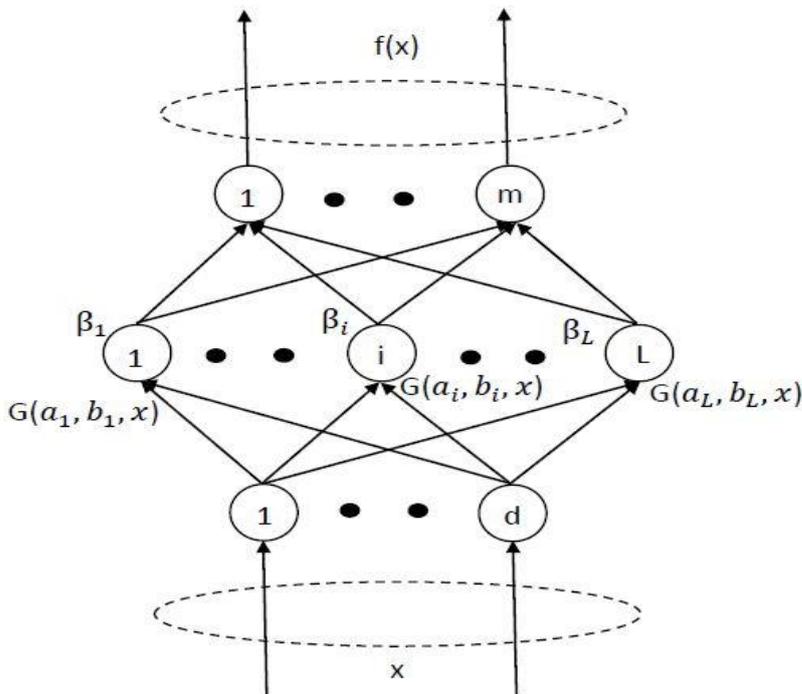


Figure 4.1: The structure of single hidden layer feedforward networks.

SLFNs are usually trained by gradient-based learning algorithms, such as BP algorithm, which typically need many iterations and typically slow. The process of training is to search

the minimum value of $\|\mathbf{H}_{N \times N} \boldsymbol{\beta}_{N \times m} - \mathbf{T}_{N \times m}\|$ by numerical optimisation methods. In this procedure, the parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{w}, \mathbf{b})$ is iteratively adjusted as below:

$$\boldsymbol{\theta} = \boldsymbol{\theta}_{k-1} - \eta \frac{\partial E(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad 4-6$$

where η is the learning rate. By using BP algorithm, the parameters are updated by error propagation from the output layer to the input layer.

4.2.2 BA-ELM

Huang et al. has proved that, if the activation function $g(x)$ is infinitely differentiable in any interval and the number of hidden nodes is large enough, it is not necessary to adjust all the weighting parameters of the network (Huang *et al.*, 2006). In other words, the weights and biases between the input and hidden layers can be randomly chosen. In order to get good performance, the required number of hidden nodes is not more than the number of input samples. Huang et al. have used a method of finding a least square solution of the linear equation $\mathbf{H}\boldsymbol{\beta}=\mathbf{T}$ to obtain the weights between the hidden and output layers.

$$\boldsymbol{\beta} = \mathbf{H}^\dagger \mathbf{T} \quad 4-7$$

where \mathbf{H}^\dagger is the generalised inverse of \mathbf{H} .

However, as the hidden layer outputs can be collinear, the modelling performance would be poor by using least square solution to find the weights between the hidden and output layers. This would be especially true for ELM as they have randomly assigned hidden layer weights and typically large number of hidden neurons are required. In this paper, PCR is used to obtain the weights between the hidden and output layers to overcome the multicollinearity problems. Instead of regressing \mathbf{H} and \mathbf{T} directly, the principal components of \mathbf{H} matrix are used as regressor.

The matrix \mathbf{H} can be decomposed into the sum of a series of rank one matrices through principal component decomposition.

$$\mathbf{H} = \mathbf{u}_1 \mathbf{p}_1^T + \mathbf{u}_2 \mathbf{p}_2^T + \dots + \mathbf{u}_N \mathbf{p}_N^T \quad 4-8$$

In the above equation, \mathbf{u}_i and \mathbf{p}_i are the i^{th} score vector and loading vector respectively. The score vectors are orthogonal, likewise the loading vectors, in addition they are of unit length. The loading vector \mathbf{p}_1 defines the direction of the greatest variability and the score vector \mathbf{u}_1 , also known as the first principal component, represents the projection of each column of \mathbf{H}

onto \mathbf{p}_1 . The first principal component is thus that linear combination of the columns in \mathbf{H} explaining the greatest amount of variability ($\mathbf{u}_1 = \mathbf{H}\mathbf{p}_1$). The second principal component is that linear combination of the columns in \mathbf{H} explaining the next greatest amount of variability ($\mathbf{u}_2 = \mathbf{H}\mathbf{p}_2$) subject to the condition that it is orthogonal to the first principal component. Principal components are arranged in decreasing order of variability explained. Since the columns in \mathbf{H} are highly correlated, the first a few principal components can explain the majority of data variability in \mathbf{H} .

$$\mathbf{H} = \mathbf{U}_k \mathbf{P}_k^T + \mathbf{E} = \sum_{i=1}^k \mathbf{u}_i \mathbf{p}_i^T + \mathbf{E} \quad 4-9$$

where $\mathbf{U}_k = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_k]$, $\mathbf{P}_k = [\mathbf{p}_1 \ \mathbf{p}_2 \ \dots \ \mathbf{p}_k]$, k represents the number of principal components to retain, and \mathbf{E} is a matrix of residuals of unfitted variation.

If the first k principal components can adequately represent the original data set \mathbf{H} , then regression can be performed on the first k principal components. The model output is obtained as a linear combination of the first k principal components of \mathbf{H} as

$$\hat{\mathbf{T}} = \mathbf{U}_k \mathbf{w} = \mathbf{H}\mathbf{P}_k \mathbf{w} \quad 4-10$$

where \mathbf{w} is a vector of model parameters in terms of principal components.

The least squares estimation of \mathbf{w} is:

$$\mathbf{w} = (\mathbf{U}_k^T \mathbf{U}_k)^{-1} \mathbf{U}_k^T \mathbf{T} = (\mathbf{P}_k^T \mathbf{H}^T \mathbf{H} \mathbf{P}_k)^{-1} \mathbf{P}_k^T \mathbf{H}^T \mathbf{T} \quad 4-11$$

The model parameters in Eq.4-6 calculated through PCR are then given by the following equation:

$$\boldsymbol{\beta} = \mathbf{P}_k \mathbf{w} = \mathbf{P}_k (\mathbf{P}_k^T \mathbf{H}^T \mathbf{H} \mathbf{P}_k)^{-1} \mathbf{P}_k^T \mathbf{H}^T \mathbf{T} \quad 4-12$$

The number of principal components, k , to be retained in the model is usually determined through cross-validation (Wold, 1978). The data set for building a model is partitioned into a training data set and a testing data set. PCR models with different numbers of principal components are developed on the training data and then tested on the testing data. The model with the smallest testing errors is then considered as having the most appropriate number of principal components.

As indicated in Chapter 3, combining several networks can improve the prediction accuracy on unseen data and give a better generalization performance. The bootstrap re-sampling replication of the original training data is used for training individual networks and the overall output of the stacked neural networks is a weighted combination of the individual neural networks (Fig. 3.1).

Therefore, the procedure of building bootstrap aggregated ELM model can be summarized as follows:

Given a activation function $g(x)$, and number of hidden nodes \tilde{N} ,

Step1: Apply bootstrap re-sampling to produce n (e.g. $n=50$) replications of the original training data, $(x_i, t_i)_{1, \dots, n} | x_i \in R^n, t_i \in R^m, i=1, \dots, N$.

Step 2: On each bootstrap replication of the original training data, build an ELM model:

Step 2(a): Randomly assign hidden layer weights w_i and bias $b_i, i=1 \dots \tilde{N}$.

Step 2(b): Calculate the hidden layer output matrix \mathbf{H} .

Step 2(c): Calculate the output weights β by PCR.

Step 3: Combine the n (e.g. $n=50$) ELM models by averaging their predictions.

It has been also suggested that, the model prediction confidence bounds can be calculated from individual predictions by using bootstrap aggregated neural networks. The standard error of the i th predicted value is detailed in Eq.3-2. A narrower confidence bound is preferred as it indicates the associated model prediction is more reliable.

4.2.3 Characteristic comparison between SLFNNs and ELM

With the comparison to SLFNNs, ELM has several advantages due to its ways to get output layer weights. Due to its fast training speed, the training time will be reduced a lot. Further, the generalisation performance is better than SLFNNs, which can be reflected by MSE values. Besides, ELM is able to overcome the shortcomings of SLFNNs such as local minimum, improper learning rate and overfitting problems. Mostly important, the activation function in ELM can be either differential or non-differentia, while for SLFNNs, it can be only differential. However, there are still existing some problems. The number of hidden nodes needs to be large enough.

4.3 Performance evaluation of static BA-ELM model

As same to Chapter 3, 90 patterns of static data including 56% training data, 24% testing data, and 20% unseen validation data (20%) are used to build BA-ELM static model. the data pre-process steps are taken as same as mentioned in Chapter 3. The output CO₂ capture Level is predicted by BA-ELM model using same 7 input variables in Section 3.2.2 to compare the performance with BA-NNs model.

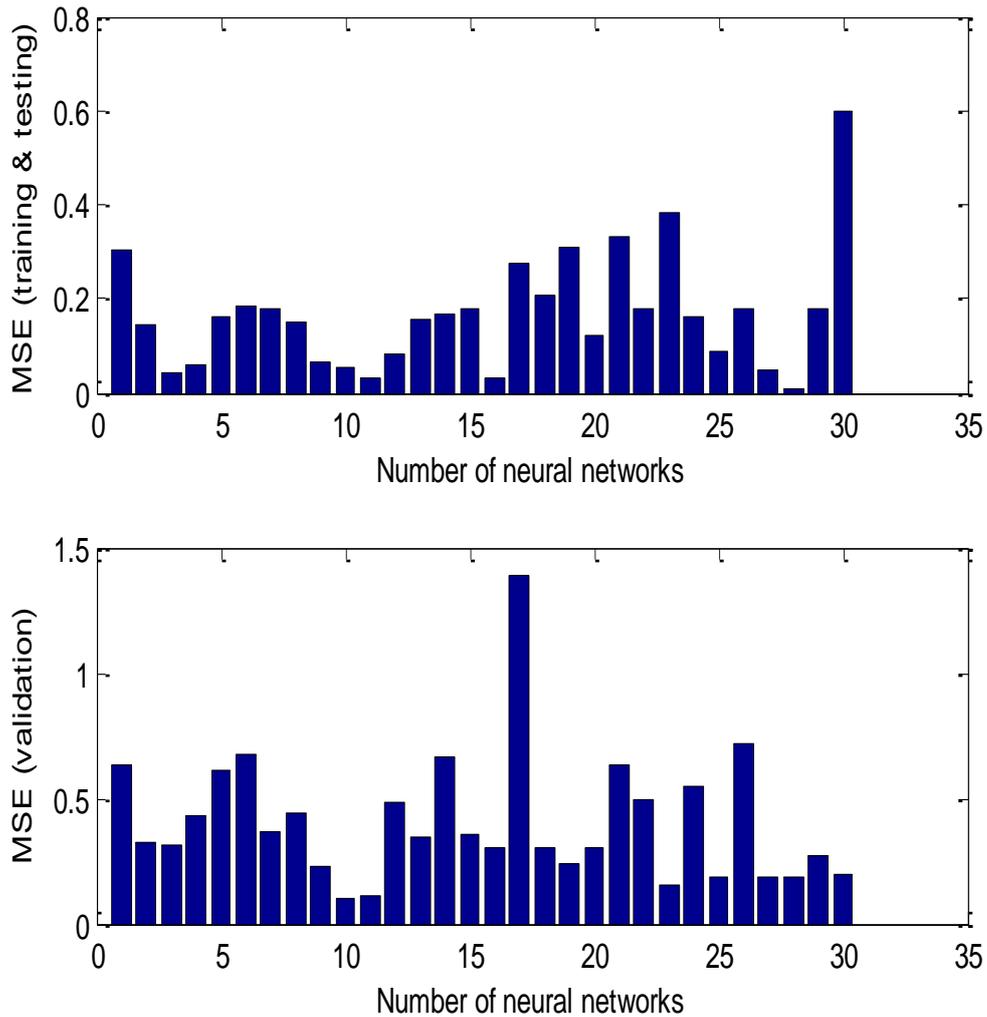


Figure 4.2: MSE of CO₂ capture level for individual ELM.

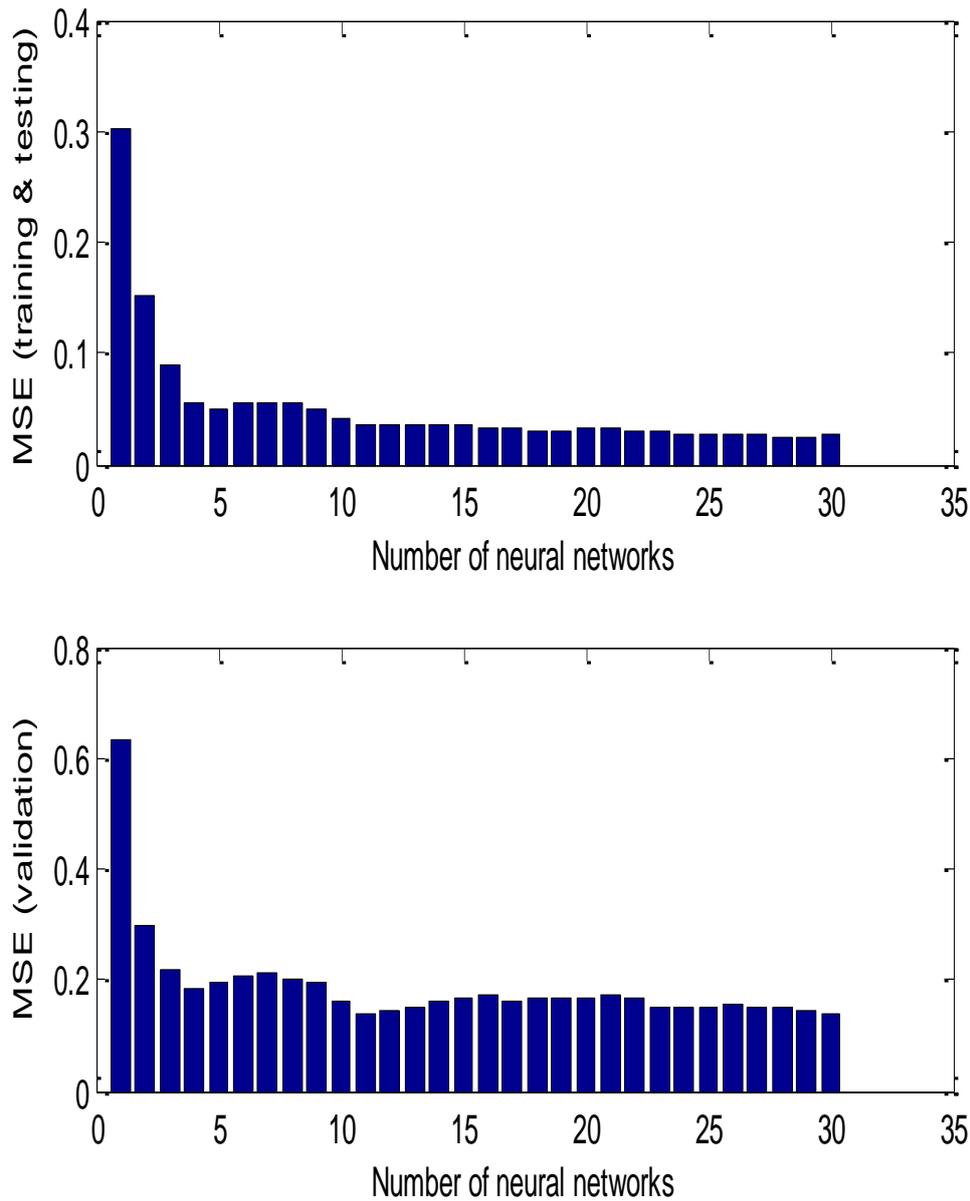


Figure 4.3: MSE of CO₂ capture level for aggregated ELM.

Figure 4.2 shows the MSE values of CO₂ capture level by 30 static individual ELM model, while figure 4.3 indicates those by BA-ELM model. As each ELM model was trained by different replication of original training data, the MSE value of each ELM model on training data is different. Compare figure 4.2 to figure 3.6, the MSE values on training & testing data by BA-ELM is much smaller, which is clearly reflected on Y-axis. It means the training accuracy of BA-ELM is much better. In addition, the time consumed by training procedure is far less than BA-NNs. Seeing from figure 4.3, with the increase of the number of stacked ELMs, the accuracy of the model will be more stable. When the number of stacking is more

than 20, there is no much improvement for accuracy both for training & testing data and the unseen validation data. When aggregate the 30 single ELM together, the MSE values on unseen validation data are similar as those by BA-NNs model. In other words, the generalisation performance of BA-ELM model is as good enough as BA-NNs model.

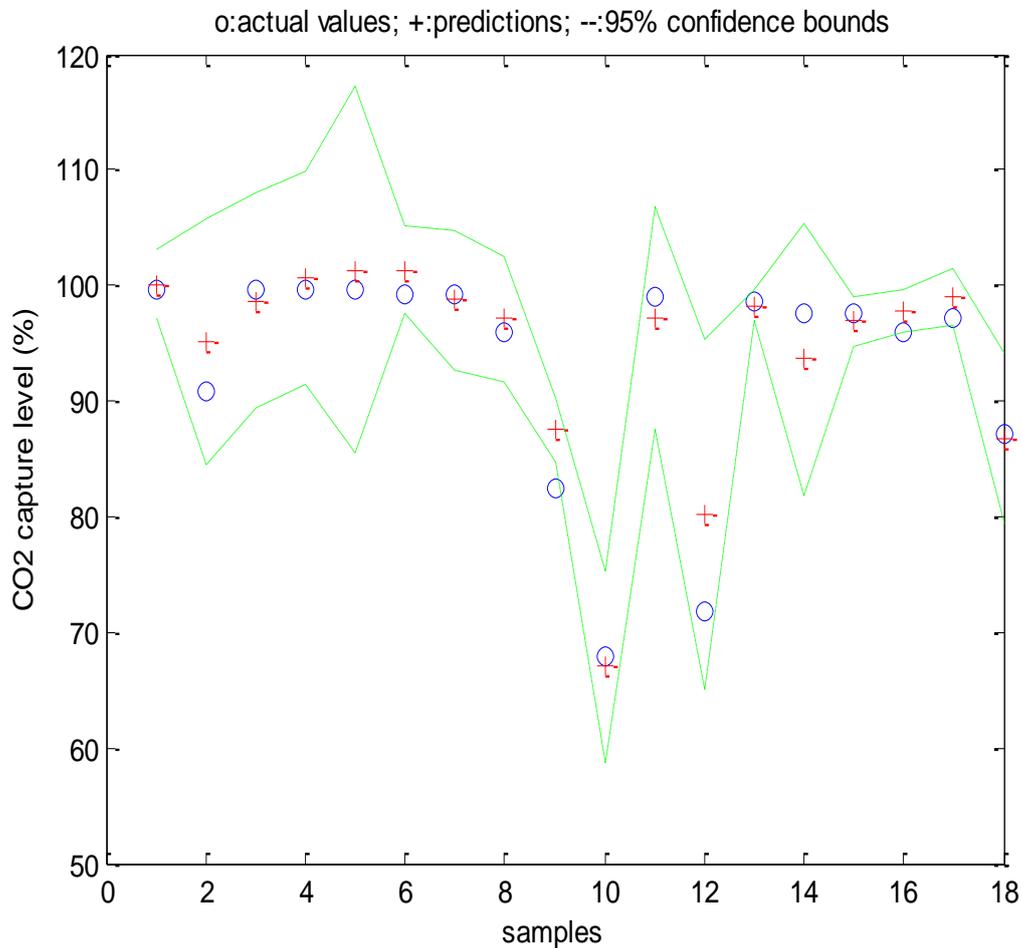


Figure 4.4: Static BA-ELM model predictions for CO2 capture level on unseen validation data.

Figure 4.4 shows the lines graph of actual values, BA-ELM model prediction value and the confidence bounds of capture level. From this graph, the top green line represents the 95% upper confidence limit and the below green line denotes the 95% under confidence limit. The confidence bounds are used to indicate the forecast reliability, by which the confidence bound is either wide or not. In specific details, the narrow prediction confidence bound means the ELM model is unreliable. As long as the prediction value of the model is within the reasonable range of 95% confidence intervals, the model's predictive value is feasible. It could be seen directly on the graph that the difference between prediction value and actual

output value is not big, except for 2nd, 9th, and 12th samples. This result appears to be similar to that predicted by BA-NNs model, while their MSE values by BA-ELM model are smaller

4.4 Performance evaluation of dynamic BA-ELM model

The simulated dynamic process operation data and same 8 variables in section 3.2.2 were used to build data-driven models. The simulated data were generated from the mechanistic model implemented in gPROMS at University of Hull with a sampling time of 5 seconds. The data were divided into three groups as same as Chapter 3: training data (56%), testing data (24%), and unseen validation data (20%). The completed model used the input data of the 7th batch (492 samples) in which the reboiler temperature has a step change, to verify its accuracy for predicting CO₂ production rate. Then, the 2nd batch (512 samples) is used to verify the model accuracy by predicting CO₂ capture level, in which there is a step change of the lean solution flowrate. To demonstrate the good performance of bootstrap aggregated ELM, its results are compared with those from Chapter 3. Before training, the data should be scaled to zero mean and unit variance. Both bootstrap aggregated neural network (BA-NNs) and BA-ELM models combine 30 neural networks. In addition, the numbers of hidden neurons used in BA-NNs and BA-ELM are selected within the range of 2-20 and 40-100 respectively. All models with the number of hidden neurons in the above ranges are developed and tested on the testing data. The models give the smallest mean squared errors (MSE) are considered as having the appropriate number of hidden neurons. The reason for ELM having more hidden neurons is due to the random nature of hidden layer weights in ELM and small number of hidden neurons would usually not be able to provide adequate function representation. The form of the dynamic model is shown in Eqs.3-6 and 3-7.

From the development procedures of these two models, it is clearly recognised that BA-ELM model is very fast because its training only needs one iteration. The performance details of the bootstrap aggregated neural networks and bootstrap aggregated ELM is shown in table 4.1. The training CPU time of BA-ELM is 163.4422 s, approximately 9 times lower than that of BA-NNs. The verifying time by BA-ELM is 0.7176 s, while by BA-NNs is only 0.2964s. This is because the number of hidden neurons in ELM is more than that in SLFNNs, which leads to the computation more complex. With respect to MSE comparison, the MSE value on training data by using BA-ELM is 0.0488, which is bigger than that with application of BA-NNs (0.0219). The reason resulting in this situation is that BA-NNs normally over-fit the training data, so that the MSE value is much smaller. Regarding to model performance on unseen validation data, MSE value of BA-NNs is much bigger than that from BA-ELM,

which means the generalisation performance of BA-ELM is better. It is concluded that BA-ELM is able to train faster and perform better than BA-NNs.

<i>Learning Algorithm</i>	<i>Time (CPU time, s)</i>		<i>Training accuracy(MSE)</i>	<i>Validation accuracy(MSE)</i>
	<i>Training time (2nd batch)</i>	<i>Verifying time</i>		
BA-ELM	163.4422	0.7176	0.0488	0.0441
BA-NNs	1726.4	0.2964	0.0219	0.0771

Table 4.1: Performance comparison of BA-ELM and BA-NNs for CO₂ production rate.

The MSE values of individual ELM model for predicting CO₂ production rate can be seen in figure 4.5. The performance on the unseen validation data is not in accordance with that on the training and testing data. For instance, the prediction on the unseen validation data by the 20th ELM is the worst, however, its performance on the training and testing data is better than many of the individual ELM models. This clearly demonstrates that single network has non-robust nature. Nevertheless, when several individual networks are combined together to build the model, the weakness can be addressed easily. Figure 4.6 indicates the MSE values on model building data by aggregating different numbers of ELM models. The first bar in figure 4.6 represents the first individual ELM model shown in figure 4.5, the second bar represents the combination of the first two individual ELM models, and the last bar represents combining all the individual ELM models. Look into the trends of top and bottom plots in figure 4.6, the prediction performance of bootstrap aggregated ELM on the unseen validation data is consistent with that on the training and testing data. In other words, combining several ELM models is able to get more accurate predictions on the training and testing data, as well as on the unseen validation data, than single ELM models. Furthermore, the MSE values in

Figure 4.6 indicates that, the aggregated ELM model provides more accurate predictions than single ELM models, when comparing with the MSE values in figure 4.5.

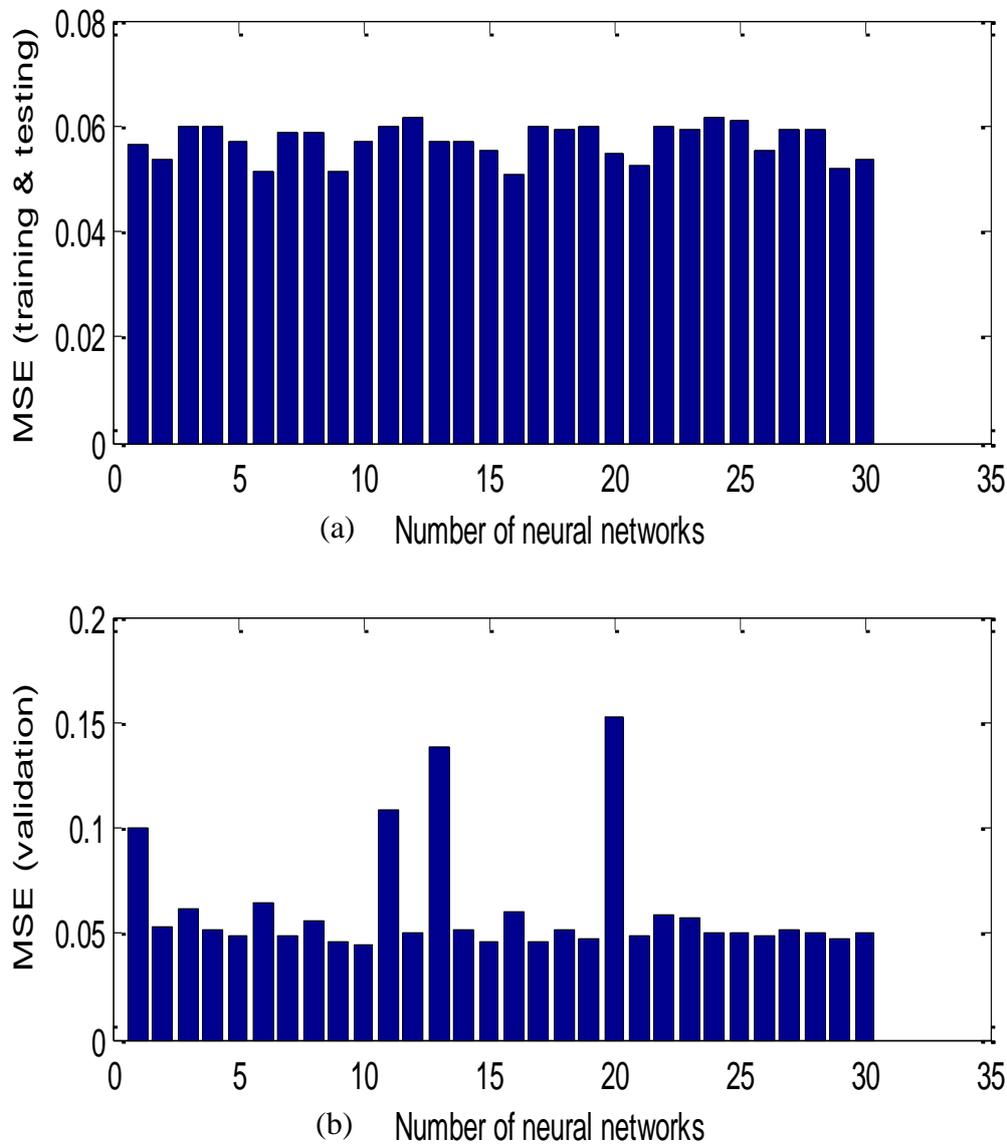


Figure 4.5: MSE of CO₂ production rate for individual ELM models on training & testing data (a) and validation data (b).

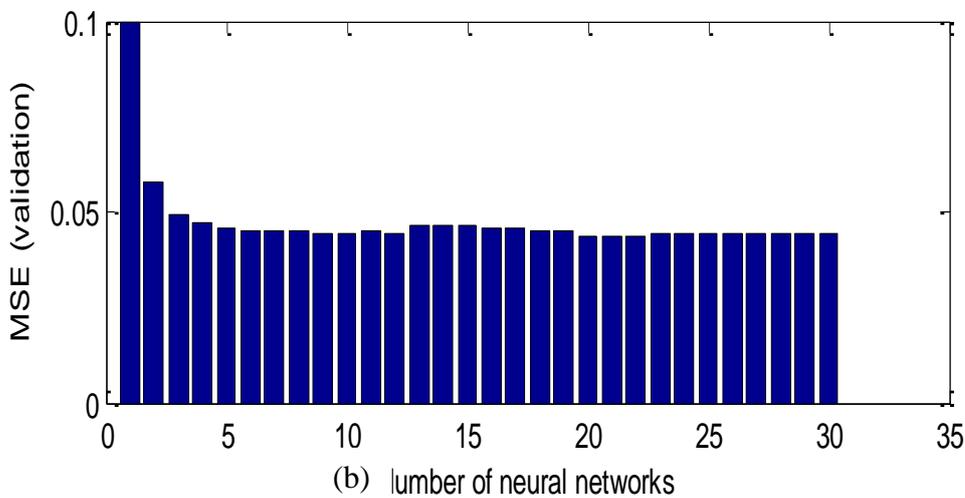
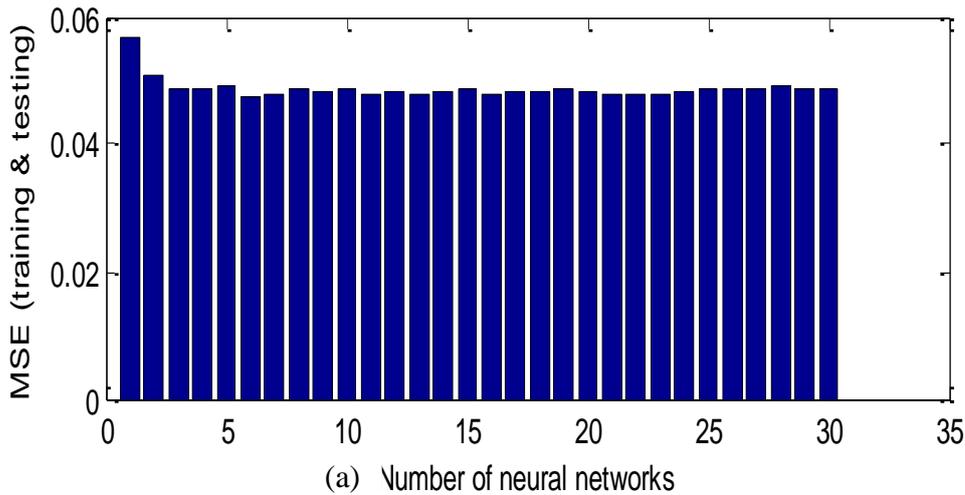
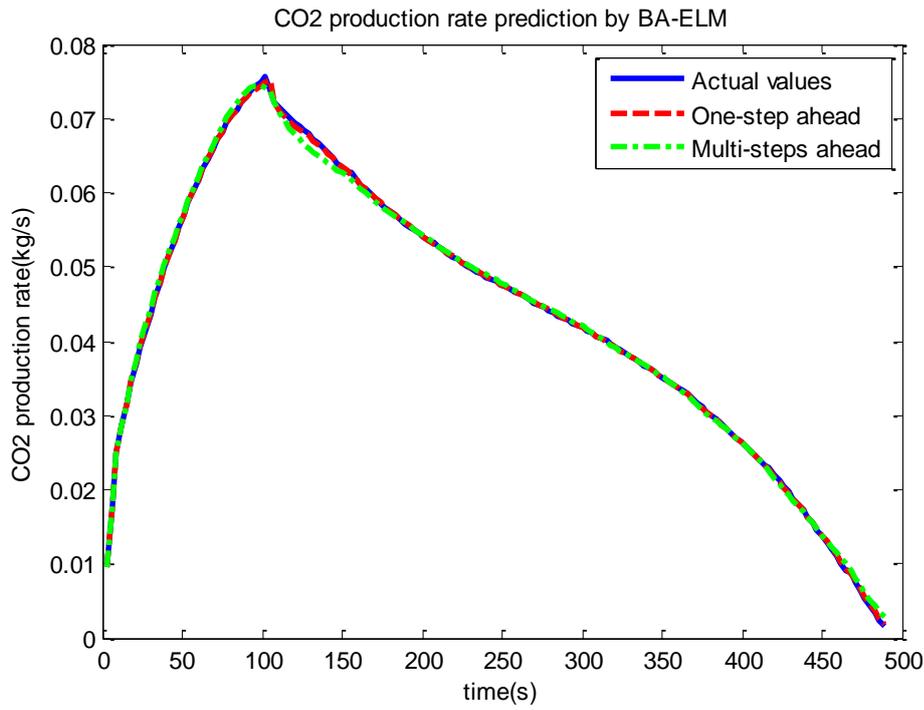
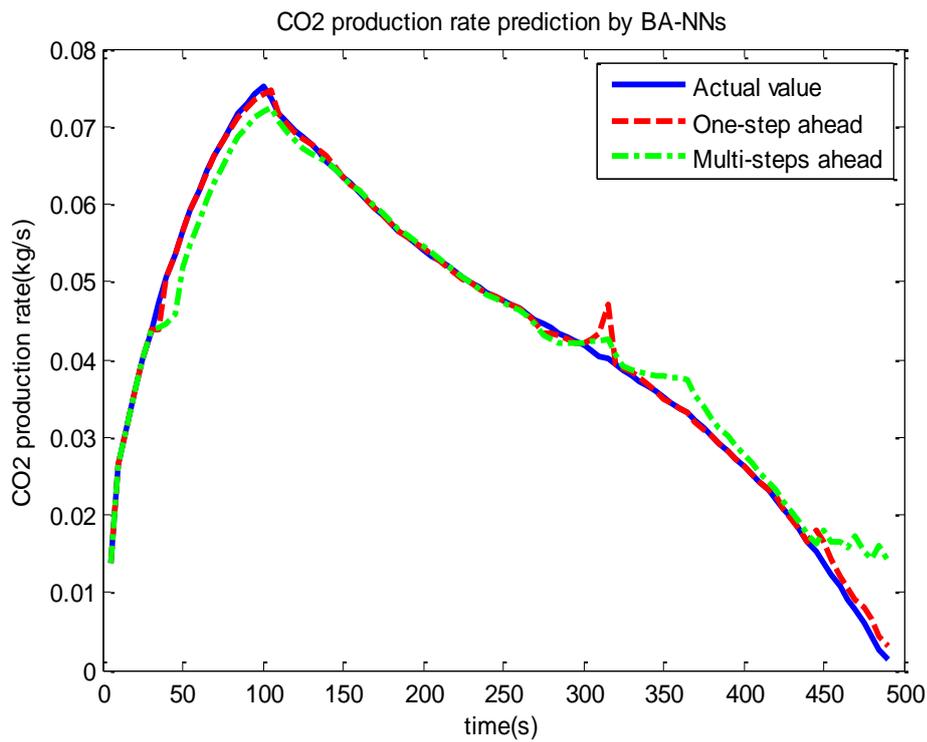


Figure 4.6: MSE of CO₂ production rate for bootstrap aggregated ELM training & testing data (a) and validation data (b).

The performance comparison of in BA-ELM and BA-NNs for one-step ahead and multi-steps ahead predictions of CO₂ production rate is shown figure 4.7. Clearly, the one-step predictions by BA-ELM model are almost consistent with actual values, while those by BA-NNs model are not accurate, especially around 310 s and 450 s. As to the long range predictions, BA-ELM performs in good situation, while BA-NNs can only estimate the outputs 157 steps ahead (see figure 4.7b). The results has agreed with the point suggested by Huang *et al.* (2004), that the generalisation performance of ELM is better than BA-NNs. This is because, BA-ELM model has capability to avoid local optimization, which BA-NNs model usually encounters.



(a)



(b)

Figure 4.7: Dynamic model predictions of CO₂ production rate on Batch 7 using BA-ELM (a) and BA-NNs (b).

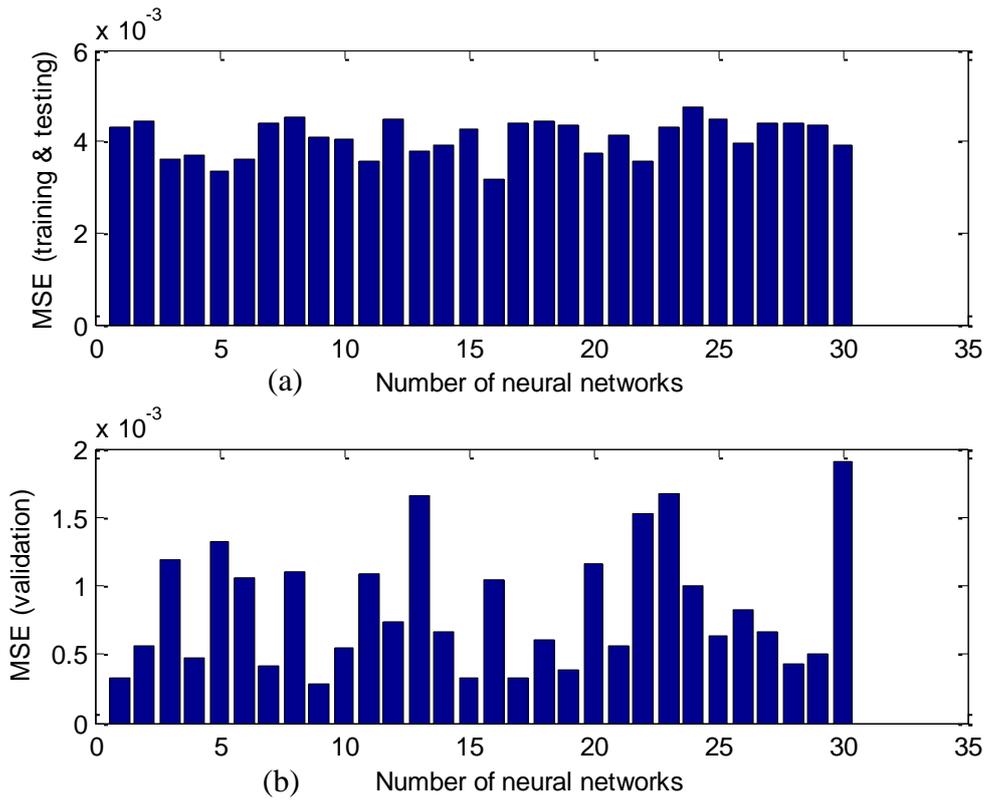


Figure 4.8: MSE of CO₂ capture level for individual ELM models on training & testing data (a) and validation data (b).

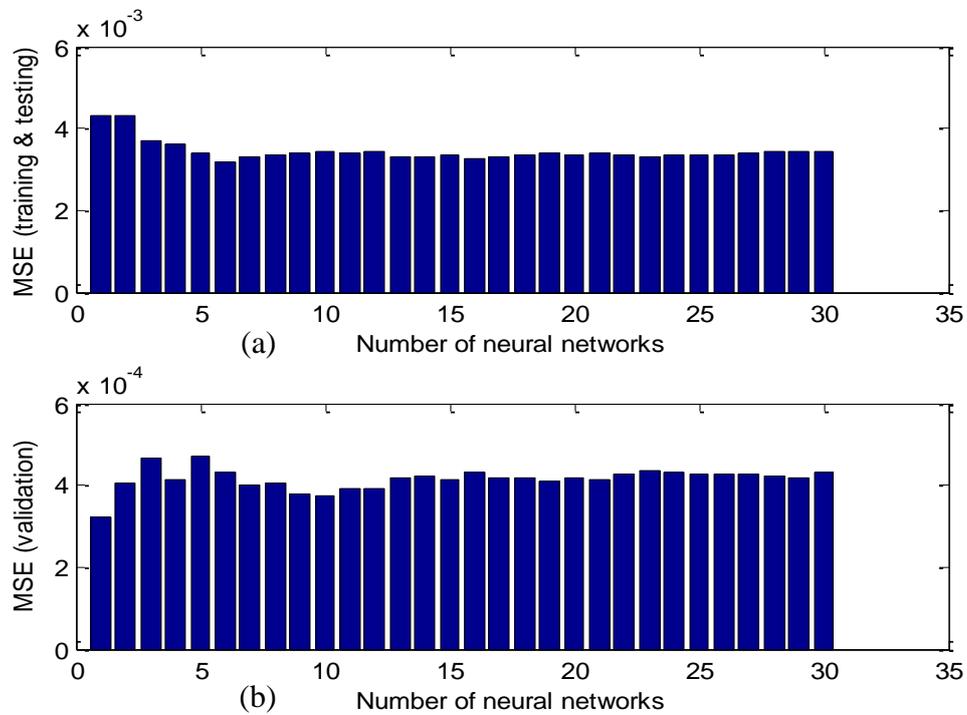
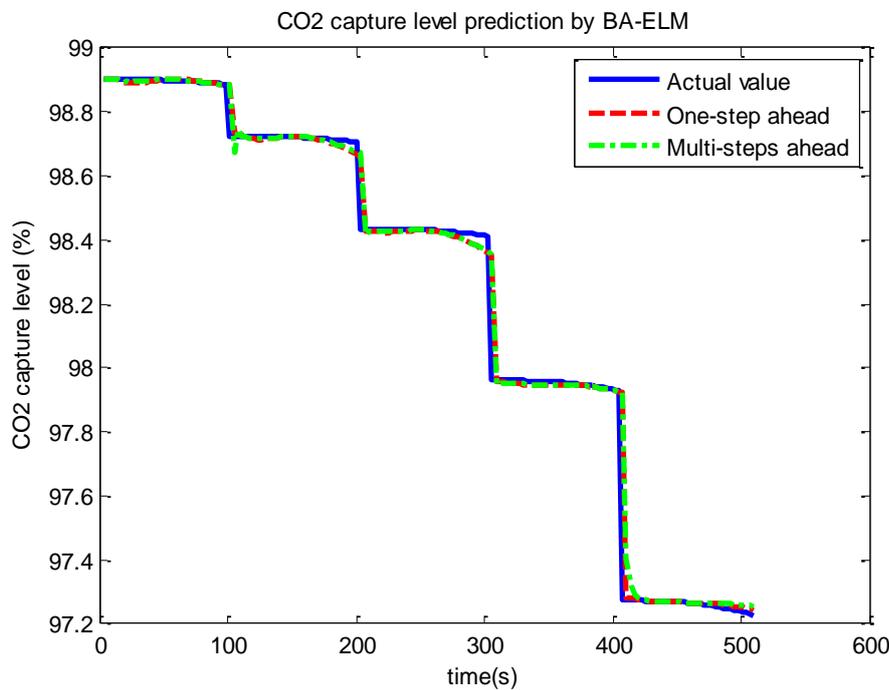


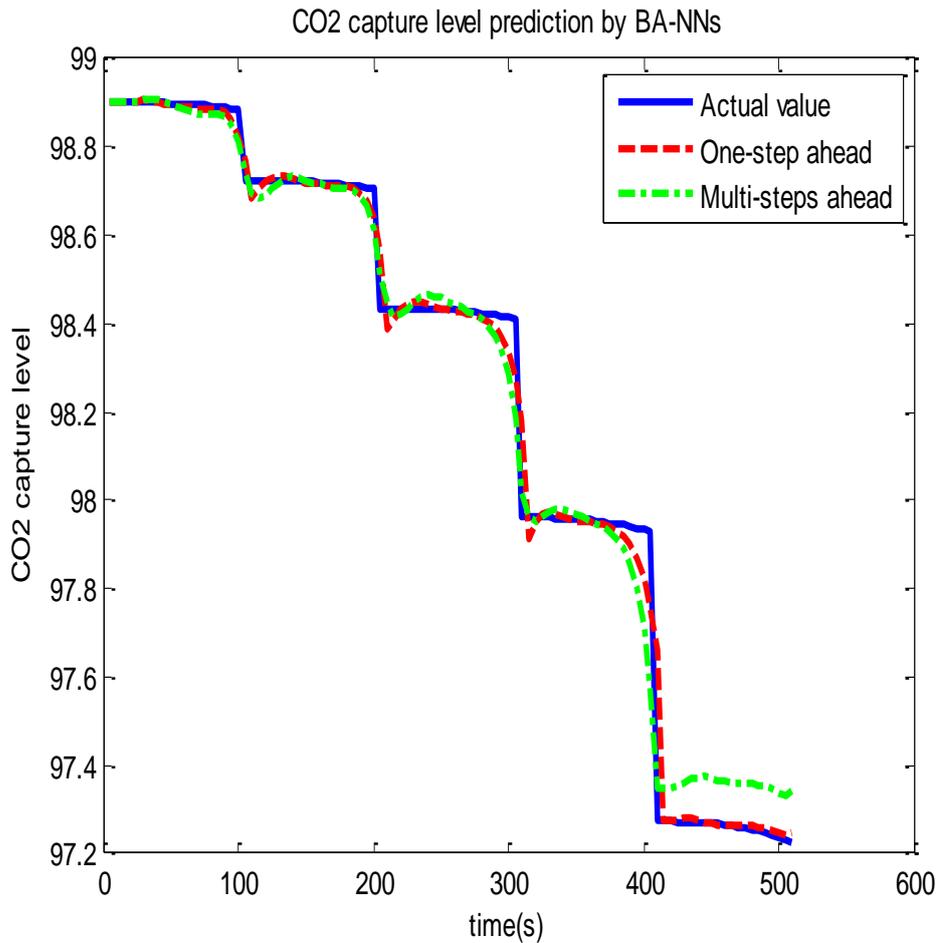
Figure 4.9: MSE of CO₂ capture level for individual ELM models on training & testing data (a) and validation data (b).

Figures 4.8 and 4.9 show the MSE values of the 30 single ELM neural network models and the 30 continuously aggregated models for CO₂ capture level. The result is almost similar with other dynamic model. From the 30 single ELM neural network models, there is a big difference between the MSE values of each model. This also shows that there is a big difference between the accuracy of the model. But in the stacked models, with the increase of the number of the stacked model, the accuracy of the model is gradually stabilized around a constant value.

Figure 4.10 shows the performance comparison of one-step-ahead predictions and multi-step-ahead predictions of CO₂ capture level using BA-ELM and BA-NNs models. It is clear seen from the bottom graph both one-step-ahead predictions and multi-step-ahead predictions from BA-NN are reasonably accurate though some errors are observable, but the long range predictions (green line) are not accurate after 82 steps (410 s). However, in the top graph, the accurate one-step-ahead predictions and multi-step-ahead predictions from BA-ELM are very encouraging, indicating that the model has captured the underlying dynamics of the process. Such accurate long range predictions can be further used for model predictive control and real-time optimisation applications.



(a)



(b)

Figure 4.10: Dynamic model prediction of CO₂ capture level using BA-ELM (a) and BA-NNs (b).

The performance comparison of the bootstrap aggregated neural networks and bootstrap aggregated ELM for CO₂ capture level is shown in table 4.2. The training CPU time of BA-ELM is 6 times lower than that of BA-NN, while the verifying CPU time is a little bit longer than the latter one. This is because each network in the BA-ELM has more hidden neurons than each network in BA-NN. Looking into the comparison of the accuracy, the mean squared error (MSE) values on training data in both models are almost same, while the MSE value of BA-ELM on validation data is 3 times lower than that of BA-NNs. This shows that BA-ELM has a faster training speed and better generalization performance than BA-NNs, which has been proved in (Huang *et al.*, 2006).

<i>Learning Algorithm</i>	<i>Time/CPU time</i>		<i>Training</i>	<i>Validation</i>
	<i>Training time</i> <i>(2nd batch)</i>	<i>Verifying time</i>	<i>accuracy(MSE)</i>	<i>accuracy(MSE)</i>
Bootstrap Aggregated ELM (BA-ELM)	292.8919	0.8112	0.0034	0.00043
Bootstrap Aggregated Neural Networks (BA- NNs)	1902.1	0.5148	0.0030	0.0015

Table 4.2: Performance comparison of BA-ELM and BA-NNs for CO₂ capture level.

4.5 Conclusions

From the theoretical ELM algorithm, we have built two types of BA-ELM models such as static and dynamic models. Two indicators represent the performance of post-combustion CO₂ capture process, such as CO₂ capture rate and CO₂ capture. In this chapter, all of the BA-ELM models indicate good model accuracy and generalisation ability. Also, the BA-ELM model has a good stability in both one-step ahead prediction and multi-steps ahead prediction. Compared with the traditional SLFNNs model, the advantages of ELM model are very prominent. The main point is the training speed improved with thousand times. Not only the training speed increases, but also the good generalization performance could be achieved by the ELM model. Especially for multi-steps ahead prediction, BA-ELM model have a better generalisation ability. Therefore, BA-ELMs is demonstrated as a powerful tool to model the post-combustion CO₂ process, which can be trained much faster and is more accurate than the BA-NNs models. It gives a good generalization performance on unseen data, because the aggregation of multiple ELM can make the model avoid being trapped into local minima and over-trained problems. The model will be used to optimize the CO₂ capture process in the future.

Nevertheless, the BA-ELM still exists some problems. For instance, the number of hidden neurons is quite large, which may affect the model accuracy. If outliers appear in the input data, the model would become unreliable. The BA-ELM model needs a further improvement, as a result of these shortcomings.

Chapter 5. Modelling of a Post-combustion CO₂ Capture Process Using Deep Belief Network

5.1 Introduction

To improve prediction accuracy and generalization ability, Chapter 3 have presented a BANNs model, which aggregates several single neural networks to model post-combustion CO₂ capture process. The modelling technique was found to be able to model the post-combustion CO₂ capture process with higher accuracy and reliability than traditional neural network models. In order to increase training speed and generalisation ability, Chapter 4 has shown a new modelling algorithm, called BA-ELM, which modify the way of calculating output layer weights in SLNNs. Both BANNs and BA-ELM can give model prediction confidence bounds, which is a useful tool to measure prediction reliability. However, the above mentioned learning algorithms have a shallow architecture, i.e., the networks have only one hidden layer. This is because, there is no successful training strategy for NNs with multiple hidden layers before 2006. As reported, there are several deficiencies for shallow neural networks. For examples, the shallow neural networks are limited to represent complex, highly-varying relationship between input and output variables, and easy to converge to local optima (Bengio and LeCun, 2007). Under such circumstances, the deep multi-layer neural networks model was proposed, which was inspired by the structure of human brain. However, due to poor results from gradient-based methods with random initialization the deep multi-layer neural networks have not been applied successfully before 2006. Hence, to develop the training algorithms of deep architecture networks appears to be a challenging problem. To solve the problems, Hinton *et al.* (2006) have put forward a greedy layer-wise unsupervised learning algorithm for Deep Belief Networks (DBN), which pre-train one layer at a time in a greedy way. With the comparison of random initialization, the results of DBN show that the initial parameters of networks are much closer to optimal solutions. Since then, increasing attention has been paid to deep learning and it contributes a lot to image recognition (Liao *et al.*, 2015) and time series forecasting (Ren *et al.*, 2017). However, the application of DBN model on regression of CO₂ capture process has not yet been much exploited.

In this work, a 3-layer DBN which is consisted of two restricted Boltzmann machines (RBMs) is proposed to modelling the inherent process relationship of post-combustion CO₂ capture plant. Theoretically, according to Hinton *et al.* (2006) and Bengio and LeCun (2007), the visible layer of RBM receives the high dimension data and then the hidden layer extracts the features of data between different classes by the connection weights.

The rest of this chapter proposed is arranged as follows. Section 5.2 presents the theoretical knowledge of DBNs and their component layers, Restricted Boltzmann Machines (RBM). Section 5.3 introduces the details of the DBN model for a post-combustion CO₂ capture plant. Then, the comparative result analysis between single-hidden layer feedforward neural networks (SLNNs) and DBNs is revealed in Section 5.4. Finally, Section 5.5 gives conclusions and future works.

5.2 Deep Belief Networks

Many researches have shown that DBNs can produce models with more accuracy and precision, especially with respect to audio and image classification (Liao *et al.*, 2015; Ren *et al.*, 2017; Uddin and Kim, 2017). In this paper, DBNs integrated with neural network model is used to model a post-combustion CO₂ capture process.

5.2.1 Restricted Boltzmann Machines

DBNs are consisted of several Restricted Boltzmann Machines (RBMs), which is a single layer of hidden units that are not connected to each other and have undirected connections to a layer of visible units (see figure 5.1). Theoretically, it is a special type of generative energy based model which can learn probability distribution over its inputs. As there are no connections between hidden units in RBMs, it has an advantage that the hidden unit is conditionally independent to each other. Both visible units (\mathbf{v}) and hidden units (\mathbf{h}) are stochastic binary variable nodes and hypothesis that the joint probability distribution of (\mathbf{v} , \mathbf{h}) fits Boltzmann distribution. \mathbf{v} is connected to \mathbf{h} through undirected weighted connections. The reason why they are restricted is that, there is no connection between hidden variables or visible variables. A probability distribution $p(\mathbf{v}, \mathbf{h})$ is defined via an energy function ($E(\mathbf{v}, \mathbf{h}; \theta)$), which can be written as:

$$-\log P(\mathbf{v}, \mathbf{h}) \propto E(\mathbf{v}, \mathbf{h}; \theta) = -\mathbf{b}^T \mathbf{v} - \mathbf{c}^T \mathbf{h} - \mathbf{h}^T \mathbf{W} \mathbf{v} \quad 5-1$$

where $\theta = (\mathbf{w}, \mathbf{b}, \mathbf{c})$ is the parameter set, \mathbf{W} is the weight vector between visible units and hidden units, and \mathbf{b} and \mathbf{c} are their bias vectors, respectively. Due to the configuration of RBMs, it is possible to compute the conditional probability distribution, when \mathbf{v} and \mathbf{h} are given, as

$$P(h_j | v_i, \theta) = \text{sigm}(\sum_{i=1}^{|\mathbf{v}|} W_{ij} v_i + b_j) \quad 5-2$$

$$P(v_i | h_j, \theta) = \text{sigm}(\sum_{j=1}^{|\mathbf{h}|} W_{ij} h_j + c_i) \quad 5-3$$

where $\text{sigm}(x) = (1/(1 + e^{-x}))$ is the sigmoid function. The parameter $\theta = (\mathbf{w}, \mathbf{b}, \mathbf{c})$ can be calculated using contrastive divergence effectively.

However, the RBMs with binary nodes can only deal with discrete inputs. When inputs are continuous values, Gaussian RBMs are suitable to apply (Shang *et al.*, 2014) as shown in Eq. 5-4.

$$E(\mathbf{v}, \mathbf{h}; \theta) = \sum_i \frac{(v_i - a_i)^2}{2\sigma_i^2} - \mathbf{c}^T \mathbf{h} - \mathbf{h}^T \mathbf{W} \mathbf{v} \quad 5-4$$

where a_i and σ_i are mean and standard deviation respectively of the Gaussian distribution for visible unit i , \mathbf{v} is the continuous valued input layer, and \mathbf{h} is the binary layer.

In some particular situations, since the input data is commonly normalized to zero mean and unit variance, the Gaussian RBMs (Eq. 5-4) is simplified as normalized Gaussian RBMs:

$$E(\mathbf{v}, \mathbf{h}; \theta) = \frac{1}{2} \mathbf{v}^T \mathbf{v} - \mathbf{b}^T \mathbf{v} - \mathbf{c}^T \mathbf{h} - \mathbf{h}^T \mathbf{W} \mathbf{v} \quad 5-5$$

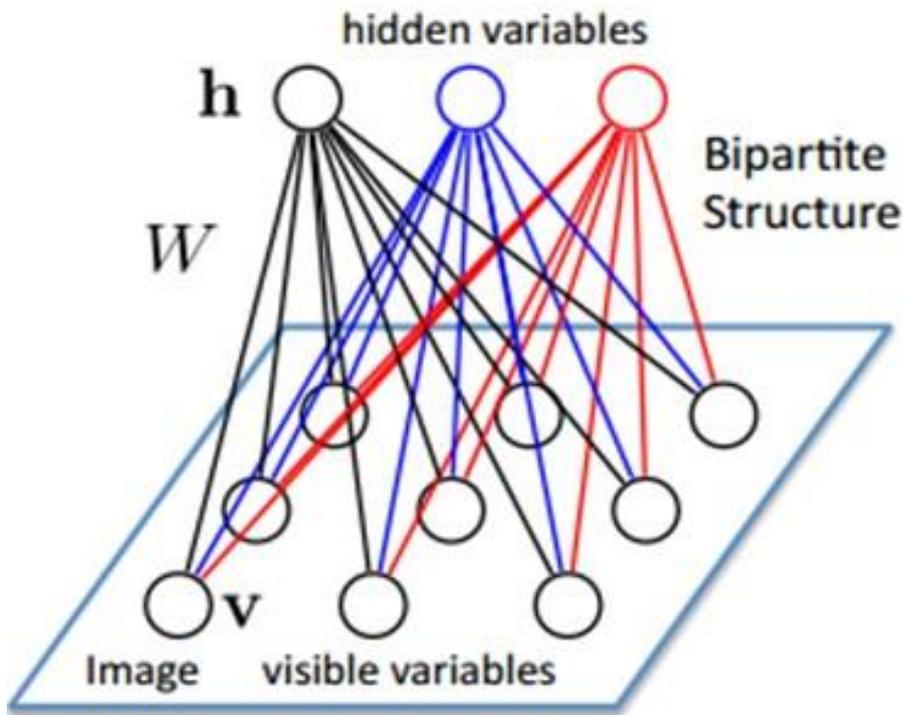


Figure 5.1: The structure of Restricted Boltzmann Machines.

5.2.2 Learning algorithm for RBMs

As to a RBMs model, \mathbf{v} is given and \mathbf{h} is to be estimated. Therefore, learning RBM means to make the probability distribution represented by RBMs ($P(\mathbf{v})$) maximally coincide with training input data, by adjusting parameters $\theta = (\mathbf{w}, \mathbf{b}, \mathbf{c})$.

For $S=\{v^1, v^2, \dots, v^{n_s}\}$, n_s is the number of training inputs, $v^i = (v_1^i, v_2^i, \dots, v_{n_v}^i)^T$, $i=1,2,\dots, n_s$, and they are independent and identically distributed. The objective of training RBM is to maximise the below likelihood function:

$$\ln L_{\theta,s} = \sum_{i=1}^{n_s} \ln P(v^i) \quad 5-6$$

Gradient ascend is a typical method to maximize Eq.5-6 It approaches the optimum via iterations, which can be formed as below:

$$\theta := \theta + \eta \frac{\partial \ln(L_{\theta})}{\partial \theta} \quad 5-7$$

where η is learning rate. In Eq.5-7, the calculation of the gradient $\frac{\partial \ln(L_{\theta})}{\partial \theta}$ is particularly important. To better understand this, the gradient of likelihood function at a single data point v is calculated as:

$$\begin{aligned} \frac{\partial \ln(P(v))}{\partial \theta} &= \frac{\partial}{\partial \theta} (\ln \sum_h e^{-E(v,h)}) - \frac{\partial}{\partial \theta} (\ln \sum_{v,h} e^{-E(\tilde{v},h)}) \\ &= -\frac{1}{\sum_h e^{-E(v,h)}} \sum_h e^{-E(v,h)} \frac{\partial E(v,h)}{\partial \theta} + \frac{1}{\sum_{v,h} e^{-E(\tilde{v},h)}} \sum_{v,h} e^{-E(\tilde{v},h)} \frac{\partial E(\tilde{v},h)}{\partial \theta} \\ &= -\sum_h P(h|v) \frac{\partial E(v,h)}{\partial \theta} + \sum_{v,h} P(\tilde{v},h) \frac{\partial E(\tilde{v},h)}{\partial \theta} \end{aligned} \quad 5-8$$

Note that there are two terms called negative term and positive term in Eq.5-8. The negative term represents the conditional expectation of $\frac{\partial E(v,h)}{\partial \theta}$, given the visible unit v , which is easy to compute.

For binary RBM, the conditional probabilities $P(h|v)$ and $P(v|h)$ are expressed as below:

$$P(h_j = 1|v) = \frac{e^{C_j + W_j v}}{1 + e^{C_j + W_j v}} = \text{sigm}(C_j + W_j v) \quad 5-9$$

$$P(v_j = 1|h) = \frac{e^{b_j + W_j^T h}}{1 + e^{b_j + W_j^T h}} = \text{sigm}(b_j + W_j^T h) \quad 5-10$$

where W_j is the j th row of \mathbf{W} , W_i is the i th column of \mathbf{W} , and $\text{sigm}(v) = 1/[1+\exp(-v)]$ is the sigmoid function. The inputs \mathbf{v} and hidden units \mathbf{h} of binary RBM is symmetrical.

For Gaussian RBM, the conditional probabilities $P(\mathbf{h}|\mathbf{v})$ and $P(\mathbf{v}|\mathbf{h})$ are stated as below:

$$P(v_i|h) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(v_i - b_i - W_i^T \mathbf{h})^2\right\} \sim \mathcal{N}(b_i + W_i^T \mathbf{h}, 1) \quad 5-11$$

However, the computation of the positive term, which is expectation of $\frac{\partial E(\tilde{\mathbf{v}}, \mathbf{h})}{\partial \theta}$ for joint distribution $P(\tilde{\mathbf{v}}, \mathbf{h})$, is an intractable problem. It is causally linked to $(2^{n_v} + n_h)$ items in $\Sigma_{\tilde{\mathbf{v}}, \mathbf{h}}$, giving rise to computation complexity of $O(n_v + n_h)$. Therefore, Gibbs Markov Chain on the pair of variables is usually considered to resolve the problem. However, it is still intricate, because a large quantity of frequency samples is always required to guarantee the precision. Hinton *et al.* (2006) proposed an idea of Contrastive Divergence, which takes initial sample $\mathbf{v}_0 = \mathbf{x}$ sampled from the training distribution and arrives the distribution of RBMs with k small steps.

The contrastive divergence (CD) deals with the approximation of positive terms of Eq. 5-8:

$$\sum_{\mathbf{v}, \mathbf{h}} P(\tilde{\mathbf{v}}, \mathbf{h}) \frac{\partial E(\tilde{\mathbf{v}}, \mathbf{h})}{\partial \theta}$$

The two-stage Gibbs sampler is used as an effective approximation approach. The Gibbs Markov chain (\mathbf{v}, \mathbf{h}) is constructed by repeating the following steps:

1. Sample $\mathbf{h}^{(t)}$ from $P(\mathbf{h}|\mathbf{v} = \mathbf{v}^{(t-1)})$;
2. Sample $\mathbf{v}^{(t)}$ from $P(\mathbf{v}|\mathbf{h} = \mathbf{h}^{(t)})$.

where $\mathbf{v}^{(0)} = \mathbf{v}$. The chain will converge to the true joint distribution $P(\mathbf{v}, \mathbf{h})$ if aperiodic and irreducible. That is to say, the $\mathbf{v}^{(\infty)}$ and $\mathbf{h}^{(\infty)}$ are ideally sampled from the joint distribution $P(\mathbf{v}, \mathbf{h})$. However, the Gibbs sampling must be performed all the time to compute new gradients. Figure 5.2 shows the flow chart of Markov chain in two-stage Gibbs sampler and CD for training RBM. For instance, to obtain $\{\mathbf{v}^{(1)}, \mathbf{h}^{(1)}\}$, second-order approximation of the positive term is adopted by CD algorithm.

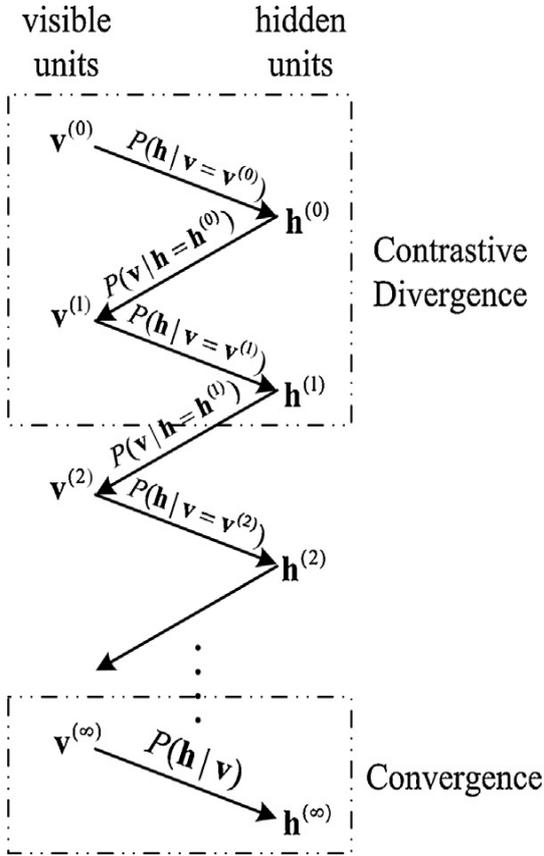


Figure 5.2: Markov chain in two-stage Gibbs sampler and CD for training RBM.

For both binary and Gaussian RBMs, the term $\frac{\partial E(\tilde{v}, h)}{\partial \theta}$ is expressed as below:

$$\frac{\partial E(v, h)}{\partial W_{ij}} = h_j v_i \quad 5-12$$

$$\frac{\partial E(v, h)}{\partial b_i} = v_i \quad 5-13$$

$$\frac{\partial E(v, h)}{\partial c_j} = h_j \quad 5-14$$

The gradient of the log-likelihood function can be stated as :

$$\begin{aligned} \Delta W_{ij} &= \frac{\partial \log P(v)}{\partial W_{ij}} = \sum_h P(\mathbf{h}|\mathbf{v}) \cdot h_j v_i - \sum_{v, h} P(\tilde{v}, h) h_j \tilde{v}_i \approx E(\mathbf{h}_j^{(0)}|\mathbf{v}^{(0)}) \cdot v_i^{(0)} - h_j^{(1)} v_i^{(1)} \approx \\ &E(\mathbf{h}_j^{(0)}|\mathbf{v}^{(0)}) v_i^{(0)} - E(\mathbf{h}_j^{(1)}|\mathbf{v}^{(1)}) \cdot E(v_i^{(1)}|\mathbf{h}^{(0)}) = \text{sigm}(C_j + W_j \mathbf{v}^{(0)}) \cdot v_i^{(0)} - \text{sigm}(C_j + \\ &W_j \mathbf{v}^{(1)}) \cdot \text{sigm}(b_i + W_i^T \mathbf{h}^{(0)}) \end{aligned} \quad 5-15$$

$$\begin{aligned} \Delta b_i &= \frac{\partial \log P(v)}{\partial b_i} = \sum_h P(\mathbf{h}|\mathbf{v}) \cdot h_j - \sum_{v, h} P(\tilde{v}, h) \cdot \tilde{v}_i \approx v_i^{(0)} - E(v_i^{(1)}|\mathbf{h}^{(0)}) = v_i^{(0)} - \text{sigm}(b_i + \\ &W_i^T \mathbf{h}^{(0)}) \end{aligned} \quad 5-16$$

$$\Delta c_j = \frac{\partial \log P(v)}{\partial c_j} = \sum_h P(\mathbf{h}|\mathbf{v}) \cdot h_j - \sum_{v,h} P(\tilde{v}, h) \cdot h_j \approx E(\mathbf{h}_j^{(0)}|\mathbf{v}^{(0)}) - E(\mathbf{h}_j^{(1)}|\mathbf{v}^{(1)}) = \text{sigm}(C_j + W_j \mathbf{v}^{(0)}) - \text{sigm}(C_j + W_j \mathbf{v}^{(1)}) \quad 5-17$$

where $\mathbf{h}^{(0)}$, $\mathbf{v}^{(1)}$ and $\mathbf{h}^{(1)}$ are sampled from the one-step Markov chain, as shown in Fig.5.2. In the above equations, the conditional expectation is used instead of the binary states sampled from the one-step Markov chain.

5.2.3 Unsupervised pre-training and supervised fine tuning of DBN

As the inputs in our case are continuous-valued and not limited to a certain range, the structure of DBN was designed as in figure 5.3. The bottom RBM layer was selected as Gaussian units and the remaining layers are Binary units. The output of each RBM was the extracting feature of previous output. In other words, the high level RBM represents the most representative feature of input data, and the low level RBM is the low-level extraction of input data.

Unsupervised pre-training of DBN is important to improve the model performance. It was interpreted by Bengio and LeCun (2007) as follows: injecting unsupervised training may help to put the parameters of that layer towards the better direction in the parameter space. A greedy layer-wised training algorithm was proposed to train each layer at one time. Specifically, start to learn from the lowest weight matrices and keep all the higher weight matrices tied. In this work, RBM is used to pre-train each layers of DBN networks to lead the initial weights to optimum solution. After the unsupervised steps of DBN are finished, the supervised fine-tuning by back-propagation method is conducted to modify the weights between each different layers. Hinton *et al.* (2006) proposed an idea of wake-up algorithm, which has capability to fine-tune the parameters of all layers together.

In general, the sampling rate of quality variables in chemical process is much slower than process variables, which is shown in figure 5.4. On account of unsupervised pre training, the fast-sampled process data can be fully utilized in building DBN model. With respect to the traditional models such as PLS, SVM and ANN, the numbers of process samples and quality samples should be equal to each other. In other words, only a small number of process samples are used and the rest amount of fast-rate process samples keep unused. However, for DBN model, the process samples abandoned by the traditional models can be used for unsupervised pre-training to extract the latent feature, thereby facilitating the supervised back-propagation process. Therefore, the more process data is used by the DBN model, the more accurate model would be obtained.

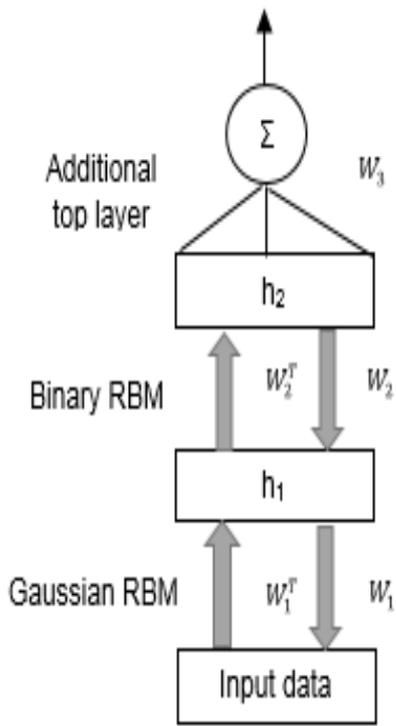


Figure 5.3: The structure of DBN with continuous-valued inputs.

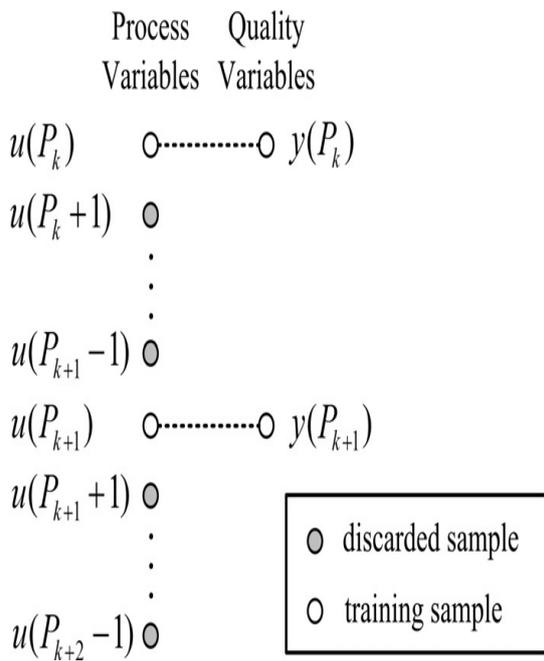


Figure 5.4: Multi-rate sampling in soft sensor modelling. Source: Shang *et al.* (2014).

5.3 Case study: Post-combustion CO₂ capture plant with chemical absorption

An amount of 690 data samples were collected from the gPROMS based simulator at University of Hull. The samples with sampling interval of 5 seconds are used to develop the model. The nonlinear dynamic models in this study are developed as following form:

$$y(t)=f[u(t), u(t-1), u(t-2), y(t-1), y(t-2)] \quad 5-18$$

where y is the process output variables (CO₂ capture level and CO₂ production rate), u represents the process input variables mentioned above, t is discrete time, and $f []$ is the nonlinear function represented by the neural network.

Prior to building the model, the data should be pre-processed to avoid missing values and outliers. As the data has different physical units, each variable should be scaled to zero mean and unit variance. In developing DBN model, all the input data is used for unsupervised training process to extract their feature, which is stated in section 2.3. Then, the data samples are split into three sets: training data (64%), test data (16%) and validation data (20%). To evaluate the performance of the continuous process, the data of Batch 1 are used to predict CO₂ production rate and CO₂ capture level, specifically. Accordingly, two DBN models are constructed for the quality predictions of CO₂ production rate and CO₂ capture level. Cross-validation is used to select the network architecture and both models is found to have the structure of 26-20-17-1. There are 26 input nodes, 20 hidden nodes in the first hidden layer, 17 hidden nodes in the second hidden layer, and 1 output layer node. As the neural network learning is a random process, it is necessary to repeat the training procedure for several times and the result with least training error is selected. In this study, the training procedure is repeated for 20 times.

5.4 Results and discussions

In this study, the performance of DNB modelling technique is compared with traditional neural network modelling technique, namely, single-hidden layer neural network (SLNN). As mentioned above, the structure of 2 hidden layer is determined for DBN, in which the bottom hidden layer is Gaussian RBM and top hidden layer is binary RBM. The neurons in these two hidden layers are 20 and 17 respectively. The learning rates of unsupervised training and supervised training for DBN are selected same as 0.1, to avoid low learning speed and local optimisation. To compare their performance, the learning rate of SLNNs is also set as 0.1, and the hidden layer is consisted of 20 hidden neurons.

As to predicting CO₂ production rate, the mean squared errors (MSE) values on training, validation and testing data of 1st batch are given in table 5.1. In details, the MSE values on training, validation, testing data are 0.0007 kg/s, 0.00047 kg/s and 0.0024kg/s by DBN model, while they are 0.0152kg/s, 0.0411 kg/s and 0.0372 kg/s by SLNNs model. It clearly shows that, DBN model gives much lower MSE values than SLNNs model, when predicting the quality variables. As a result, DBN model has ability to predict more accurately than the SLNN model. This is because DBN can extract the data characteristics by unsupervised learning, thereby accelerate the learning convergence and avoid local minimum. To further prove this point, the data of 1st batch is used to verify the DBN model. Figure 5.5 compares the one-step-ahead prediction performance on CO₂ production rate by SLNN model (top) and DBN model (bottom). The red dashed line represents the prediction values, while the blue solid line represents actual values. It can be seen clearly that the predictions from the DBN model are closer to the actual values than the SLNN model. The MSE values of DBN on 1st batch is 0.0012 kg/s, while that of SLNNs model is 0.0018 kg/s which is slightly higher. When consider long term prediction, figure 5.6 shows their performance comparison on the data of 1st batch. Obviously, both models can predict the quality variables long steps ahead, but DBN model perform much better. The predictions by DBN model are much closer to the actual values, in which the MSE value is only 0.0027 kg/s. Compare to the MSE value of short term prediction, it is not different a lot. However, as to the long term prediction by SLNNs model, the MSE value is 0.009 kg/s, which is 4 times higher than that of short term prediction. This demonstrates that DBN model can not only give the accurate short term predictions, but also predict the long term values with high performance.

	DBN	SLNNs
Training MSE (kg/s)	0.0007	0.0152
Validation MSE (kg/s)	0.00047	0.0411
Testing MSE (kg/s)	0.0024	0.0372

Table 5.1: Comparison of modelling results of DBN and SLNNs on CO₂ production rate.

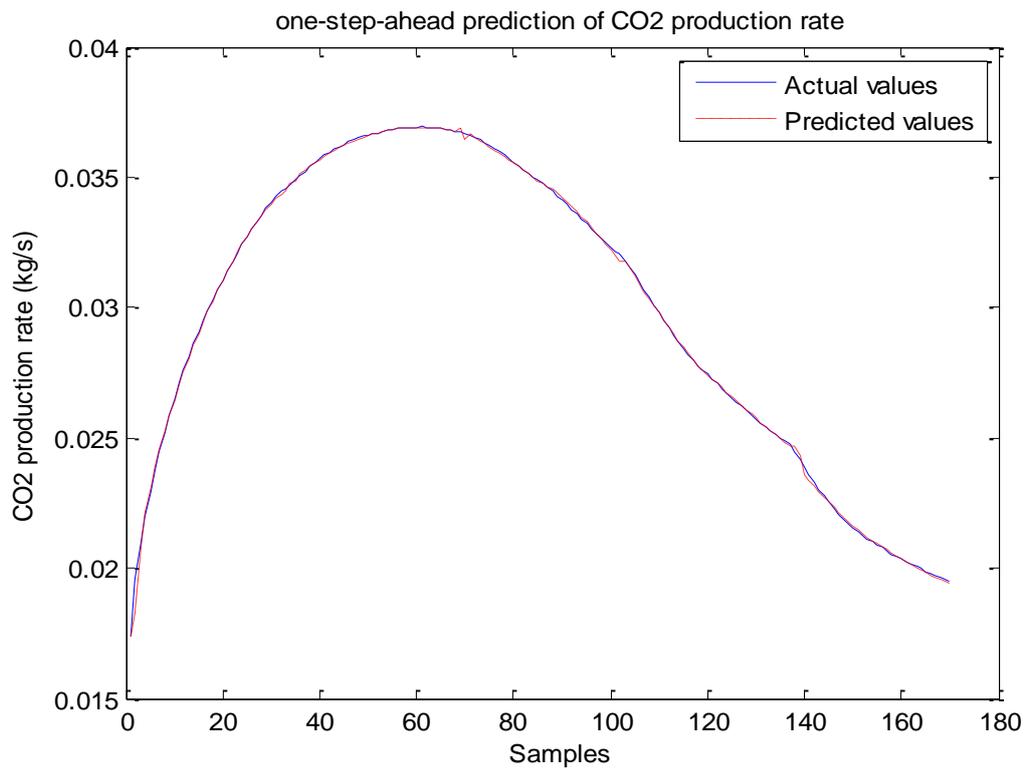
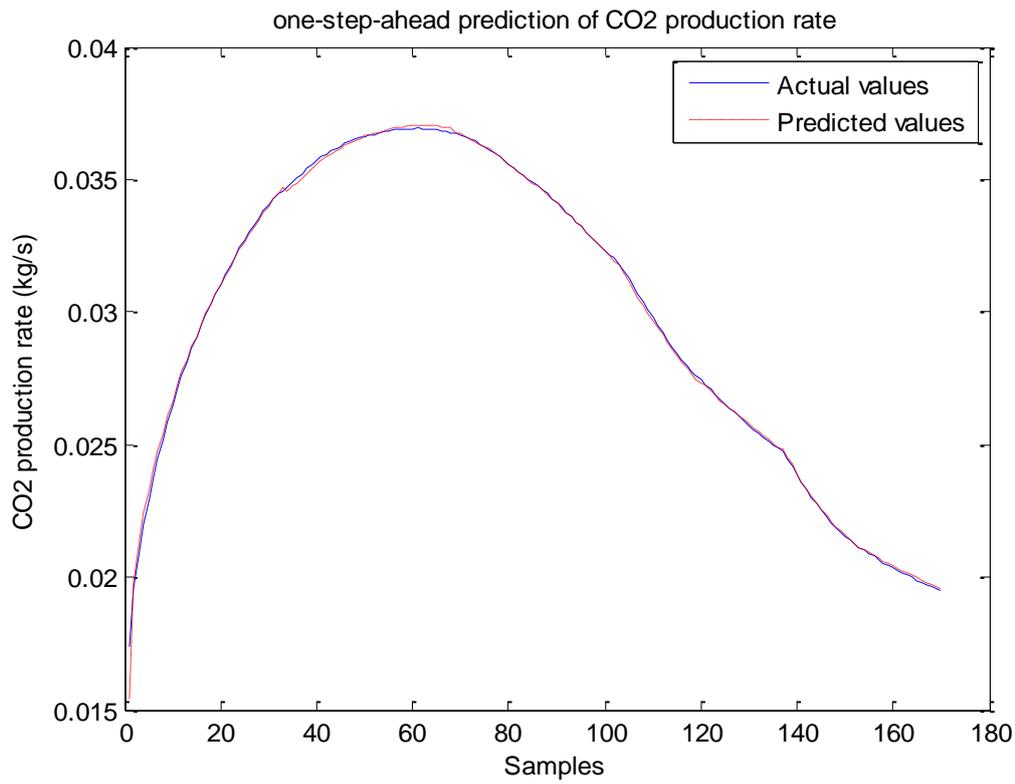


Figure 5.5: One-step-ahead predictions of CO₂ production rate by SLNNs (top) and DBN (bottom).

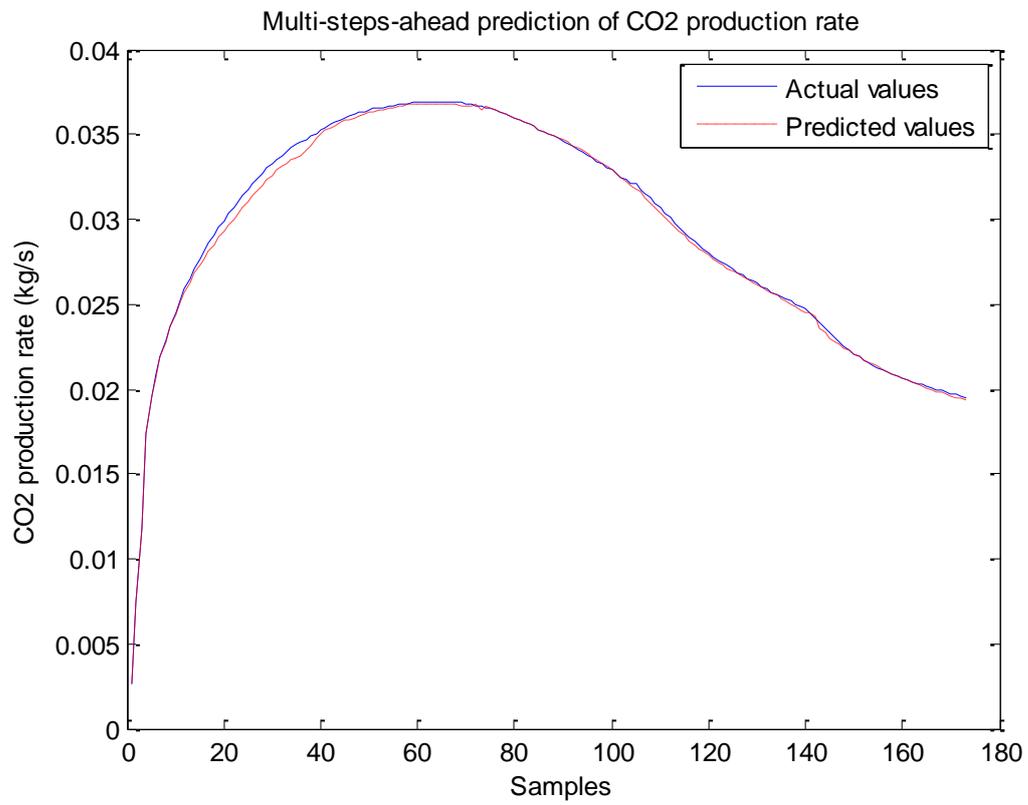
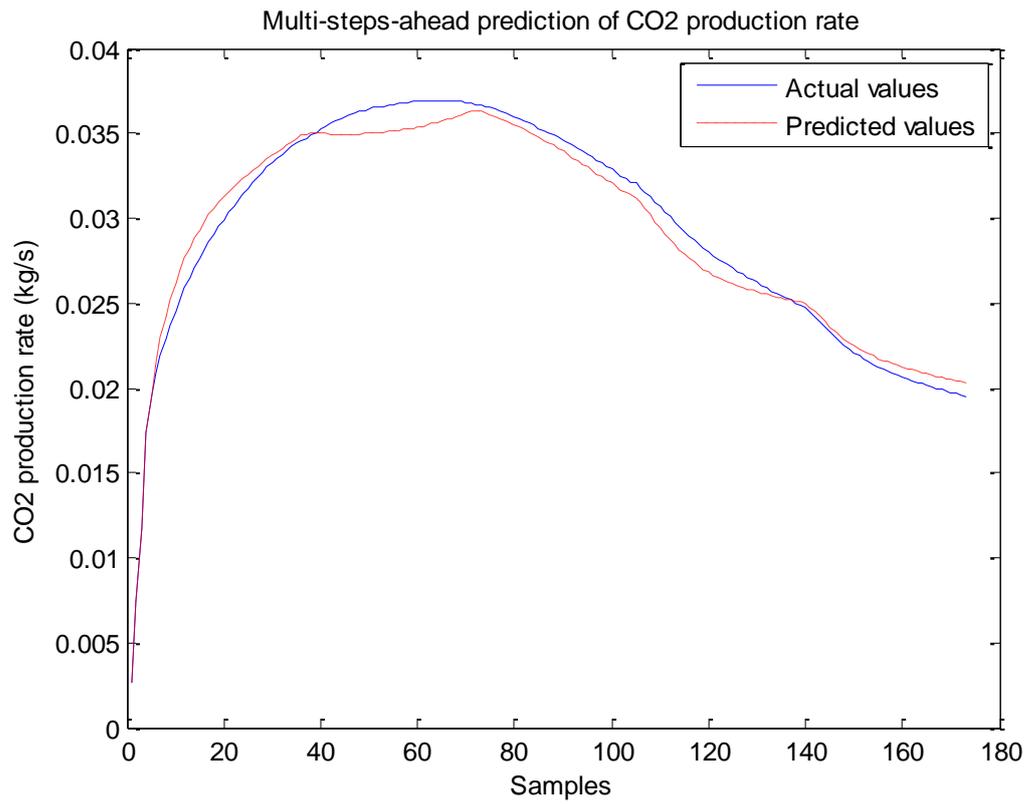


Figure 5.6: Multi-step-ahead predictions of CO₂ production rate by SLNNs (top) and DBN (bottom).

The MSE values of CO₂ capture level prediction by DBN and SLNNs models are given in table 5.2. As can be seen clearly, the MSE values on training, validation and testing data by DBN model are 0.0121%, 0.0171% and 0.0175% respectively. However, as to SLNNs model, the MSE values of all 3 sets are much higher. Especially, the MSE values on validation and testing data by SLNNs model are approximately 10 times higher than those by DBN model. This indicates that DBN model has a better generalisation ability than SLNNs model. The reason is that, the unsupervised training procedure is using principal component analysis (PCA) to analyse the underlying structure of the input data, in which the reduced-dimensionality feature is captured. Therefore, it can extract the most important feature of data and works well in modelling variables.

The one-step-ahead prediction performance comparison of SLNNs and DBN models on 1st batch is shown in figure 5.7. The red dashed line (predictions) is almost identical to the blue solid line (true values) in both plots. However, slightly large prediction errors are seen clearly in the top plot when there are step changes in inputs. Specifically, the MSE values of SLNNs and DBN models are 0.0170% and 0.0141%, respectively, which are not different a lot. This demonstrates that the DBN is able to catch the underlying feature of the data and represent the dynamics of process accurately. Turn to figure 5.8, it shows the long range predictions of CO₂ capture level by SLNNs and DBN models. It can be seen clearly from the top graph, the multi-step-ahead predictions of SLNNs model are not accuracy as a result of extremely large errors. However, in the DBN technique, the long range predictions are much closer to the actual values of CO₂ capture level. In details, the MSE values of long term predictions by SLNNs and DBN models are 0.1929% and 0.0485%, respectively. The latter one is 4 times lower than the former one. It further proves that DBN model is able to predict with higher generalization ability than SLNNs model. For model predictive control and real-time optimisation applications, the long range prediction ability is generally more than significant.

	DBN	SLNNs
Training MSE (%)	0.0121	0.0803
Validation MSE (%)	0.0171	0.2004
Testing MSE (%)	0.0175	0.1400

Table 5.2: Comparison of modelling results of DBN and SLNNs on CO₂ capture level.

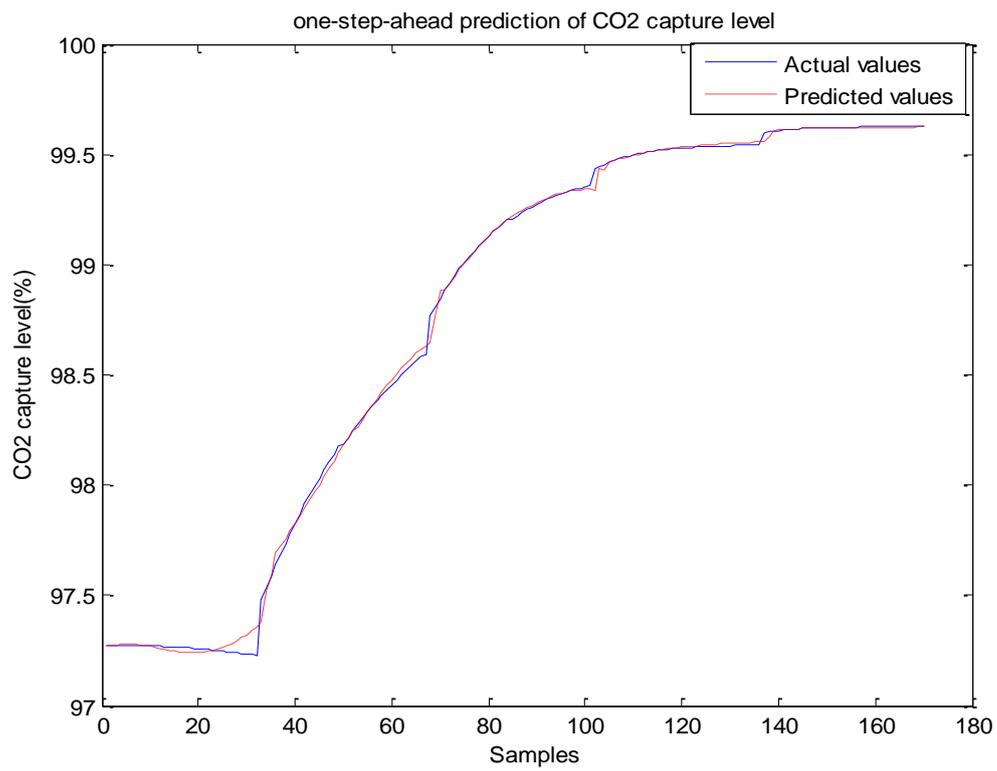
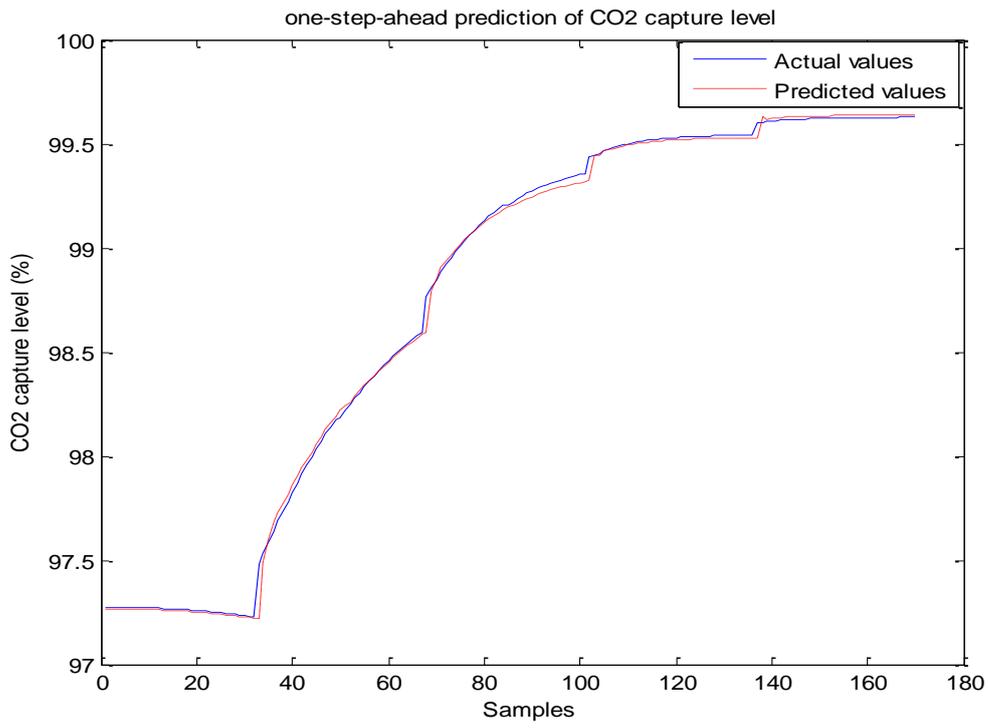


Figure 5.7: One-step-ahead predictions of CO₂ capture level by SLNNs (top) and DBN (bottom).

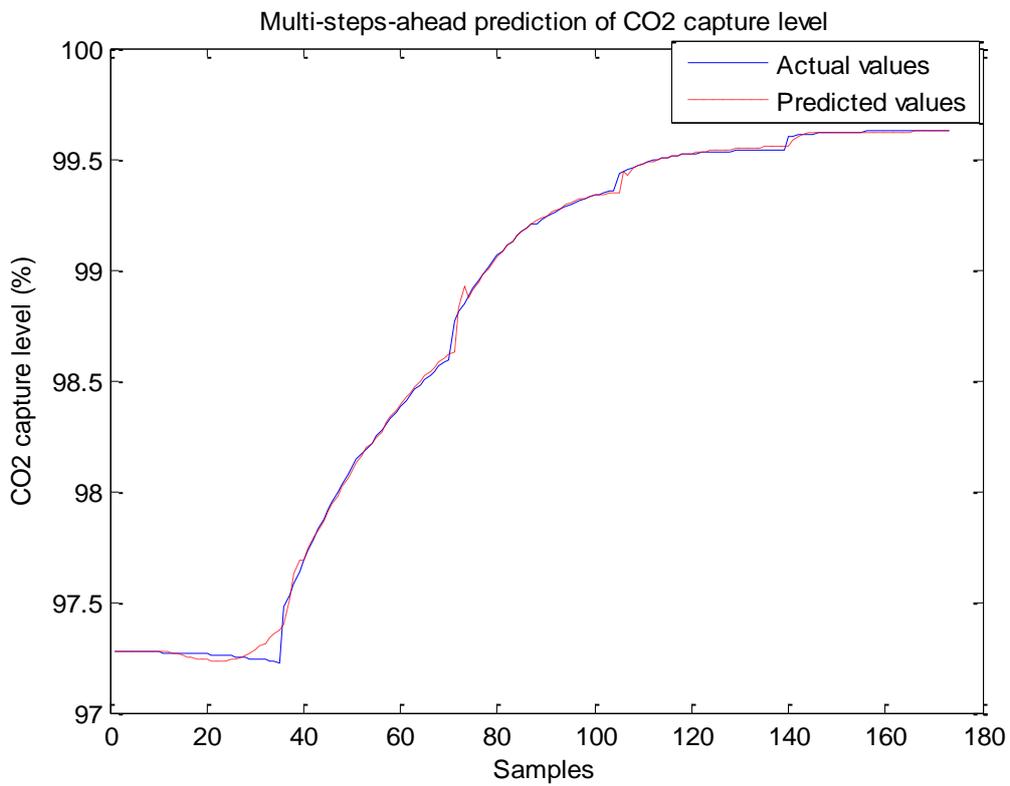
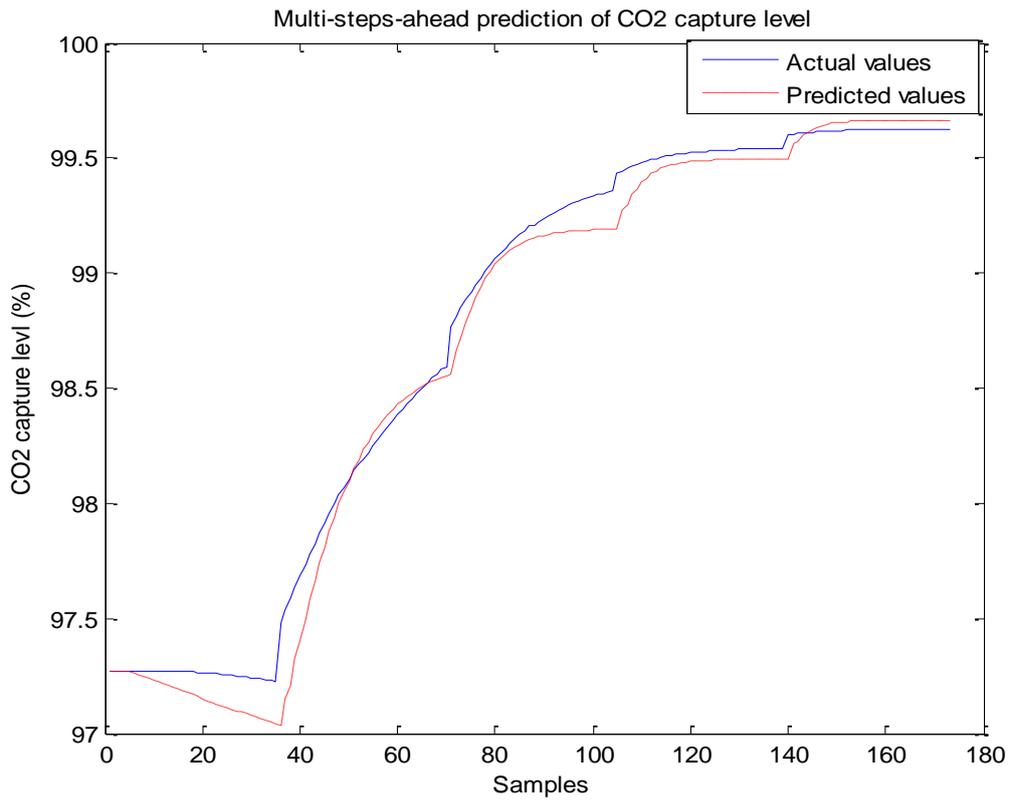


Figure 5.8: Multi-step-ahead predictions of CO₂ capture level by SLNNs (top) and DBN (bottom).

5.5 Conclusions

The deep learning technique is employed as a new modelling method for post-combustion CO₂ capture plant. It is identified to be more accurate than the traditional neural networks as a result of their multi-layer structure. The advantages and characteristics of DBN is analysed, and stated in details in this chapter. The results indicated that DBN can extract nonlinear latent variables, making the neural networks as a latent variable model. Nevertheless, it still exits some problems. For instance, the training of DBN procedure requires much more time than SLNN. In addition, the modelling parameters is expected to be adjusted for the further results improvement.

Chapter 6. Post-combustion CO₂ capture plant control structure selection

6.1 Introduction

In this project, the post-combustion CO₂ capture (PCC) plant is integrated with the supercritical coal-fired power plant, which would possibly be operated at full load in peak hours and part load in off peak hours. These circumstances will lead to changes in inlet gas flow rate. In addition, the upstream power plants are also subjected to shut-up and shut-down scenarios. Due to the aforementioned transient changes in the power plant operation conditions, the performance of PCC process will be affected negatively. In order to cope with the above phenomenon and gain the insights on the transient behaviour of PCC process, a flexible effective operation strategy of PCC plant is required.

In past decades, a large amount of controllability studies have been developed for PCC plant (see section 2.4). Harun *et al.* (2012) and Lin *et al.* (2012) both suggested that CO₂ removal efficiency and lean loading were key parameters to represent the performance of PCC plant, which have been treated as CVs. So, they should be maintained as close as possible to their set-points. In these studies, the results showed the two CVs were controlled by adjusting the lean solvent flow rate and reboiler duty, respectively. Lawal *et al.* (2010) have raised an opinion that the CO₂ removal efficiency is more sensitive to L/G ratio (the ratio of lean amine flow rate to flue gas flow rate) and, therefore L/G ratio should be maintained at a specific value to achieve a desired CO₂ removal efficiency. However, the settling time to reach the desired set point of CO₂ removal efficiency will be longer by using L/G ratio controller (Gaspar *et al.*, 2015). Nittaya *et al.* (2014) have presented three 6×6 control structures based on RGA and heuristic approaches. By using heuristic approach, CO₂ removal efficiency was paired with reboiler heat duty and reboiler temperature was paired with the rich solvent flow rate, thereby reducing the settling time. Panahi and Skogestad (2012) used RGA and dynamic RGA to determine the control loops. The performance of four proposed control structures were evaluated by changing flue gas flow rate. In their study, they also compared the performance of decentralized control structures with a 2×2 MPC. Further, Sahraei and Ricardez-Sandoval (2014) compared a RGA-based control structure to 6×6 MPC scheme, under the step changes in flue gas flow rate, set-point tracking of CO₂ capture level, and constrained heat supply. The results showed that MPC-based control scheme was able to response to disturbance with faster speed and minimum deviation. The variable of energy performance (specific heat duty) was introduced as a key indicator for PCC plant (Luu *et al.*, 2015). They developed three control schemes: a standard PID feedback control scheme, a

cascade-PID control scheme and a MPC control scheme. The performance of these three control schemes were evaluated by set-point tracking and disturbance rejection. Conclusively, MPC strategy was found to be the best control structure to regulate the operation, economic and environmental constraints. In addition, the performance of MPC scheme under the oscillated changes in upstream plant load was investigated by He *et al.* (2016), and an optimal operation scenario was concluded.

This chapter is to apply new methods of GRDA and DRAG to determine the control loops of decentralized control structure, and compare to a model-based MPC scheme. It is consisted of following sections: the system identification introduced in Section 2; the results and discussion of GRDA and DRGA based control loops are presented in Section 3; the results of MPC scheme is discussed in Section 4. Finally, the conclusion is drawn in Section 4.

6.2 Multivariate decentralized control analysis

From the PCC system, the variables to be controlled were selected as CO₂ capture level (y_1) and lean loading (y_2), while the manipulated variables were selected as lean solvent flow rate (u_1) and reboiler temperature (u_2). It is a 2×2 system shown in figure 6.1.

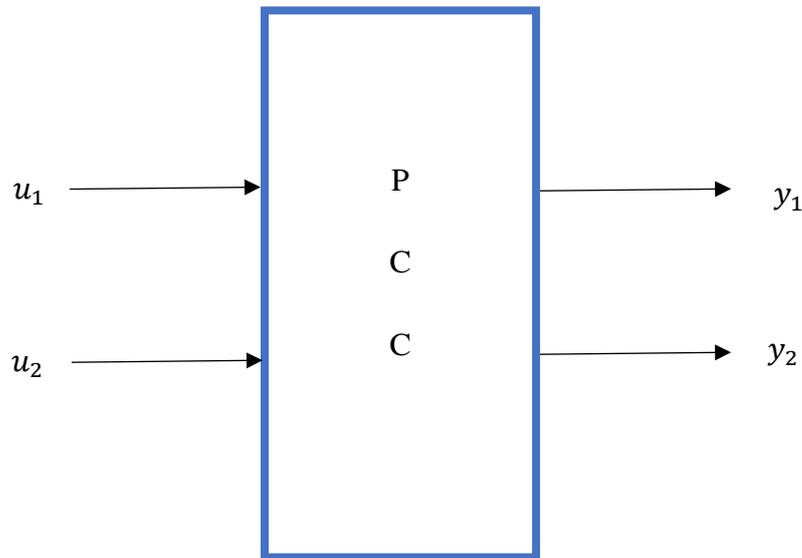


Figure 6.1: A simplified PCC block diagram.

6.2.1 RGA range analysis via optimisation

The transfer function obtained by Matlab coding is shown as follows:

$$\mathbf{G}(s) = \begin{bmatrix} \frac{135.9}{s^3+0.7988s^2+4.963s+1.912} e^{-s} & \frac{0.06245}{s+0.1248} e^{-3s} \\ \frac{0.0005328 s^2+0.004752s+0.00003363}{s^3+0.4361s^2+0.3351s+0.01602} e^{-s} & \frac{-0.001015}{s+0.06413} e^{-3s} \end{bmatrix} \quad 6-1$$

The nominal steady state gain matrix \mathbf{K} is given as:

$$\mathbf{K} = \mathbf{G}(0) = \begin{bmatrix} 71 & 0.5 \\ 0.0022 & -0.0158 \end{bmatrix} \quad 6-2$$

The expression to obtain RGA from steady state gain matrix \mathbf{K} is shown as:

$$\text{RGA} = \mathbf{K} \otimes (\mathbf{K}^{-1})^T \quad 6-3$$

where \otimes represents multiplication of element by element.

According to the above equation, RGA values are:

$$\text{RGA} = \begin{bmatrix} 0.999 & 0.001 \\ 0.001 & 0.999 \end{bmatrix} \quad 6-4$$

The RGA values suggest the diagonal pairing control structure ($y_1 - u_1, y_2 - u_2$) is selected, because all the diagonal values are positive and close to 1.

As RGA is calculated by the steady-state gains, the model uncertainties will result in the uncertainties of RGA. In this chapter, the worst-case bound has been applied to describe the RGA uncertainties.

According to Chen and Seborg (2002), the expression of uncertainty for each steady state gain is assumed as:

$$|\Delta K_{ij}| \leq \alpha |\hat{K}_{ij}| \quad 6-5$$

Case 1: $\alpha = 0.01$

The uncertainty ranges of RGA elements are calculated by optimisation method according to Section 2.4.1, which are show as below:

$$\begin{bmatrix} 0.99899 \leq \lambda_{11} \leq 0.99905 & 0.00095 \leq \lambda_{12} \leq 0.00101 \\ 0.00095 \leq \lambda_{21} \leq 0.00101 & 0.99899 \leq \lambda_{22} \leq 0.99905 \end{bmatrix} \quad 6-6$$

From Eq.6-6, the pair selection is $y_1 - u_1, y_2 - u_2$, which is similar to the above conclusion. Table 6-1 shows the steady state gains \mathbf{K} corresponding to their maximum and minimum values of RGA elements, within constraint $\alpha = 0.01$.

λ_{ij}	\mathbf{K}
$\lambda_{11}\text{min} = 0.99899$	$\begin{bmatrix} 71 & 0.5050 \\ 0.002222 & -0.015642 \end{bmatrix}$
$\lambda_{11}\text{max} = 0.99905$	$\begin{bmatrix} 71 & 0.4950 \\ 0.002178 & -0.015958 \end{bmatrix}$
$\lambda_{12}\text{min} = 0.00095$	$\begin{bmatrix} 71 & 0.4950 \\ 0.002178 & -0.015958 \end{bmatrix}$
$\lambda_{12}\text{max} = 0.00101$	$\begin{bmatrix} 71 & 0.5050 \\ 0.002222 & -0.015642 \end{bmatrix}$
$\lambda_{21}\text{min} = 0.00095$	$\begin{bmatrix} 71 & 0.4950 \\ 0.002178 & -0.015958 \end{bmatrix}$
$\lambda_{21}\text{max} = 0.00101$	$\begin{bmatrix} 71 & 0.5050 \\ 0.002222 & -0.015642 \end{bmatrix}$
$\lambda_{22}\text{min} = 0.99899$	$\begin{bmatrix} 71 & 0.5050 \\ 0.002222 & -0.015642 \end{bmatrix}$
$\lambda_{22}\text{max} = 0.99905$	$\begin{bmatrix} 71 & 0.4950 \\ 0.002178 & -0.015958 \end{bmatrix}$

Table 6.1: Lower band and upper band of RGA and their corresponding steady state gains for $\alpha = 0.01$.

Case 2: $\alpha = 0.1$

The uncertainty ranges of RGA elements calculated via optimisation method are show as below:

$$\begin{bmatrix} 0.9987 \leq \lambda_{11} \leq 0.9993 & 0.0007 \leq \lambda_{12} \leq 0.0013 \\ 0.0007 \leq \lambda_{21} \leq 0.0013 & 0.9987 \leq \lambda_{22} \leq 0.9993 \end{bmatrix} \quad 6-7$$

As to 0.1 uncertainty range of model, the values of RGA elements suggest the sample pair selection as mentioned above. Again, the steady state gains \mathbf{K} corresponding to lower band and upper band of RGA is shown in Table 6-2.

λ_{ij}	K
$\lambda_{11}\text{min} = 0.9987$	$\begin{bmatrix} 71 & 0.5500 \\ 0.00242 & -0.01422 \end{bmatrix}$
$\lambda_{11}\text{max} = 0.9993$	$\begin{bmatrix} 71 & 0.4500 \\ 0.00198 & -0.01738 \end{bmatrix}$
$\lambda_{12}\text{min} = 0.0007$	$\begin{bmatrix} 71 & 0.4500 \\ 0.00198 & -0.01738 \end{bmatrix}$
$\lambda_{12}\text{max} = 0.0013$	$\begin{bmatrix} 71 & 0.5500 \\ 0.00242 & -0.01422 \end{bmatrix}$
$\lambda_{21}\text{min} = 0.0007$	$\begin{bmatrix} 71 & 0.4500 \\ 0.00198 & -0.01738 \end{bmatrix}$
$\lambda_{21}\text{max} = 0.0013$	$\begin{bmatrix} 71 & 0.5500 \\ 0.00242 & -0.01422 \end{bmatrix}$
$\lambda_{22}\text{min} = 0.9987$	$\begin{bmatrix} 71 & 0.5500 \\ 0.00242 & -0.01422 \end{bmatrix}$
$\lambda_{22}\text{max} = 0.9993$	$\begin{bmatrix} 71 & 0.4500 \\ 0.00198 & -0.01738 \end{bmatrix}$

Table 6.2: Lower band and upper band of RGA and their corresponding steady state gains for $\alpha = 0.1$.

Case 3: $\alpha = 0.25$

The uncertainty ranges of RGA elements calculated via optimisation method are show as below:

$$\begin{bmatrix} 0.9980 \leq \lambda_{11} \leq 0.9996 & 0.0004 \leq \lambda_{12} \leq 0.0020 \\ 0.0004 \leq \lambda_{21} \leq 0.0020 & 0.9980 \leq \lambda_{22} \leq 0.9996 \end{bmatrix} \quad 6-8$$

As to 0.25 uncertainty ranges of model, the ranges of RGA elements are wider, while it is still recommended that the pairs selection is determined as $y_1 - u_1, y_2 - u_2$. The details of steady stated gains corresponding to the RGA bonds are shown as below:

λ_{ij}	\mathbf{K}
$\lambda_{11\min} = 0.9980$	$\begin{bmatrix} 71 & 0.625 \\ 0.00275 & -0.01185 \end{bmatrix}$
$\lambda_{11\max} = 0.9996$	$\begin{bmatrix} 71 & 0.3750 \\ 0.00165 & -0.01975 \end{bmatrix}$
$\lambda_{12\min} = \mathbf{0.0004}$	$\begin{bmatrix} 71 & 0.3750 \\ 0.00165 & -0.01975 \end{bmatrix}$
$\lambda_{12\max} = \mathbf{0.0020}$	$\begin{bmatrix} 71 & 0.625 \\ 0.00275 & -0.01185 \end{bmatrix}$
$\lambda_{21\min} = \mathbf{0.0004}$	$\begin{bmatrix} 71 & 0.3750 \\ 0.00165 & -0.01975 \end{bmatrix}$
$\lambda_{21\max} = \mathbf{0.0020}$	$\begin{bmatrix} 71 & 0.625 \\ 0.00275 & -0.01185 \end{bmatrix}$
$\lambda_{22\min} = 0.9980$	$\begin{bmatrix} 71 & 0.625 \\ 0.00275 & -0.01185 \end{bmatrix}$
$\lambda_{22\max} = 0.9996$	$\begin{bmatrix} 71 & 0.3750 \\ 0.00165 & -0.01975 \end{bmatrix}$

Table 6.3: lower band and upper band of RGA and their corresponding steady state gains for $\alpha = 0.25$.

6.2.2 RDGA range analysis via optimisation

The step change of flue gas flow rate is introduced as a disturbance to the system. The disturbance transfer function obtained by Matlab coding, $G_d(s)$, is expressed as below

$$\mathbf{G}_d(s) = \begin{bmatrix} \frac{-48.23}{s+0.9619} \\ \frac{-0.003255s^2+0.005436s+0.0001026}{s^3+0.3887s^2+0.2659s+0.005797} \end{bmatrix} \quad 6-9$$

The steady state vector gain matrix can be obtained as:

$$\mathbf{K}_d = \mathbf{G}_d(0) = \begin{bmatrix} -50.1403 \\ 0.0177 \end{bmatrix} \quad 6-10$$

The nominal RDGA can be calculated according to Eq. 2-52 and 2-53:

$$\text{RDGA} = \begin{bmatrix} 0.9879 & 0.0121 \\ -0.0867 & 1.0867 \end{bmatrix} \quad 6-11$$

Also, the three cases of uncertainty bounds was considered for RDGA matrix, which is similar to RGA analysis.

Case 1: $\alpha = 0.01$

$$\begin{bmatrix} 0.9877 \leq \beta_{11} \leq 0.9880 & 0.012 \leq \beta_{12} \leq 0.0123 \\ -0.0903 \leq \beta_{21} \leq -0.0833 & 1.0833 \leq \beta_{22} \leq 1.0903 \end{bmatrix} \quad 6-12$$

Case 2: $\alpha = 0.1$

$$\begin{bmatrix} 0.9864 \leq \beta_{11} \leq 0.9892 & 0.0108 \leq \beta_{12} \leq 0.0136 \\ -0.13 \leq \beta_{21} \leq -0.0577 & 1.0577 \leq \beta_{22} \leq 1.13 \end{bmatrix} \quad 6-13$$

Case 3: $\alpha = 0.25$

$$\begin{bmatrix} 0.9840 \leq \beta_{11} \leq 0.9920 & 0.008 \leq \beta_{12} \leq 0.016 \\ -0.2334 \leq \beta_{21} \leq -0.034 & 1.034 \leq \beta_{22} \leq 1.2334 \end{bmatrix} \quad 6-14$$

To measure the interaction for each loops with the disturbance injection, GRDG was applied. It is able to evaluate the load effect under a specific controller structure (closed-loop load effect) over the open load effect.

$$\delta_i = \sum_{j=1}^n \beta_{ij} \gamma_{ij} \quad 6-15$$

where β_{ij} is the element of RDGA matrix and γ_{ij} is the element of a structure election matrix, Γ , shown as below:

$$\Gamma = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1n} \\ \vdots & \ddots & \vdots \\ \gamma_{n1} & \cdots & \gamma_{nn} \end{bmatrix} \quad 6-16$$

where

$\gamma_{ij} = 1$, element is picked up for the controller structure

$\gamma_{ij} = 0$, element ignored.

For this process, the structure selection matrix Γ is specified as:

- $\Gamma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\text{GRDG} = [0.9879, 1.0867]^T$
- $\Gamma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\text{GRDG} = [0.0121, -0.0867]^T$

According to GRDG rule, the small positive value of elements is preferable because it is the ratio of net load effect over the open loop load effect. Therefore, the pairs of $y_1 - u_1$ and $y_2 - u_2$ are selected, which is same as RGA.

6.2.3 DRGA analysis

On many occasions, the steady state RGA has no access to the dynamic information for the control structures. Hence, it is extended to frequency-dependent RGA in a straightforward way, shown as:

$$\text{DRGA}(s) = \mathbf{K}(s) \otimes (\mathbf{K}(s)^{-1})^T \quad 6-17$$

The definition here is exactly the same as Eq.6-3, except that DRGA is a function of frequency by setting $s=i\omega$. The perfect control is assumed in the situation as same as RGA.

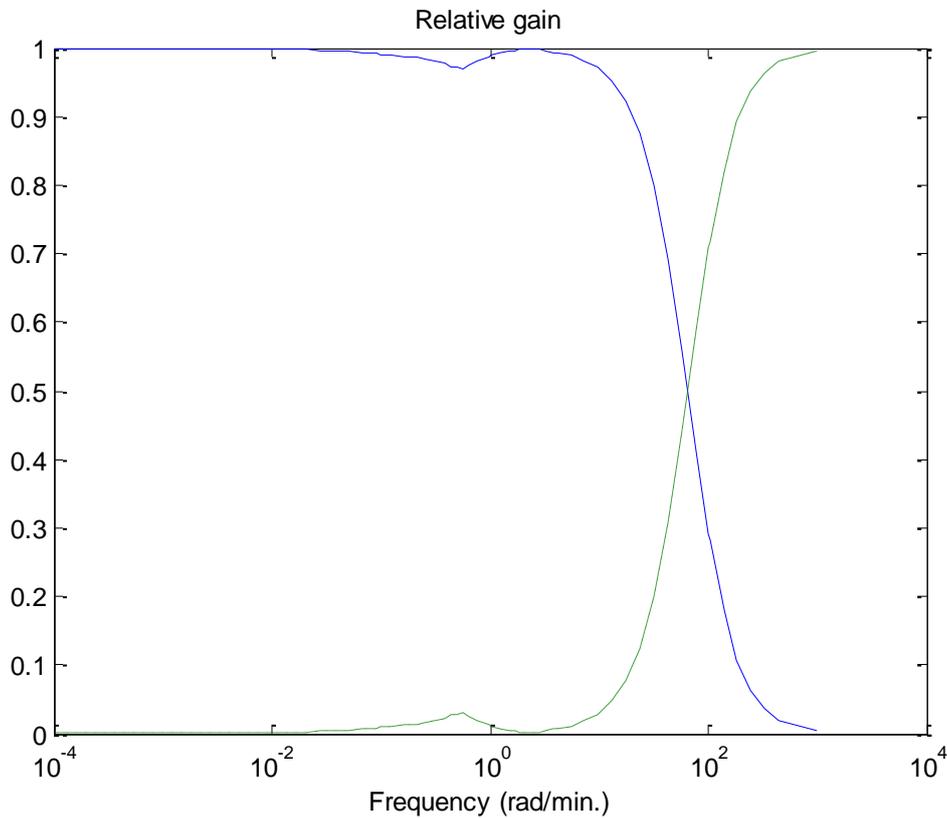


Figure 6.2: Frequency-dependent RGA analysis.

The magnitudes of the RGA generated at different frequencies are depicted in figure 6.2. The solid line represents the RGA value of diagonal pair, while the dashed line is the RGA value of off-diagonal pair. From figure 6.2, the frequencies in the range $1 < \omega < 10^2$ are of particular interest. For the post-combustion CO_2 capture process, the possible pairing recommendations are given in the following table. It is demonstrated that DRGA suggests

same pairing as steady-state RGA at low frequencies, while at high frequencies the pairing changes corresponding to different closed-loop characteristics (see table 6.4).

Frequency range (rad/min)	Steady-state RGA	Dynamic RGA
Low ($\omega < 10^2$)	$y_1 - u_1, y_2 - u_2$	$y_1 - u_1, y_2 - u_2$
High ($\omega > 10^2$)	$y_1 - u_1, y_2 - u_2$	$y_1 - u_2, y_2 - u_1$

Table 6.4: Recommended pairing for a frequency range.

6.3 Results and analysis of decentralized control structure

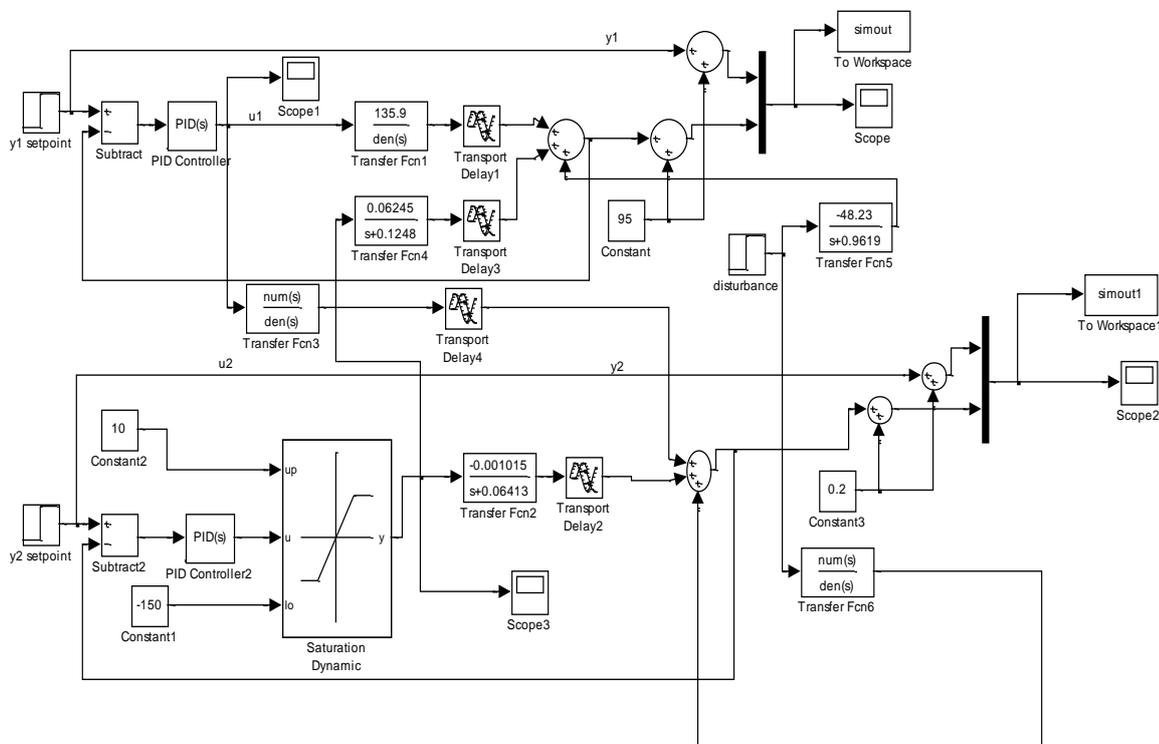


Figure 6.3: The control structure based on RGA analysis using PID controller.

Figure 6.3 shows the selected closed loop control structure based on GRDG and DRGA analysis using PID controllers in Matlab, within step changes in the set points of the carbon capture level and lean loading. The controllers were tuned using Ziegler-Nichols tuning combined with the BLT tuning method. In details, the controller were first tuned using the Ziegler-Nichols rules as if they are for single input and single output systems without control

loop interactions and then detuned using BLT tuning method to consider control loop interactions. For BLT tuning method, a factor F was introduced to modify P, I and D parameters. The updated gains of all controllers K_{ci} are calculated as:

$$K_{ci} = \frac{K_{ZNi}}{F} \quad 6-18$$

where

$$K_{ZNi} = \frac{K_{ui}}{2.2} \quad 6-19$$

The updated reset time τ_{Ii} is shown as below:

$$\tau_{Ii} = \tau_{ZNi} F \quad 6-20$$

where

$$\tau_{ZNi} = \frac{2\pi}{1.2w_{ui}} \quad 6-21$$

Thus, the closed-loop system is described as:

$$\mathbf{X} = \mathbf{GM} = \mathbf{GB}(\mathbf{X}^{set} - \mathbf{X}) \quad 6-22$$

$$\mathbf{X} = (\mathbf{1} + \mathbf{GB})^{-1} \mathbf{GBX}^{set} \quad 6-23$$

As stated by Luyben (1986), a multivariable closed-loop log modulus L_{cm} was defined to obtain the suitable factor F (Eq.6-24)

$$L_{cm} = 20 \log \left| \frac{W}{1+W} \right| \quad 6-24$$

where

$$W(s) = -1 + \det(\mathbf{I} + \mathbf{GB})(s) \quad 6-25$$

By varying F factor, the suitable biggest log modulus L_{cm}^{max} is determined as:

$$L_{cm}^{max} = 2\mathbf{N} \quad 6-26$$

where \mathbf{N} is number of square matrix of control structure.

In this case study, as the control structure is previously selected as 2×2 , the biggest log modulus L_{cm}^{max} is determined as 4. From figure 6.4, the magnitude reached to 4 when F factor is 2.1. Therefore, F factor (2.1) was introduced to adjust the P, I and D parameters.

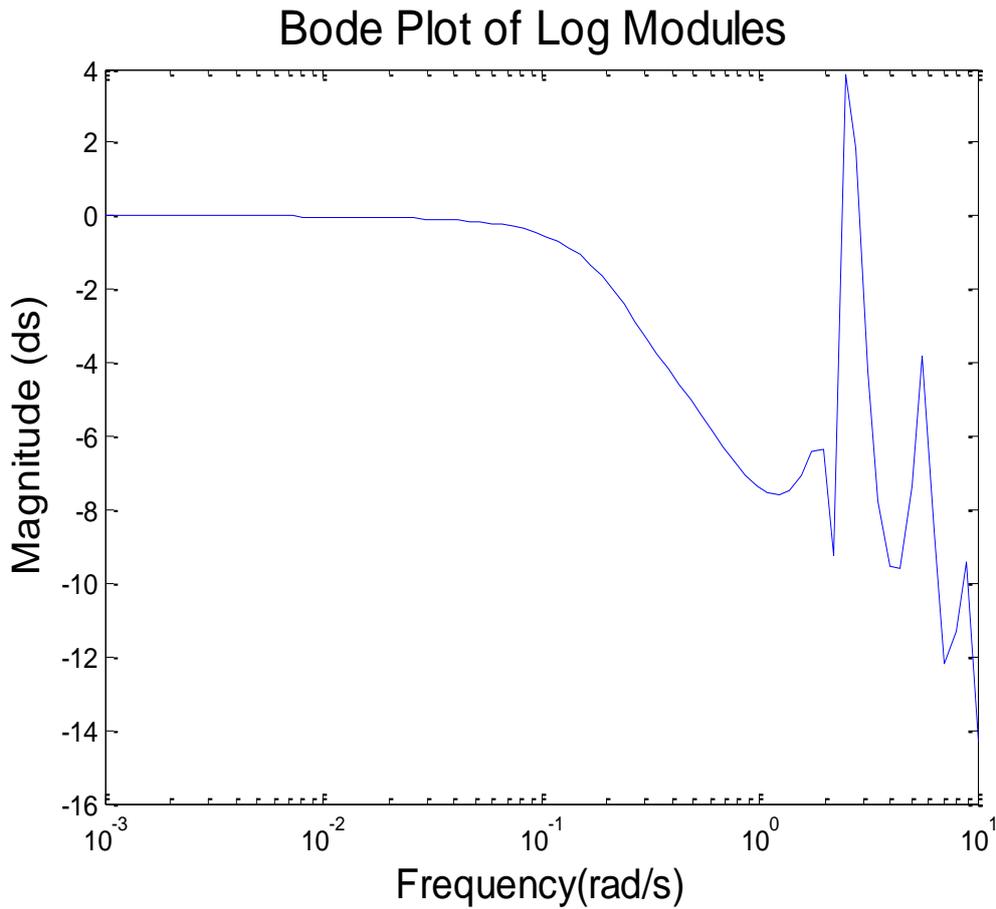
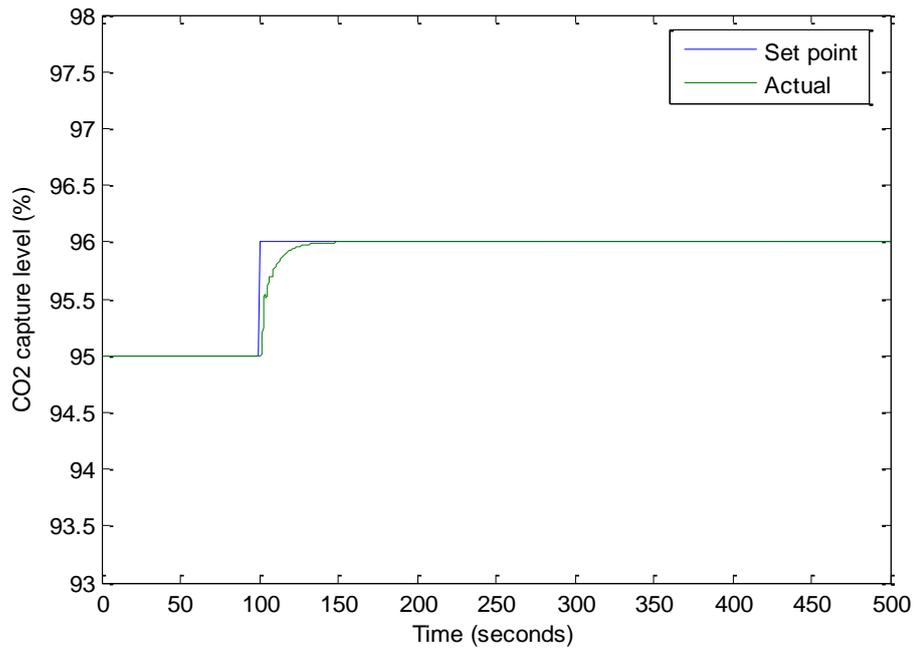
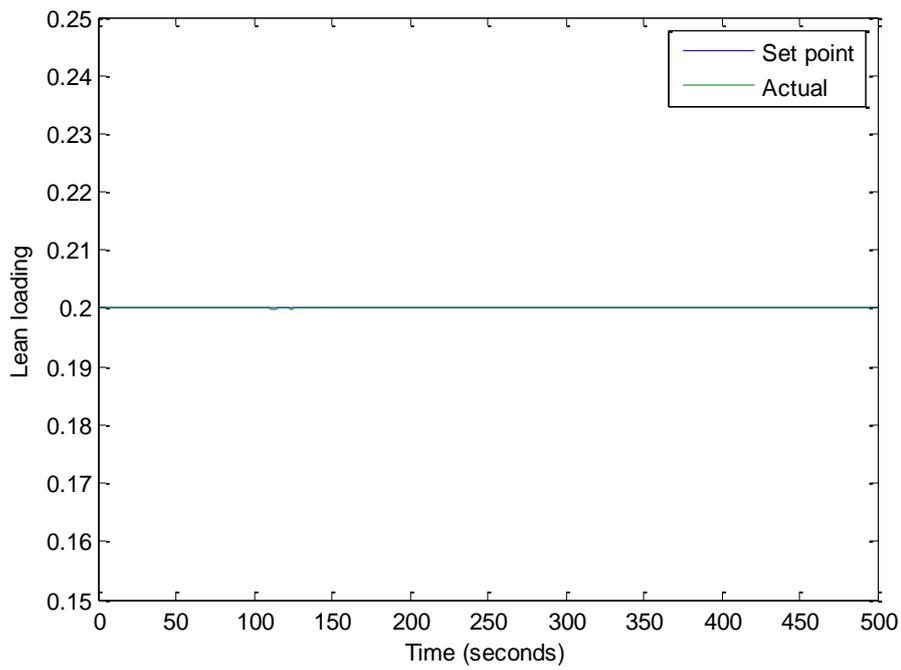


Figure 6.4: Magnitudes of L_{cm}^{max} with different frequencies.

In order to validate the RGA, GRDA and DRGA results, the closed loop response of the control structure ($y_1 - u_1, y_2 - u_2$) to set point changes in CO₂ capture level and lean loading were performed. The set point positive changes of CO₂ capture level were introduced to the system at 100 seconds. These scenarios represent a decrease demand of power plant. When the upstream power plant encounter an off-peak duration, The energy it supplies to capture plant could be increased. The positive changes of capture level and lean loading will be achieved. As shown in figure 6.5a, the response of CO₂ capture level requires approximately 50 seconds to attain the new set point. In addition, it has little interaction on lean loading, as there is no change on the manitude at 100s on figure 6.5b.

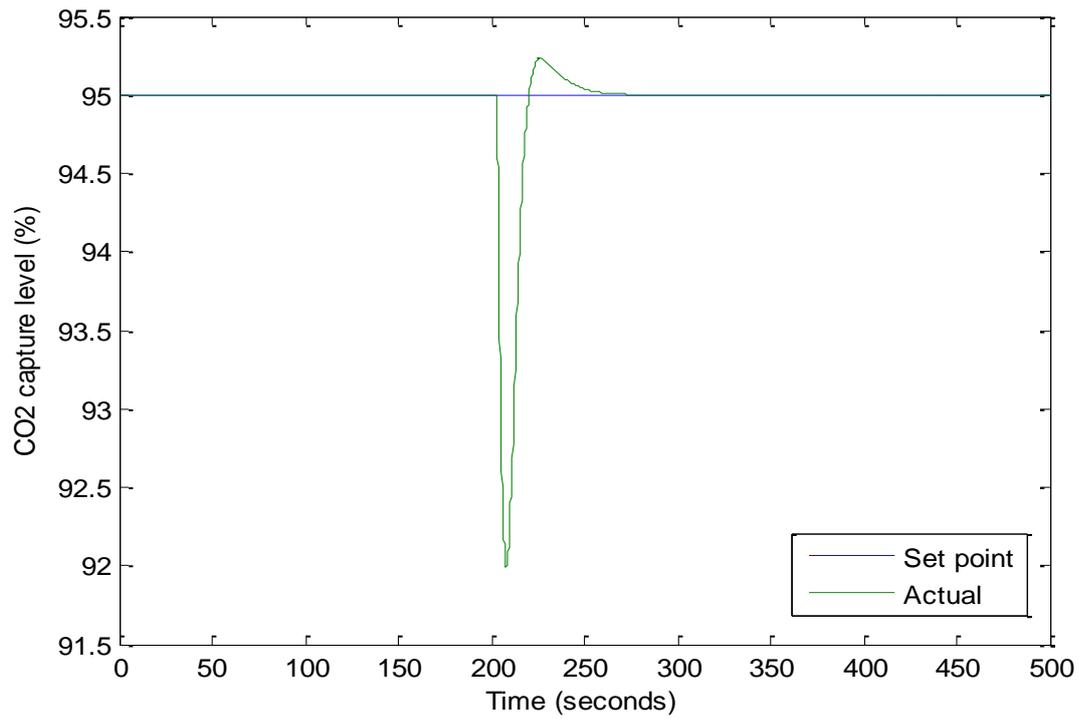


(a)

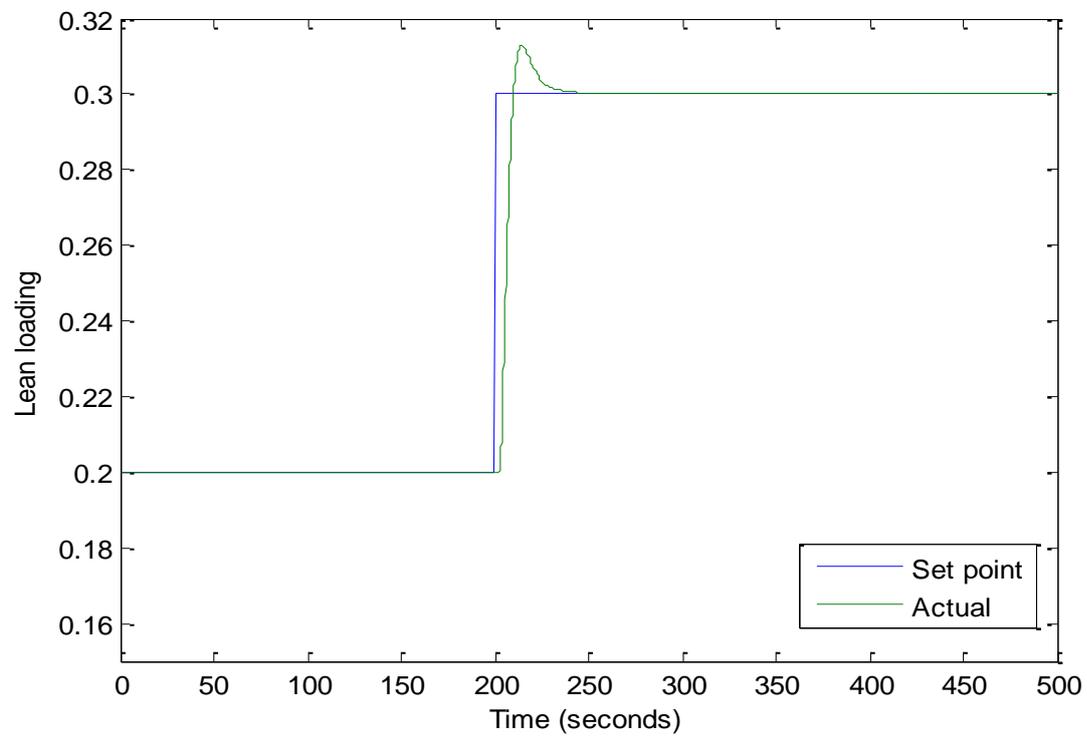


(b)

Figure 6.5: Response to set point changes in CO₂ capture level at 100s.



(a)

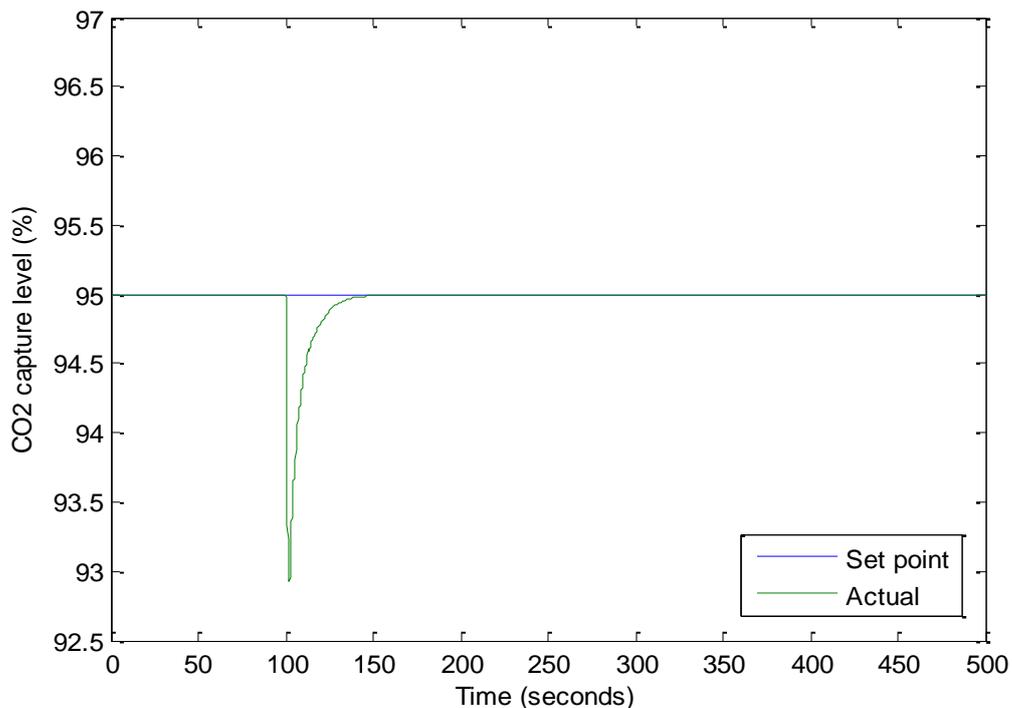


(b)

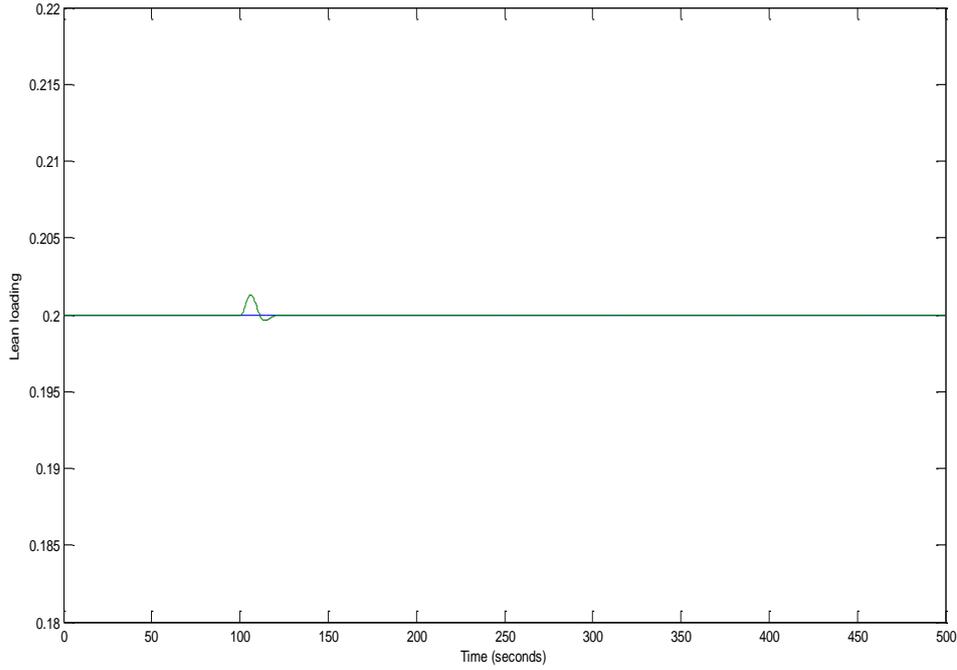
Figure 6.6: Response to set point changes in lean loading at 200s.

At 200s, a 50% increase in the set point of lean loading will change both CO₂ capture level and lean loading with respect to their base conditions. Figure 6.6a indicates the actual value of CO₂ capture level has experienced a 3% decrease and then reach its set point with 70 seconds. It reflects the control loop ($y_2 - u_2$) affects the control loop ($y_1 - u_1$) a lot. This is because, once the CO₂ concentration in lean MEA solution increases, the absorption capacity of lean solution will decrease. In this case, the CO₂ capture level will be reduced temporarily and then recover by control loop. As shown in figure 6.6b, it takes 50 seconds to allow the actual lean loading achieve its new set point, with 5% excess.

A disturbance of inlet flue gas flow rate, 0.05 kg/s increase, was introduced to the system at 100s, which is a common scenario in the operation of the power plant. For instance, when the upstream power plant is at the stage of start-up, or at the peak duration, the inlet flue gas flowrate coming from the power plant will be increased. Figure 6.7a shows the operation performance of CO₂ capture level and lean loading under the disturbance. The CO₂ level is decreasing and then reached to the set point with control action, which requires about 50 seconds. The reason is that, the CO₂ composition is increased with the increasing inlet flue gas flow rate, while the absorption capacity in the column is unchanged. Hence, the removing percentage of CO₂ is decreasing. Figure 6.7b states the lean loading is almost unchanged with the increasing inlet flue gas flow rate.



(a)



(b)

Figure 6.7: Response of CO₂ capture level (a) and lean loading (b) to disturbance in inlet flue gas flowrate.

6.4 MPC control analysis

6.4.1 State space model for MPC controller

To obtain the suitable MPC controller, the s-domain transfer functions should be transformed into state space form. It is extended from Eqs.2-72, 2-73 and 2-74 that, an integrator is introduced by using the differentiated state vector as below:

$$\Delta_k = \mathbf{x}_k - \mathbf{x}_{k-1} \quad 6-27$$

By integrating with the controlled output \mathbf{z}_k of Eq.2-73:

$$\begin{bmatrix} \Delta \mathbf{x}_{k+1} \\ \mathbf{z}_{k+1} \end{bmatrix} = \begin{bmatrix} A & 0 \\ DA & I \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_k \\ \mathbf{z}_k \end{bmatrix} + \begin{bmatrix} B \\ DB \end{bmatrix} \Delta u_k + \begin{bmatrix} C \\ DC \end{bmatrix} \Delta v_k + \begin{bmatrix} I \\ D \end{bmatrix} \Delta w_k \quad 6-28$$

$$\mathbf{z}_k = \begin{bmatrix} 0 & I \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_k \\ \mathbf{z}_k \end{bmatrix} \quad 6-29$$

$$\mathbf{y}_k = \mathbf{z}_k + \mathbf{3}_k \quad 6-30$$

The above Eqs.6-28, 6-29 and 6-30 can be written as:

$$\hat{x}_{k+1} = \check{A}\hat{x}_k + \check{B}\Delta u_k + \check{C}\Delta v_k + \Delta\check{w}_k \quad 6-31$$

$$z_k = \check{D}\hat{x}_k \quad 6-32$$

$$y_k = Dx_k + z_k \quad 6-33$$

A state observer is used to estimate the state vector, which is shown as below:

$$\beta_k = y_k - \check{D}\hat{x}_{k|k-1} \quad 6-34$$

$$\hat{x}_{k+1|k} = \check{A}\hat{x}_{k|k-1} + \check{B}\Delta u_k + \check{C}\Delta v_k + E\beta_k \quad 6-35$$

The above equations give one-step ahead predictions of state vector. Z is predicted by multiplication \check{D} with estimated state. The output vector can be predicted with p samples ahead (prediction horizon) and control actions are taken into account for m samples (control horizon), which denotes $m \leq p$. It is simplified as:

$$\mathbf{u}_k = \begin{bmatrix} u_k \\ \vdots \\ u_{k+m-1} \end{bmatrix}, \quad \mathbf{z}_k = \begin{bmatrix} z_k \\ \vdots \\ z_{k+p-1} \end{bmatrix} \quad 6-36$$

The MVs over the control horizon and process variables over prediction horizon are collected in above expressions. Based on these, the predicted process variables over the prediction horizon are:

$$\begin{aligned} z_{k+1|k} = & \begin{bmatrix} \check{D}\check{A} \\ \vdots \\ \check{D}\check{A}^p \end{bmatrix} \hat{x}_{k|k-1} + \begin{bmatrix} \check{D}\check{B} & \dots & 0 \\ \vdots & \ddots & \vdots \\ \check{D}\check{A}^{p-1}\check{B} & \dots & \check{D}\check{A}^{p-m}\check{B} \end{bmatrix} \Delta \mathbf{u}_k \\ & + \begin{bmatrix} \check{D}\check{C} \\ \vdots \\ \check{D}\check{A}^{p-1}\check{C} \end{bmatrix} \Delta v_k + \begin{bmatrix} \check{D}\check{E} \\ \vdots \\ \check{D}\check{A}^{p-1}\check{E} \end{bmatrix} \beta_k \end{aligned} \quad 6-37$$

The feedback error β_k is based on the most recent measurement of y_k .

The control error over the prediction horizon is the difference between predictions and the set points.

$$E_{k+1} = z_{k+1|k} - Rf_{k+1} \quad 6-38$$

It is consisted of three parts, such as set-point error, feed forward error and remaining error, which is sum up to $\Delta \mathbf{u}_k$, the increments of MVs. The optimisation process of MPC controller is to find the minimum values of increments of each manipulate variables.

6.4.2 Objective function

The performance criteria of control scheme is determined by the plant operation requirements. The criteria can be expressed by a mathematical term. In MPC, a quadratic objective function is used which can be introduced as following form:

$$\min_{\Delta u(k)} \sum_{l=1}^p \|\mathbb{T}_l^y [y(k+l|k) - R(k+l)]\|^2 + \sum_{l=1}^m \|\mathbb{T}_l^u [\Delta u(k+l|k)]\|^2 \quad 6-39$$

$$\text{constraints: } c^u \Delta u(k) \cong c(k+1|k) \quad 6-40$$

where

$$\mathbb{T}^u = \text{diag} \{ \mathbb{T}_1^u, \dots, \mathbb{T}_m^u \} \quad 6-41$$

and

$$\mathbb{T}^y = \text{diag} \{ \mathbb{T}_1^y, \dots, \mathbb{T}_p^y \} \quad 6-42$$

are the weight matrices in block diagonal form, and

$$R(k+1) = \begin{bmatrix} r(k+1) \\ \vdots \\ r(k+p) \end{bmatrix} \quad 6-43$$

is the vector of reference trajectory.

The criteria minimize the sum of squared deviations of the predicted CV values from a time-varying reference trajectory or setpoint $r(k+l)$ over p future time steps.

6.4.3 Tuning rules of MPC

The parameters tuning method is of great importance in designing MPC controllers. The following steps are detailed to achieve a satisfied MPC. The first step is to determine the suitable sampling time conforming to the actual plant process. There is an agreement that less sampling time result in more unknown disturbance rejections and more computational efforts. In this case study, the sampling time is 1 second. Then, the prediction horizon (p) was concluded as 40 because it should be twice or triple greater than the dominant time constant of the process (Love, 2007). According the tuning rules of MPC, the control horizon (m) is determined between inherent time delay and prediction horizon. It is because, if the control horizon is higher than the prediction horizon, some changes of MVs which perfect the predictions, will not be considered. Eq.6-1 indicates the highest delay time is 3, hence the control horizon was determined as 5 in this case.

6.5 Results and analysis of MPC control

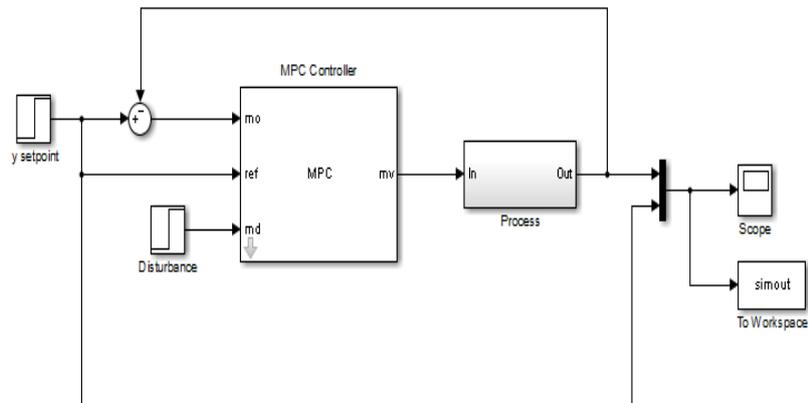


Figure 6.8: Simulink model with MPC controller.

Figure 6.8 shows the simplified graph of MPC control structure in Matlab. It can be seen clearly that, unlike PID control structure, the MVs are centralized controlled by MPC controller. Here, the same actions, such as set point changes of CO₂ capture level, lean loading, and step change in inlet gas flow rate, were taken to evaluate the performance.

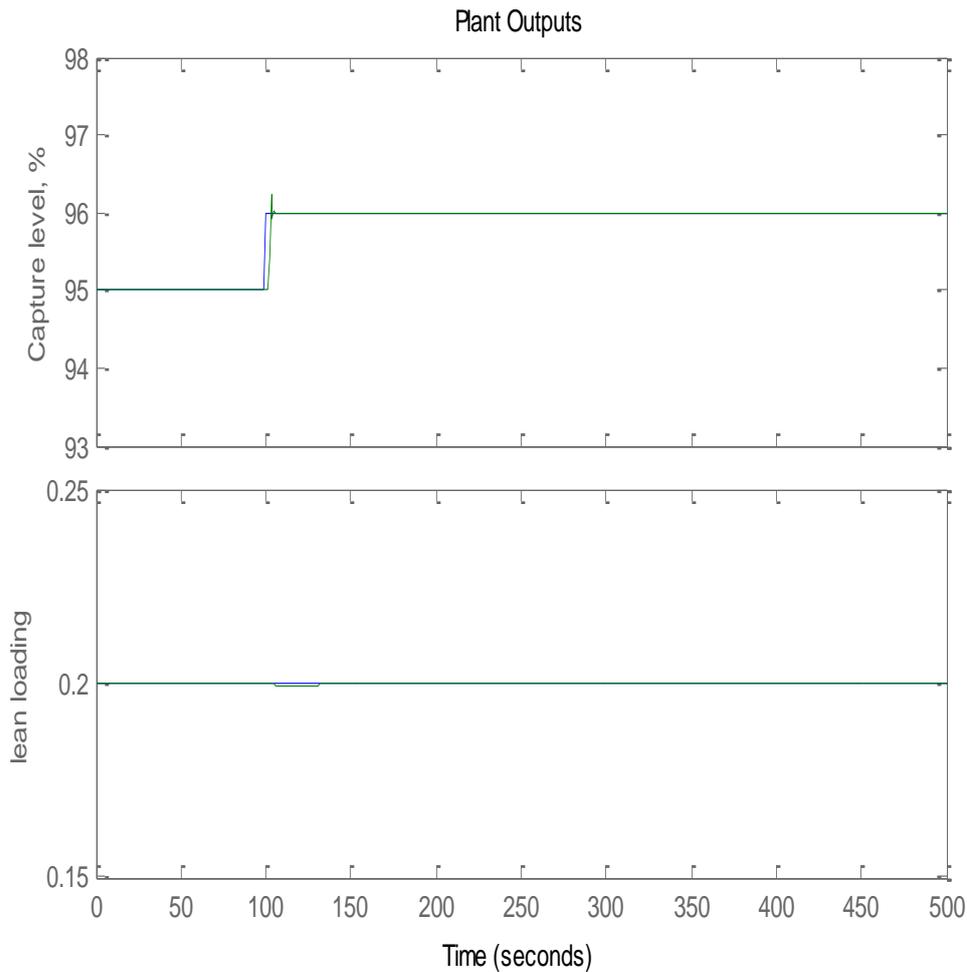


Figure 6.9: Response to set point changes in CO₂ capture level at 100s by MPC control.

Blue: Set point; Green: Actual.

Firstly, the set point change of CO₂ capture level at 100s was implemented to the system, which was same as PID control structure. From figure 6.9, the actual value arrives its new set point values with only 5 seconds, while the lean loading is almost keep steady at same time. Compare figure 6.9 with figure 6.5, the rising time by MPC controller is less than that by PID controller. Furthermore, the changes in manipulated variables (lean flowrate and reboiler temperature) are shown in figure 6.10. Specifically, in the top graph, the lean flow rate is increasing to provide more amount of lean solution, thereby capturing more CO₂ every second. With the operation, it is steady at a new point. In the bottom graph, the reboiler temperature is initially increased as well to meet the new set point of CO₂ capture level but finally reduced to its set point. This is because, the increasing reboiler temperature will result in more CO₂ vaporising from solution in stripper and then generate less lean loading for the

solution for the complement of capturing more CO₂ in the absorber. However, due to the unchanged set point of lean loading, the reboiler temperature is reverted to its original state.

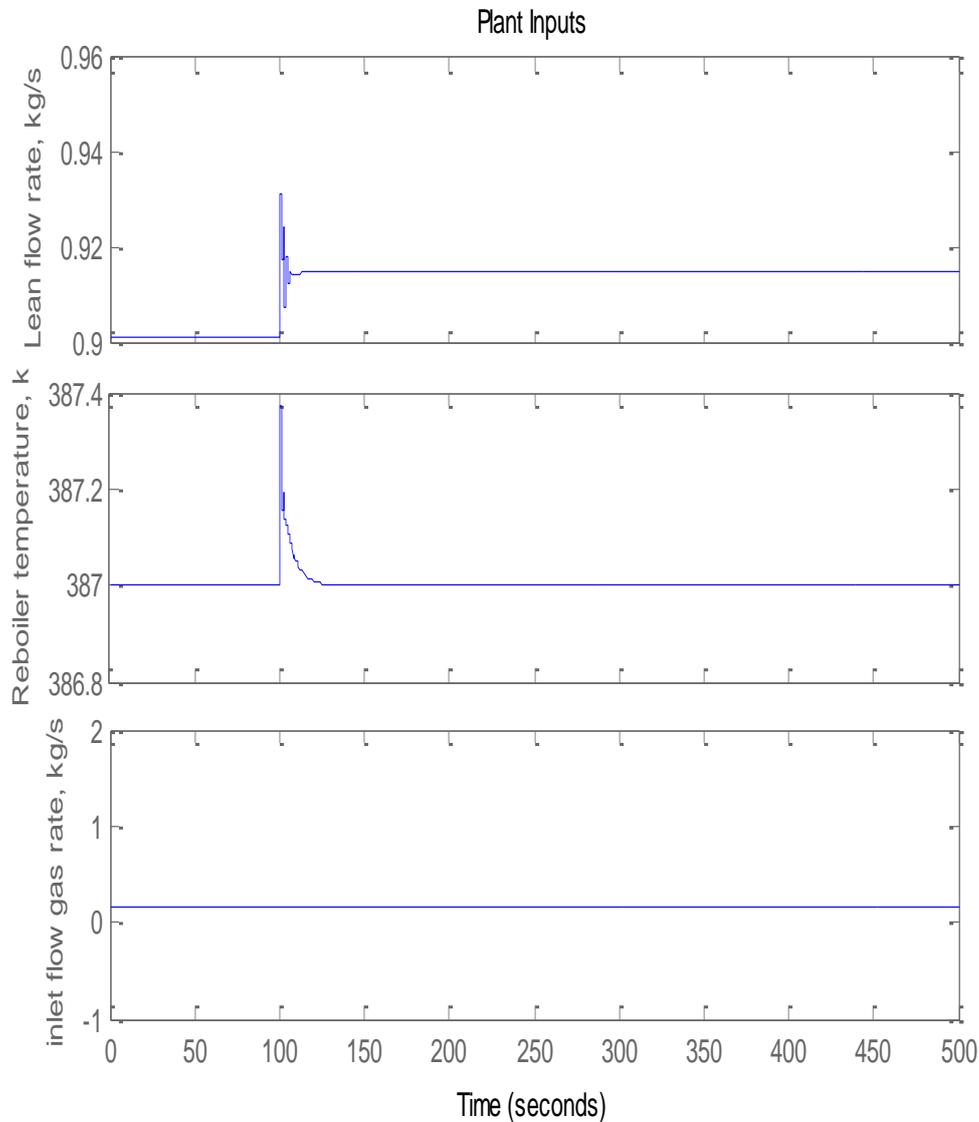


Figure 6.10: Changes in manipulated variables during the set point changes of CO₂ capture level.

Figure 6.11 indicates the performance of outputs while the set point change of lean loading was introduced to the system at 200s. It can be seen clearly that the CO₂ capture level remains almost steady and the lean loading is increasing to its new set point within 50 seconds. Compare figure 6.11 with figure 6.6, the MPC controller performs better as it can keep CO₂ capture level unchanged and lean loading not exceeding its set point. The settling time by MPC controller and PID controller are almost the same. Figure 6.12 shows the changes of

manipulated variables corresponding to the set point change of lean loading at 200s. In details, the increasing lean loading will lead to the reduced absorption capacity of solution. In this case, the lean solvent flow rate should be increased to keep the capture level unchanged. To achieve the larger lean loading, the reboiler temperature is operated to reduce for vaporising less CO₂ from rich solution in the stripper. All these two MVs were changed to new points to satisfy this action.

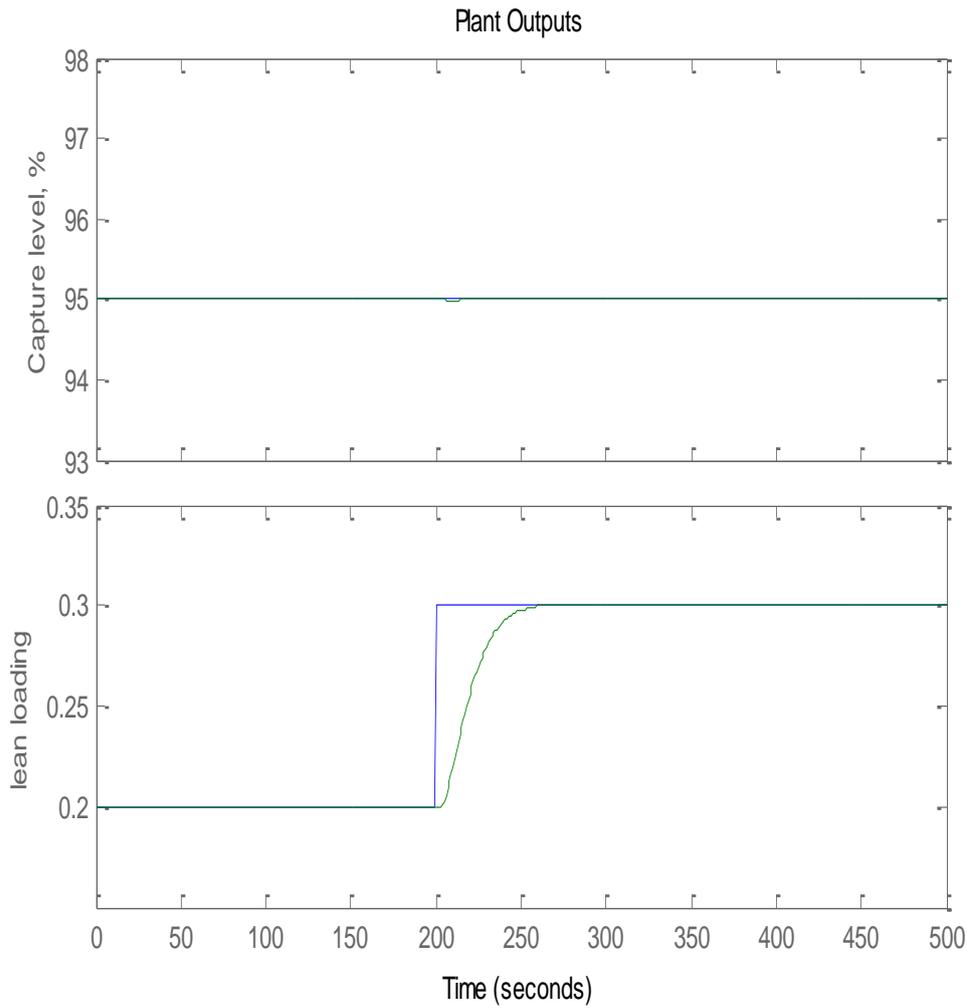


Figure 6.11: Response to set point changes in lean loading at 200s by MPC control.

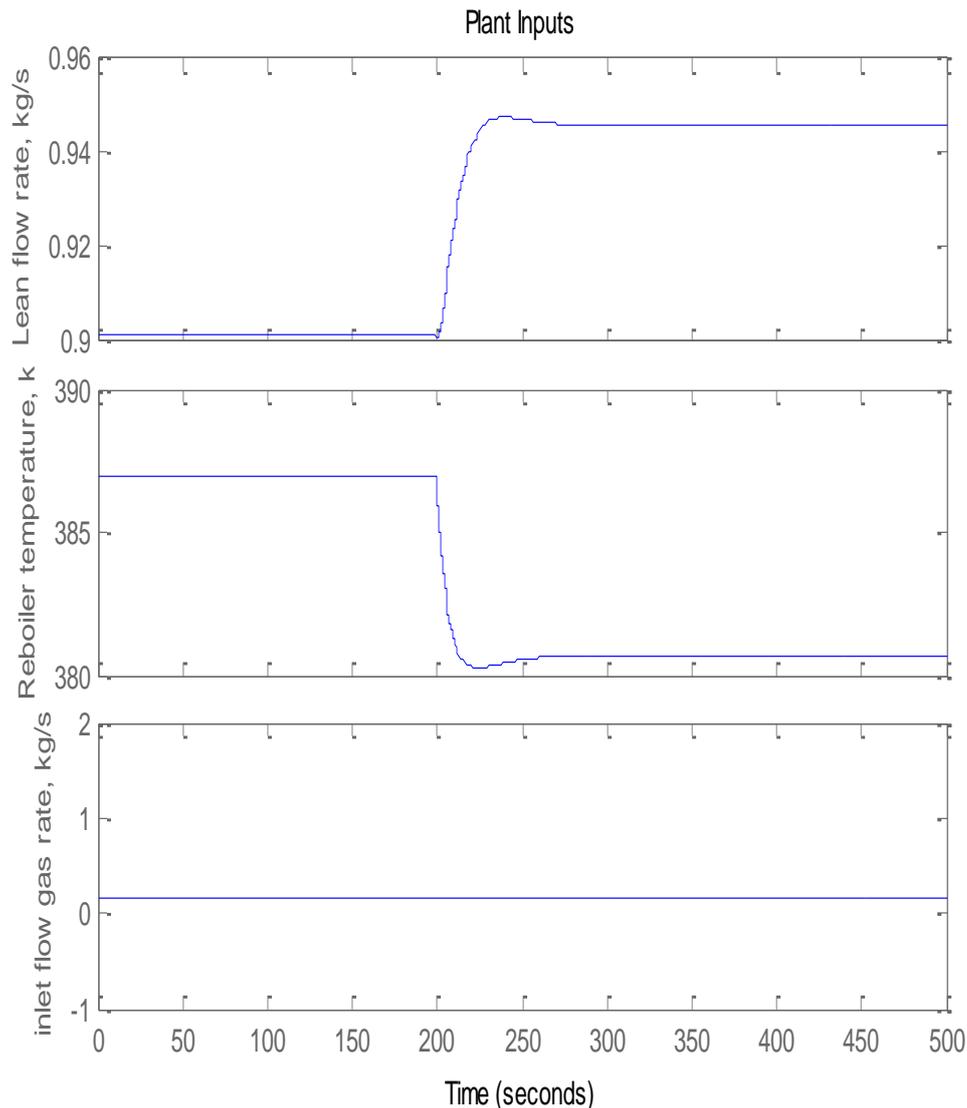


Figure 6.12: Changes in manipulated variables during the set point changes of lean loading.

Figure 6.13 shows the performance of CO₂ capture level and lean loading under a disturbance of inlet gas flow rate. The CO₂ capture level is first reduced and then recovered to its set point within only 5 seconds. The lean loading experience slightly fluctuation during the period of disturbance rejection. Compare figure 6.13 with figure 6.7, the setting time of CO₂ capture level by MPC controller is less than that by PID controller (50s), while the output of lean loading performs similar. This is to say, the MPC controller has a higher controlling capacity than PID controller. To keep the outputs at their set points, the manipulated variables will be changed with operation corresponding to the disturbance. As the inlet gas flow rate is increased, the CO₂ component to the system will be increased as well. Therefore, the lean solvent flow rate should be operated to increase for absorbing more CO₂, as well as the

reboiler temperature. However, since the set point of lean loading is unchanged, the reboiler temperature which can take a great impact on lean loading, is recovered to its original value at last.

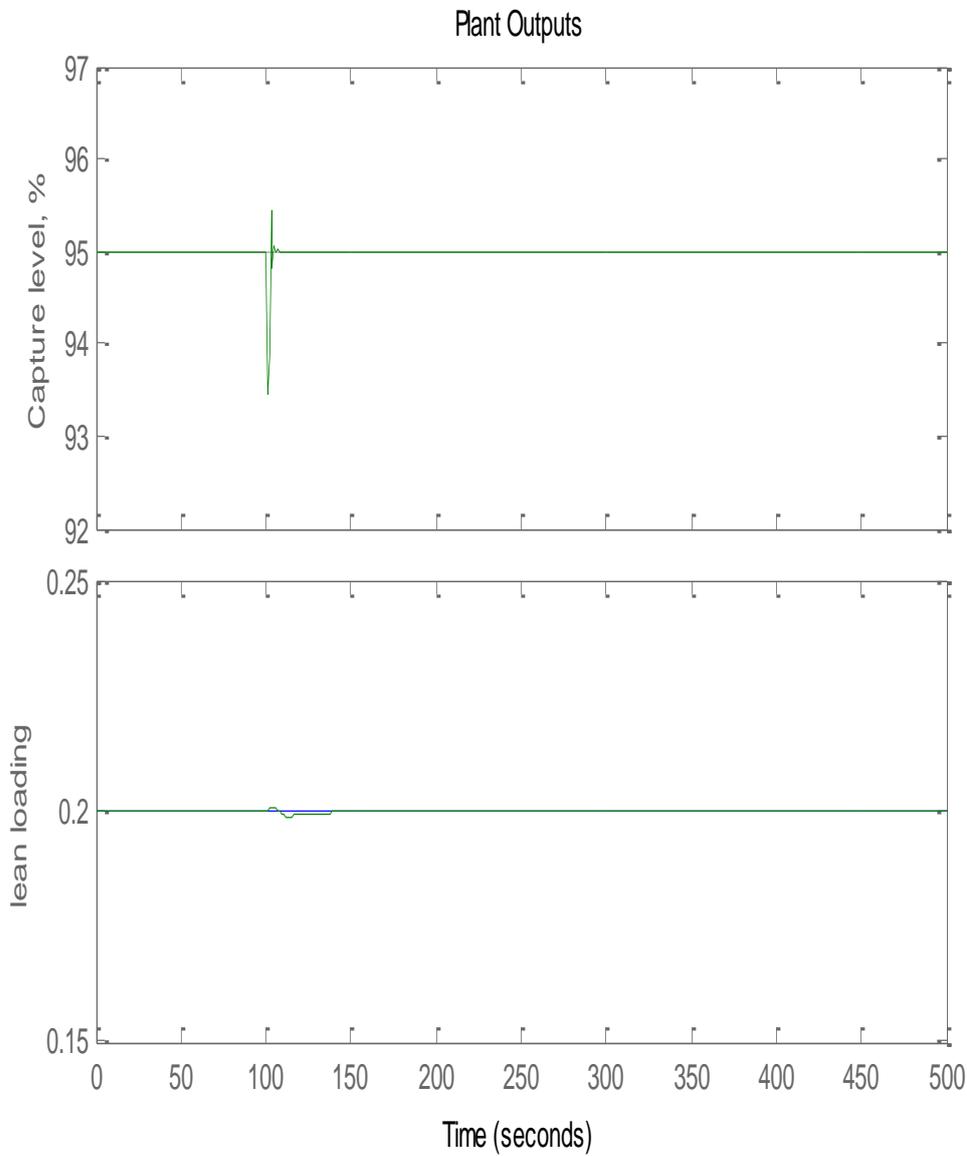


Figure 6.13: Response to step change in inlet flow gas rate at 100s by MPC control.

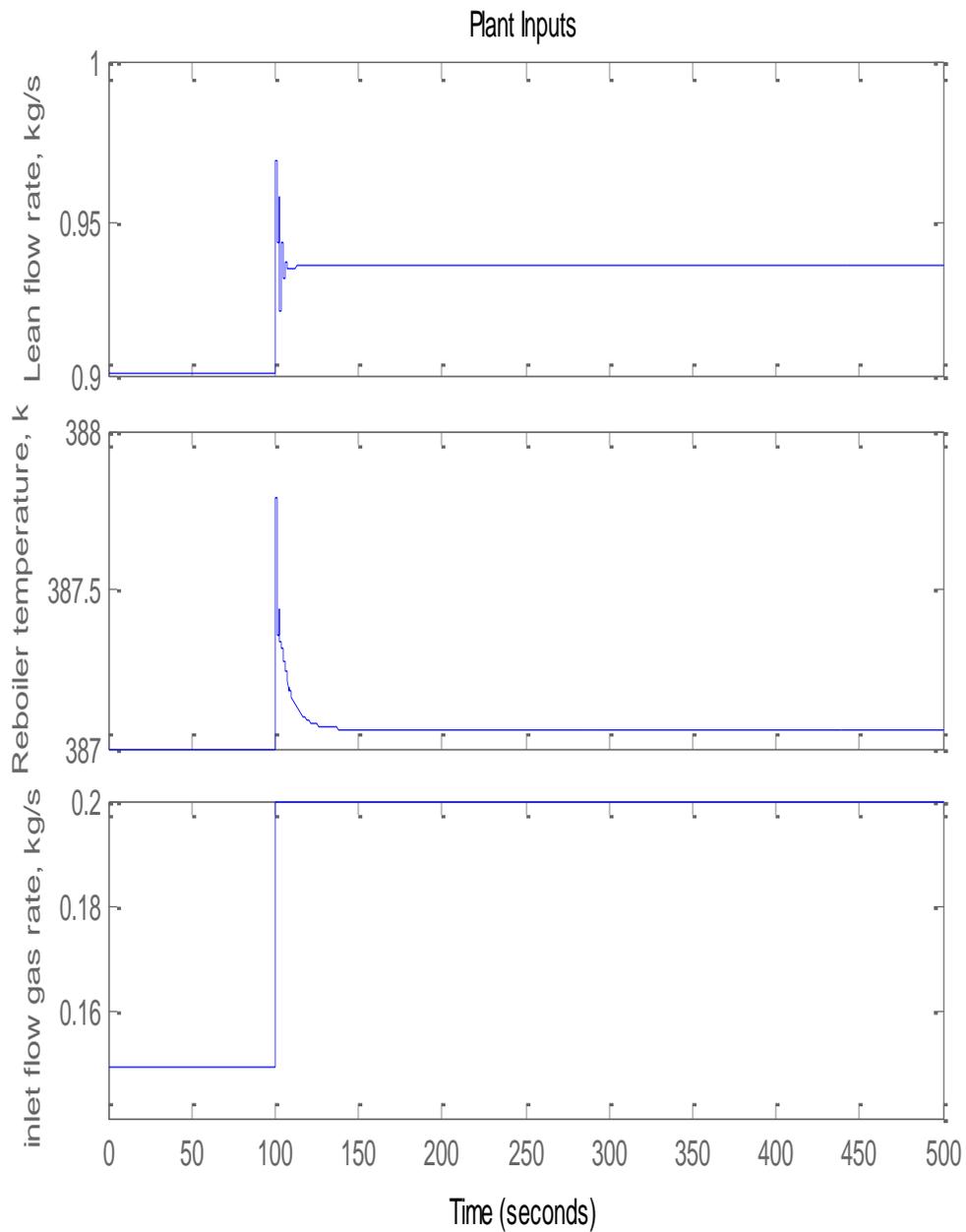


Figure 6.14: Changes in manipulated variables during the set point changes of inlet flow gas flow rate.

To sum up, MPC controller is demonstrated as a better choice to establish an effective control system for post-combustion CO₂ capture plant. In addition, it is also proved that the system including time delay and high order can be handled by MPC control structure.

6.6 Conclusions

In this work, two types of control structures (decentralized and centralized) for MEA-based post-combustion CO₂ capture plant were presented. For decentralized control structure, the general relative disturbance gain (GRDG) and dynamical relative gain array (DRGA) analysis represent the base for the design of the control schemes. It was determined to pair CO₂ capture level and lean solvent flow rate as a control loop, as well as lean loading and reboiler temperature. The Ziegler-Nichols rules integrated with BLT tuning method are used to tune PID parameters. For centralized control structure, MPC controller was used to design the control scheme. The effect of step change disturbance in inlet gas flow rate, set point tracking of CO₂ capture level and lean loading for the plant were studied for the controllability analysis. The results shows that the multi-variable MPC-based control scheme recover faster (short closed-loop settling time) when implement disturbance and set point change to the system with minimum deviation regarding to the plants set-point than PID-based control scheme. In addition, the constraints of manipulated variables can be applied to MPC controller to keep them under feasible limits, while PID controller would possibly violate the saturation limits on those variables for the same change in the plant's operating conditions. All these indicate MPC control scheme is more preferable choice for post-combustion CO₂ capture plant to main the dynamic operability and controllability in the presence of the process disturbance and set-point tracking.

Chapter 7. Conclusions and Recommendations for Future Works

7.1 Conclusions

MEA-based post-combustion CO₂ capture process plays a vital role in removing CO₂ emissions coming from coal-fired power plants, as it has capacity of capturing the low-pressure CO₂ in flow gas and retrofitting the existing power plants easily. The modelling and optimisation of this process has been treated as a main target to improve the capture performance such as CO₂ capture level and energy consumption. In previous research, more attentions were paid to mechanistic model analysis, which need a deepened knowledge of process underlying physics. It would possibly result in some time issues, even though the process input-output relationships could be expressly represented by the mechanistic model. However, in this research, the computational intelligence techniques, including BA-NNs, BA-ELM and DBN models, were applied instead of mechanistic strategies to model the post-combustion CO₂ capture process. Furthermore, the BA-NNs model was used in the process optimisation for energy efficiency operation. Following this part, the decentralized and centralized control schemes with the aim of enhancing energy efficiency were presented with the application of post-combustion CO₂ capture plant.

In order to model the post-combustion CO₂ capture process integrated with power plant, a majority of computational intelligent models were explored. In this work, they were categorized into two groups: static model and dynamic model. For static model, seven parameters of the process, namely inlet gas flow rate, CO₂ mass fraction in inlet flow gas, inlet gas flow pressure, inlet gas flow temperature, lean solvent circulation rate, MEA concentration and lean solution temperature, were investigated as input variables, while one parameter, capture level, was treated as output variable. As to dynamic model, eight key process variables, such as inlet gas flow rate, CO₂ mass fraction in inlet flow gas, inlet gas flow temperature, inlet gas flow pressure, lean solvent circulation rate, lean solvent temperature, reboiler temperature and lean loading, were considered as input variables, while CO₂ capture level and CO₂ production rate were adopted as output variables.

ANNs were found to be the appropriate technology to model the process no matter how complicated the underlying process relationships will be. Seven To improve model accuracy and reliability, multiple feedforward neural network models were developed from bootstrap re-sampling replications of the original training data and were combined. BA-NNs model can offer more accurate predictions than a single neural network according to MSE values of testing data, as well as provide model prediction confidence bounds. The model prediction

confidence bounds reflect the reliability of model. The computational intelligence based model was further employed to save the energy consumption in the regenerator column. With the profound understanding of the post-combustion CO₂ capture process, the energy efficiency is improved as the lean solvent flow rate is reduced. Therefore, the lean solvent flow rate was minimized as the optimisation objective, as well as the CO₂ capture level was operated above 90%. The optimisation technique based on BA-NNs model is presented in this work. The optimum operating condition of lean solvent flow rate was found as 0.5698 kg/s.

To further improve the learning speed and generalization performance, the BA-ELM was used in this work. The input weights and biases were randomly assigned, while the weights between the hidden layer and output layer were obtained using regression type approach in one step. This feature allows the BA-ELM to be developed very quickly. This work proposes principal component regression to calculate the weights between the hidden layer and output layer to address the collinearity issue among hidden neuron outputs. Since the input weights and biases were randomly assigned, the BA-ELM model had variations in performance. By comparing the training time between BA-ELM and BA-NNs models, it was found that the former CPU time was several times less than the latter one. In addition, the prediction accuracy for unseen testing data by BA-ELM model was better than that by BA-NNs model, which presented a better generalisation performance. The one-step ahead and multi-steps ahead predictions by BA-ELM model were also conducted, which appear to be both better than those by BA-NNs model. The problems of local minima and over-trained encountered by BA-NNs model could be resolved by the implementation of BA-ELM model.

However, both of the two above mentioned neural networks have only one hidden layer, which are narrow neural networks. More hidden layers constitute the neural networks, more accurate relationships between input and output it can represent. Therefore, DBN model with many layers of RBM were proposed then. By using DBN model, a deep hierarchical representation of training data was extracted in unsupervised pre-training stage. A greedy layer-wise unsupervised learning algorithm is used to obtain initial weights of the subsequent supervised phase. This can result in global optimum, which gives better generalisation than the single hidden layer neural network. Then, the network weights were fined tuned in a supervise manner by the supervised back-propagation phase. With comparing the MSE values of unseen testing data between SLFN model and DBN model, it was found that the former one was much larger than the latter one. This indicates that the multiple hidden layer neural networks have the capacity to catch the underlying feature of the data more easily, thereby achieving the better generalisation ability

The controllability analysis of post-combustion CO₂ capture process is implemented following the model developments. Two key process variables, lean solvent flow rate and reboiler temperature, were considered as manipulated variables, while the CO₂ capture level and lean loading were used as control variables. As to decentralized scheme, the control structure selection for the process was determined with application of GRDG and DRGA in addition to RGA for operability analysis. GRDG was developed from analysing the process disturbance, inlet flue gas flow rate, showing the same result of control structure selection to RGA. With a consideration of dynamic information, DRGA was employed in this work. It was a frequency-dependent RGA analysis, which indicated the different result as RGA analysis. In details, at low frequencies, the same pairing was suggested as steady-state RGA, while at high frequencies, it was found the opposite pair selections. The uncertainty ranges of RGA, GRDG and DRGA were calculated, as there could possibly exist model uncertainties. The PID controllers in decentralized control structures were tuned using Ziegler-Nichols tuning combined with the BLT tuning method. With respect to centralized control strategy, MPC based control structure was applied in this research, with the adjustment of prediction horizon and control horizon. Finally, to compare the performance of PID-based and MPC-based control structures, the closed loop response to set point changes and disturbance was presented. The results have shown MPC-based control scheme is more appropriate to control the dynamic process in presence of the process disturbance and set-point change.

7.2 Recommendations for future work

The data used for developing the model was generated from the gPROMs software in Hull University. The dynamic models in gPROMs were developed using the rate-based approach. However, they assumed that all chemical reactions attained equilibrium. To improve the accuracy of the model, the reaction kinetics could be involved (moving from model 2 to model 3 in Fig. 2.10).

In ELM model, the weights and bias between hidden layer and output layer was calculated by principal component regression method. To further explore ELM algorithms, partial latent square method could be used. With respect to DBNs model, it only accounted for individual DBNs model in this project. Next, the bootstrap aggregated DBN model could be developed to compare with BA-NNs and BA-ELM.

The model techniques developed in this work were limited to feedforward neural networks, which take no consideration of sequential information. With the inclusion of a 'memory' which captures information about what has been calculated so far, the accuracy of dynamic

models could be improved. Therefore, it should be interesting to apply recurrent neural networks (RNNs) to model the post-combustion CO₂ capture process. In addition, to process high-dimensional input signals and extract complex features, another new method, slow feature analysis (SFA), could be further used to learn invariant or slow varying features from the input signal. With respect to DBNs model, it only accounts for individual DBNs model in this project. Next, the bootstrap aggregated DBN model could be developed to compare with BA-NNs and BA-ELM.

The main focus of optimisation was on energy efficiency. The further work could be concentrated on the other possible consideration of optimisation objectives and real time optimisation.

The transfer function for control studies are linear and in future nonlinear dynamic process model should be used to test control performance. The decoupler was not introduced to the PID control scheme, which would result in the interference between control loops. To eliminate the interference, the decoupler could be considered. The MPC control scheme in this project also used the linear transfer function. To improve control performance, the neural network models developed here could be used in MPC controller design in the future

There are only two controlled variables and two manipulated variables considered in this project. It is far too simplified to apply the control scheme to a real post-combustion carbon capture plant. The control of reboiler heat duty, the temperature of absorber column, the inlet temperature of lean solvent L/G ratio could be considered in the future to develop the control structure.

Some chemical reaction phenomenon, such as solvent degradation, have been ignored in this project. It is necessary to investigate the effect of varying compositions of O₂, SO₂ and NO_x. O₂ and SO₂ should be separated before capture technology. However, extracting O₂ from flue gas is much more difficult. Under such circumstance, it is necessary to study the effect of O₂ on capture process, as it may result in the degradation of solvent.

References

- Aaron, D. and Tsouris, C. (2005) 'Separation of CO₂ from flue gas: A review', *Separation Science and Technology*, 40(1-3), pp. 321-348.
- Abdulhameed, M.A., Othman, M.H.D., Ismail, A.F., Matsuura, T., Harun, Z., Rahman, M.A., Puteh, M.H., Jaafar, J., Rezaei, M. and Hubadillah, S.K. (2017) 'Carbon dioxide capture using a superhydrophobic ceramic hollow fibre membrane for gas-liquid contacting process', *Journal of Cleaner Production*, 140, pp. 1731-1738.
- Abu-Zahra, M.R.M., Schneiders, L.H.J., Niederer, J.P.M., Feron, P.H.M. and Versteeg, G.F. (2007) 'CO₂ capture from power plants. Part I. A parametric study of the technical-performance based on monoethanolamine', *International Journal of Greenhouse Gas Control*, 1(1), pp. 37-46.
- Adeiwale Adeosuna, M.R.M.A.-Z. (2013) 'Evaluation of amine-blend solvent systems for CO₂ post-combustion capture applications', *Energy Procedia*, (37), pp. 211-218.
- Agustriyanto, R. and Zhang, J. (2007) 'Obtaining the worst case RGA and RDGA for uncertain systems via optimization', *2007 American Control Conference, Vols 1-13*, pp. 1642-1647.
- Ahmad, Z. and Zhang, J. (2005) 'Combination of multiple neural networks using data fusion techniques for enhanced nonlinear process modelling', *Computers & Chemical Engineering*, 30(2), pp. 295-308.
- Akesson, J., Laird, C.D., Lavedan, G., Prolss, K., Tummescheit, H., Velut, S. and Zhu, Y. (2012) 'Nonlinear Model Predictive Control of a CO₂ Post-Combustion Absorption Unit', *Chemical Engineering & Technology*, 35(3), pp. 445-454.
- Alatiqi, I., Sabri, M.F., Bouhamra, W. and Alper, E. (1994) 'Steady-State Rate-Based Modeling for CO₂/Amine Absorption Desorption Systems', *Gas Separation & Purification*, 8(1), pp. 3-11.
- Anderson, M. (2016) 'Climate Change 2014: Synthesis Report', *Library Journal*, 141(9), pp. 28-28.
- Arce, A., Mac Dowell, N., Shah, N. and Vega, L.F. (2012) 'Flexible operation of solvent regeneration systems for CO₂ capture processes using advanced control techniques: Towards operational cost minimisation', *International Journal of Greenhouse Gas Control*, 11, pp. 236-250.

- Arias, A.M., Mores, P.L., Scenna, N.J. and Mussati, S.F. (2016) 'Optimal design and sensitivity analysis of post-combustion CO₂ capture process by chemical absorption with amines', *Journal of Cleaner Production*, 115, pp. 315-331.
- Aroonwilas, A. and Veawab, A. (2007) 'Integration of CO₂ capture unit using single- and blended-amines into supercritical coal-fired power plants: Implications for emission and energy management', *International Journal of Greenhouse Gas Control*, 1(2), pp. 143-150.
- Astarita, G. and Savage, D.W. (1980) 'Theory of Chemical Desorption', *Chemical Engineering Science*, 35(3), pp. 649-656.
- Bates, J.M. and Granger, C.W.J. (1969) 'The Combination of Forecasts', *Oper. Res. Q.*, 20(4), p. 451.
- Bengio, Y. and LeCun, Y. (2007) *Scaling learning algorithms towards AI*. MIT Press.
- Biliyok, C., Lawal, A., Wang, M.H. and Seibert, F. (2012a) 'Dynamic modelling, validation and analysis of post-combustion chemical absorption CO₂ capture plant', *International Journal of Greenhouse Gas Control*, 9, pp. 428-445.
- Biliyok, C., Lawal, A., Wang, M.H. and Seibert, F. (2012b) 'Dynamic Validation of Model for Post-Combustion Chemical Absorption CO₂ Capture Plant', *22 European Symposium on Computer Aided Process Engineering*, 30, pp. 807-811.
- Bishop, C. (1991) 'Improving the generalisation properties of radial basis function neural networks', *Neural Computation*, 13, pp. 579-588.
- Bishop, C. (1995) *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press.
- Bristol, E.H. (1966) 'On a new measure of interaction for multivariable process control', *IEEE Transaction on Automatic Control*, 11(1), pp. 133-134.
- Caplow, M. (1968) 'Kinetics of carbamate formation and breakdown', *Journal of the American Chemical Society*, 90(24), pp. 6795-6803.
- Chachuat, B., Srinivasan, B. and Bonvin, D. (2009) 'Adaptation strategies for real-time optimization', *Computers & Chemical Engineering*, 33(10), pp. 1557-1567.
- Chen, D. and Seborg, D.E. (2002) 'Relative gain array analysis for uncertain process models', *Aiche Journal*, 48(2), pp. 302-310.

- Chu, F.M., Yang, L.J., Du, X.Z. and Yang, Y.P. (2016) 'CO₂ capture using MEA (monoethanolamine) aqueous solution in coal-fired power plants: Modeling and optimization of the absorbing columns', *Energy*, 109, pp. 495-505.
- Colantuono, G. and Cockerill, T. (2017) 'Selective strategy for solid sorbent replacement in CCS', *Chemical Engineering Research and Design* 120, pp. 82-91.
- Crooks, J.E. and Donnellan, J.P. (1989) 'Kinetics and Mechanism of the Reaction between Carbon-Dioxide and Amines in Aqueous-Solution', *Journal of the Chemical Society-Perkin Transactions 2*, (4), pp. 331-333.
- Darby, M.L. and White, D.C. (1988) 'On-Line Optimization of Complex Process Units', *Chem. Eng. Prog.*, 67(10), pp. 51-59.
- Davidson, R.M. (2007) 'Post-combustion Carbon Capture from Coal Fired Plants-Solvent Scrubbing', *IEA Clean Coal Centre*, (GGG/125).
- Davis, J. and Rochelle, G. (2009) 'Thermal degradation of monoethanolamine at stripper conditions', *Greenhouse Gas Control Technologies* 9, 1(1), pp. 327-333.
- Davison, J. (2007) 'Performance and costs of power plants with capture and storage of CO₂', *Energy*, 32(7), pp. 1163-1176.
- de Riva, J., Suarez-Reyes, J., Moreno, D., Ferro, V. and Palomar, J. (2017) 'Ionic liquids for post-combustion CO₂ capture by physical absorption: Thermodynamic, kinetic and process analysis', *International Journal of Greenhouse Gas Control*, 61, pp. 61-70.
- Decoursey, W.J. and Thring, R.W. (1989) 'Effects of Unequal Diffusivities on Enhancement Factors for Reversible and Irreversible Reaction', *Chemical Engineering Science*, 44(8), pp. 1715-1721.
- Donaldson, T.L. and Nguyen, Y.N. (1980) 'Carbon-Dioxide Reaction-Kinetics and Transport in Aqueous Amine Membranes', *Industrial & Engineering Chemistry Fundamentals*, 19(3), pp. 260-266.
- Ebrahimpour, R., Amini, M. and Sharifzadehi, F. (2011) 'Farsi handwritten recognition using combining neural networks based on stacked generalization', *International Journal on Electrical Engineering and Informatics*, 3(2), p. 1460164.
- Efron, B. (1982) *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia: Society for Industrial and Applied Mathematics.

- EIA (2016) *International Energy Outlook 2016* (DOE/EIA-0484(2016)). Washington DC, U.S.: U.S. Energy Information Administration.
- Ellis, M., Durand, H. and Christofides, P.D. (2014) 'A tutorial review of economic model predictive control methods', *Journal of Process Control*, 24(8), pp. 1156-1178.
- Elvira, G.-B., Francisco, G.-C., Víctor, S.-M. and Alberto, M.L.R. (2016) 'MgO-based adsorbents for CO₂ adsorption: Influence of structural and textural properties on the CO₂ adsorption performance', *Journal of Environmental Sciences (China)* p. 11.
- Engell, S. (2007) 'Feedback control for optimal process operation', *Journal of Process Control*, 17(3), pp. 203-219.
- Favre, E. and Svendsen, H.F. (2012) 'Membrane contactors for intensified post-combustion carbon dioxide capture by gas-liquid absorption processes', *Journal of Membrane Science*, 407, pp. 1-7.
- Frenay, B. and Verleysen, M. (2016) 'Reinforced Extreme Learning Machines for Fast Robust Regression in the Presence of Outliers', *IEEE Transactions on Cybernetics*, 46(12), pp. 3351-3363.
- Gabelman, A. and Hwang, S.T. (1999) 'Hollow fiber membrane contactors', *Journal of Membrane Science*, 159(1-2), pp. 61-106.
- Gaspar, J. and Cormos, A.M. (2011) 'Dynamic modeling and validation of absorber and desorber columns for post-combustion CO₂ capture', *Computers & Chemical Engineering*, 35(10), pp. 2044-2052.
- Gaspar, J., Jorgensen, J.B. and Fosbol, P.L. (2015) 'Control of a post-combustion CO₂ capture plant during process start-up and load variations', *Ifac Papersonline*, 48(8), pp. 580-585.
- Hart, A. and Gnanendran, N. (2009) 'Cryogenic CO₂ Capture in Natural Gas', *Energy Procedia*, 1(1), pp. 697-706.
- Harun, N., Nittaya, T., Douglas, P.L., Croiset, E. and Ricardez-Sandoval, L.A. (2012) 'Dynamic simulation of MEA absorption process for CO₂ capture from power plants', *International Journal of Greenhouse Gas Control*, 10, pp. 295-309.
- He, Z.R., Sahraei, M.H. and Ricardez-Sandoval, L.A. (2016) 'Flexible operation and simultaneous scheduling and control of a CO₂ capture plant using model predictive control', *International Journal of Greenhouse Gas Control*, 48, pp. 300-311.

- Hinton, G.E., Osindero, S. and Teh, Y.W. (2006) 'A fast learning algorithm for deep belief nets', *Neural Computation*, 18(7), pp. 1527-1554.
- Hu, X.F., Zhao, Z., Wang, S., Wang, F.L., He, D.K. and Wu, S.K. (2008) 'Multi-stage extreme learning machine for fault diagnosis on hydraulic tube tester', *Neural Computing & Applications*, 17(4), pp. 399-403.
- Huang, G.B. and Siew, C.K. (2004) 'Extreme learning machine: RBF network case', *2004 8th International Conference on Control, Automation, Robotics and Vision, Vols 1-3*, pp. 1029-1036.
- Huang, G.B., Zhu, Q.Y. and Siew, C.K. (2004) 'Extreme learning machine: A new learning scheme of feedforward neural networks', *2004 Ieee International Joint Conference on Neural Networks, Vols 1-4, Proceedings*, pp. 985-990.
- Huang, G.B., Zhu, Q.Y. and Siew, C.K. (2006) 'Extreme learning machine: Theory and applications', *Neurocomputing*, 70(1-3), pp. 489-501.
- Huynh, H.T., Won, Y. and Kim, J.J. (2008) 'An Improvement of Extreme Learning Machine for Compact Single-Hidden-Layer Feedforward Neural Networks', *International Journal of Neural Systems*, 18(5), pp. 433-441.
- International Energy Agency. and Organisation for Economic Co-operation and Development. (2004) *Prospects for CO₂ capture and storage*. Paris: International Energy Agency/Organisation for Economic Co-operation and Development.
- Iosifidis, A., Tefas, A. and Pitas, I. (2015) 'Large-Scale Nonlinear Facial Image Classification Based on Approximate Kernel Extreme Learning Machine', *2015 Ieee International Conference on Image Processing (Icip)*, pp. 2449-2453.
- Kailath, T. (1980) *Linear systems*. Englewood Cliffs, NJ: Prentice-Hall.
- Kenarsari, S.D., Yang, D.L., Jiang, G.D., Zhang, S.J., Wang, J.J., Russell, A.G., Wei, Q. and Fan, M.H. (2013) 'Review of recent advances in carbon dioxide separation and capture', *Rsc Advances*, 3(45), pp. 22739-22773.
- Kenig, E.Y., Schneider, R. and Gorak, A. (2001) 'Reactive absorption: Optimal process design via optimal modelling', *Chemical Engineering Science*, 56(2), pp. 343-350.
- Klimes, J., Bulatov, I. and Cockerill, T. (2007) 'Techno-economic modelling and cost functions of CO₂ capture processes', *Computers & Chemical Engineering*, 31(5-6), pp. 445-455.

- Koivo, H.N. and Tantt, J.T. (1991) 'Tuning of PID Controllers: Survey of SISO and MIMO Techniques', *Proceedings of the IFAC Intelligent Tuning and Adaptive Control Symposium (Singapore)*, p. 758.
- Kucka, L., Muller, I., Kenig, E.Y. and Gorak, A. (2003) 'On the modelling and simulation of sour gas absorption by aqueous amine solutions', *Chemical Engineering Science*, 58(16), pp. 3571-3578.
- Kvamsdal, H.M., Jakobsen, J.P. and Hoff, K.A. (2009) 'Dynamic modeling and simulation of a CO₂ absorber column for post-combustion CO₂ capture', *Chemical Engineering and Processing*, 48(1), pp. 135-144.
- Lawal, A., Wang, M., Stephenson, P., Koumpouras, G. and Yeung, H. (2010) 'Dynamic modelling and analysis of post-combustion CO₂ chemical absorption process for coal-fired power plants', *Fuel*, 89(10), pp. 2791-2801.
- Lawal, A., Wang, M., Stephenson, P. and Yeung, H. (2009a) 'Dynamic modelling of CO₂ absorption for post combustion capture in coal-fired power plants', *Fuel*, 88(12), pp. 2455-2462.
- Lawal, A., Wang, M.H., Stephenson, P. and Obi, O. (2012) 'Demonstrating full-scale post-combustion CO₂ capture for coal-fired power plants through dynamic modelling and simulation', *Fuel*, 101, pp. 115-128.
- Lawal, A., Wang, M.H., Stephenson, P. and Yeung, H. (2009b) 'Dynamic Modeling and Simulation of CO₂ Chemical Absorption Process for Coal-Fired Power Plants', *10th International Symposium on Process Systems Engineering*, 27, pp. 1725-1730.
- Lee, A.S., Eslick, J.C., Miller, D.C. and Kitchin, J.R. (2013) 'Comparisons of amine solvents for post-combustion CO₂ capture: A multi-objective analysis approach', *International Journal of Greenhouse Gas Control*, 18, pp. 68-74.
- Lee, S.C., Choi, B.Y., Lee, T.J., Ryu, C.K., Soo, Y.S. and Kim, J.C. (2006) 'CO₂ absorption and regeneration of alkali metal-based solid sorbents', *Catalysis Today*, 111(3-4), pp. 385-390.
- Liao, B., Xu, J.G., Lv, J.T. and Zhou, S.L. (2015) 'An Image Retrieval Method for Binary Images Based on DBN and Softmax Classifier', *Iete Technical Review*, 32(4), pp. 294-303.

- Lin, Y.J., Wong, D.S.H., Jang, S.S. and Ou, J.J. (2012) 'Control strategies for flexible operation of power plant with CO₂ capture plant', *Aiche Journal*, 58(9), pp. 2697-2704.
- Love, J. (2007) *Process Automation Handbook : A Guide to Theory and Practice*. London: Springer-Verlag London Limited.
- Luu, M.T., Manaf, N.A. and Abbas, A. (2015) 'Dynamic modelling and control strategies for flexible operation of amine-based post-combustion CO₂ capture systems', *International Journal of Greenhouse Gas Control*, 39, pp. 377-389.
- Luyben, W.L. (1986) 'Simple Method for Tuning Siso Controllers in Multivariable Systems', *Industrial & Engineering Chemistry Process Design and Development*, 25(3), pp. 654-660.
- Mac Dowell, N. and Shah, N. (2013) 'Identification of the cost-optimal degree of CO₂ capture: An optimisation study using dynamic process models', *International Journal of Greenhouse Gas Control*, 13, pp. 44-58.
- MacKay, D.J.C. (1992) 'Bayesian interpolation', *Neural Computation* 4(415-447).
- Manaf, N.A., Cousins, A., Feron, P. and Abbas, A. (2016) 'Dynamic modelling, identification and preliminary control analysis of an amine-based post-combustion CO₂ capture pilot plant', *Journal of Cleaner Production*, 113, pp. 635-653.
- Marquardt, D. (1963) 'An algorithm for least squares estimation of nonlinear parameters', *SIAM J. Appl. Math.* , 11, pp. 431-441.
- Mc Avoy, T., Arkun, Y., Chen, R., Robinson, D. and Schnelle, P.D. (2003) 'A new approach to defining a dynamic relative gain', *Control Engineering Practice*, 11(8), pp. 907-914.
- Meindersma, G.W. and de Haan, A.B. (2008) 'Conceptual process design for aromatic/aliphatic separation with ionic liquids', *Chemical Engineering Research & Design*, 86(7a), pp. 745-752.
- Meisen, A. and Shuai, X.S. (1997) 'Research and development issues in CO₂ capture', *Energy Conversion and Management*, 38, pp. S37-S42.
- Metz, B. and Intergovernmental Panel on Climate Change. Working Group III. (2005) *IPCC special report on carbon dioxide capture and storage*. Cambridge: Cambridge University Press, for the Intergovernmental Panel on Climate Change.

- Metz, B. and Intergovernmental Panel on Climate Change. Working Group III. (2007) *Climate change 2007 : mitigation of climate change : contribution of Working Group III to the Fourth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge ; New York: Cambridge University Press.
- Minchener, A. (2016) 'Coal-fired power generation in Japan and the world', *Cornerstone*, 4(4), pp. 24-29.
- Monemian, S.A., Shahsavan, H., Bolouri, O., Taranejoo, S., Goodarzi, V. and Torabi-Angaji, M. (2010) 'A Stacked Neural Network Approach for Yield Prediction of Propylene Polymerization', *Journal of Applied Polymer Science*, 116(3), pp. 1237-1246.
- Morari, M. and Lee, J.H. (1999) 'Model predictive control: past, present and future', *Computers & Chemical Engineering*, 23(4-5), pp. 667-682.
- Mores, P., Rodriguez, N., Scenna, N. and Mussati, S. (2012a) 'CO₂ capture in power plants: Minimization of the investment and operating cost of the post-combustion process using MEA aqueous solution', *International Journal of Greenhouse Gas Control*, 10, pp. 148-163.
- Mores, P., Scenna, N. and Mussati, S. (2012b) 'CO₂ capture using monoethanolamine (MEA) aqueous solution: Modeling and optimization of the solvent regeneration and CO₂ desorption process', *Energy*, 45(1), pp. 1042-1058.
- Mosadegh-Sedghi, S., Rodrigue, D., Brisson, J. and Iliuta, M.C. (2014) 'Wetting phenomenon in membrane contactors - Causes and prevention', *Journal of Membrane Science*, 452, pp. 332-353.
- Naceur, K.B. and McCulloch, S. (2016) 'The Role of CCS in a Well-Below 2 °C World', *Cornerstone*, 4(4), pp. 13-16.
- Nittaya, T., Douglas, P.L., Croiset, E. and Ricardez-Sandoval, L.A. (2014) 'Dynamic modelling and control of MEA absorption processes for CO₂ capture from power plants', *Fuel*, 116, pp. 672-691.
- O'Dwyer, A. (2003) *Handbook of PI and PID Controller Tuning Rules*. London: Imperial College Press.
- O'Dwyer, A. (2006) 'PI and PID controller tuning rules: an overview and personal perspective', *Proceedings of the IET Irish Signals and Systems Conference*. Dublin. Dublin Institute of Technology, pp. 161-166.

- O'Dwyer, A. (2009) *Handbook of PI and PID controller tuning rules*. 3rd edn. London: Imperial College Press.
- Panahi, M. and Skogestad, S. (2011) 'Economically efficient operation of CO₂ capturing process part I: Self-optimizing procedure for selecting the best controlled variables', *Chemical Engineering and Processing*, 50(3), pp. 247-253.
- Panahi, M. and Skogestad, S. (2012) 'Economically efficient operation of CO₂ capturing process. Part II. Design of control layer', *Chemical Engineering and Processing*, 52, pp. 112-124.
- Pintola, T.T., P. Meisen, A. (1993) 'Simulation of pilot plant and industrial CO₂-MEA absorbers ', *Gas Separation and Purification*, 7(1), pp. 47-52.
- Pontes, K.V., Wolf, I.J., Embirucu, M. and Marquardt, W. (2015) 'Dynamic Real-Time Optimization of Industrial Polymerization Processes with Fast Dynamics', *Industrial & Engineering Chemistry Research*, 54(47), pp. 11881-11893.
- Prolss, K., Tummescheit, H., Velut, S. and Akesson, J. (2011) 'Dynamic model of a post-combustion absorption unit for use in a non-linear model predictive control scheme', *10th International Conference on Greenhouse Gas Control Technologies*, 4, pp. 2620-2627.
- Qin, S.J. and Badgwell, T.A. (2003) 'A survey of industrial model predictive control technology', *Control Engineering Practice*, 11(7), pp. 733-764.
- Rao, A.B. and Rubin, E.S. (2006) 'Identifying cost-effective CO₂ control levels for amine-based CO₂ capture systems', *Industrial & Engineering Chemistry Research*, 45(8), pp. 2421-2429.
- Rao, A.B., Rubin, E.s., Berkenpas, M.B (2004) 'An Integrated Modelling Framework for Carbon Management Technologies'.
- Ren, Y., Mao, J., Liu, Y. and Li, Y. (2017) 'A novel DBN model for time seires forecasting', *IAENG International Journal of Computer Science*, 44(1), pp. 79-86.
- S. B. Fredriksen¹, K.-J.J. (2013) 'Oxidative degradation of aqueous amine solutions of MEA, AMP, MDEA, Pz: A review', *Energy Procedia*, 37, pp. 1770-1777.
- Sahraei, M.H. and Ricardez-Sandoval, L.A. (2014) 'Controllability and optimal scheduling of a CO₂ capture plant using model predictive control', *International Journal of Greenhouse Gas Control*, 30, pp. 58-71.

- Sequeira, S.E., Graells, M. and Puigjaner, L. (2002) 'Real-time evolution for on-line optimization of continuous processes', *Industrial & Engineering Chemistry Research*, 41(7), pp. 1815-1825.
- Sexton, A.J. and Rochelle, G.T. (2009) 'Catalysts and Inhibitors for MEA Oxidation', *Greenhouse Gas Control Technologies* 9, 1(1), pp. 1179-1185.
- Shang, C., Yang, F., Huang, D.X. and Lyu, W.X. (2014) 'Data-driven soft sensor development based on deep learning technique', *Journal of Process Control*, 24(3), pp. 223-233.
- Shao R, S.A. (2009) 'Amines used in CO₂ capture – health and environmental impacts', *The Bellona Foundation*, p. 49.
- Sharma, S.D. and Azzi, M. (2014) 'A critical review of existing strategies for emission control in the monoethanolamine-based carbon capture process and some recommendations for improved strategies', *Fuel*, 121, pp. 178-188.
- Shiflett, M.B., Drew, D.W., Cantini, R.A. and Yokozeki, A. (2010) 'Carbon Dioxide Capture Using Ionic Liquid 1-Butyl-3-methylimidazolium Acetate', *Energy & Fuels*, 24, pp. 5781-5789.
- Sipocz, N., Tobiesen, F.A. and Assadi, M. (2011) 'The use of Artificial Neural Network models for CO₂ capture plants', *Applied Energy*, 88(7), pp. 2368-2376.
- Sreedhar, I., Nahar, T., Venugopal, A. and Srinivas, B. (2017) 'Carbon capture by absorption – Path covered and ahead ', *Renewable and Sustainable Energy Reviews*, 76, pp. 1080-1107.
- Sreenivasulu, B., Sreedhar, I., Suresh, P. and Raghavan, K.V. (2015) 'Development Trends in Porous Adsorbents for Carbon Capture', *Environmental Science & Technology*, 49(21), pp. 12641-12661.
- Sridhar, D.V., Seagrave, R.C. and Bartlett, E.B. (1996) 'Process modeling using stacked neural networks', *Aiche Journal*, 42(9), pp. 2529-2539.
- Stanley, G., Marinogalarraga, M. and Mcavoy, T.J. (1985) 'Shortcut Operability Analysis .1. The Relative Disturbance Gain', *Industrial & Engineering Chemistry Process Design and Development*, 24(4), pp. 1181-1188.
- Sun, Z.L., Choi, T.M., Au, K.F. and Yu, Y. (2008) 'Sales forecasting using extreme learning machine with applications in fashion retailing', *Decision Support Systems*, 46(1), pp. 411-419.

- Thiruvengkatachari, R., Su, S., An, H. and Yu, X.X. (2009) 'Post combustion CO₂ capture by carbon fibre monolithic adsorbents', *Progress in Energy and Combustion Science*, 35(5), pp. 438-455.
- Tock, L. and Marechal, F. (2014) 'Process design optimization strategy to develop energy and cost correlations of CO₂ capture processes', *Computers & Chemical Engineering*, 61, pp. 51-58.
- Uddin, M.Z. and Kim, J. (2017) 'A Robust Approach for Human Activity Recognition Using 3-D Body Joint Motion Features with Deep Belief Network', *Ksii Transactions on Internet and Information Systems*, 11(2), pp. 1118-1133.
- Vaidya, P.D. and Kenig, E.Y. (2007) 'CO₂-alkanolamine reaction kinetics: A review of recent studies', *Chemical Engineering & Technology*, 30(11), pp. 1467-1474.
- von Harbou, I., Mangalapally, H.P. and Hasse, H. (2013) 'Pilot plant experiments for two new amine solvents for post-combustion carbon dioxide capture', *International Journal of Greenhouse Gas Control*, 18, pp. 305-314.
- Wang, M., Lawal, A., Stephenson, P., Sidders, J. and Ramshaw, C. (2011) 'Post-combustion CO₂ capture with chemical absorption: A state-of-the-art review', *Chemical Engineering Research & Design*, 89(9), pp. 1609-1624.
- Wang, X., Mahalec, V. and Qian, F. (2017) 'Globally optimal dynamic real time optimization without model mismatch between optimization and control layer ', *Computers and Chemical Engineering* 104, pp. 64-75.
- WeRek, R.M., Brunson, R.J. and Law, F.H. (1978) 'Enhancement factors for gas absorption with second-order irreversible chemical reaction', *Can J Chem Eng*, 56, pp. 181-186.
- Witcher, M.F. and Mcavoy, T.J. (1977) 'Interacting Control-Systems - Steady-State and Dynamic Measurement of Interaction', *Isa Transactions*, 16(3), pp. 35-41.
- Wold, S. (1978) 'Cross validatory estimation of the number of components in factor and principal components models', *Technometrics*, 20, pp. 397-404.
- Wolpert, D.H. (1992) 'Stacked generalization', *Neural Networks* 5, pp. 241-259.
- Wu, Y.X., Zhou, Q. and Chan, C.W. (2010) 'A comparison of two data analysis techniques and their applications for modeling the carbon dioxide capture process', *Engineering Applications of Artificial Intelligence*, 23(8), pp. 1265-1276.

- Zaidiza, D.A., Wilson, S.G., Belaisaoui, B., Rode, S., Castel, C., Roizard, D. and Favre, E. (2016) 'Rigorous modelling of adiabatic multicomponent CO₂ post-combustion capture using hollow fibre membrane contactors', *Chemical Engineering Science*, 145, pp. 45-58.
- Zanganeh, K.E., Shafeen, A. and Salvador, C. (2009) 'CO₂ Capture and Development of an Advanced Pilot-Scale Cryogenic Separation and Compression Unit', *Energy Procedia*, 1(1), pp. 247-252.
- Zhang, J. (1999) 'Developing robust non-linear models through bootstrap aggregated neural networks', *Neurocomputing*, 25(1-3), pp. 93-113.
- Zhang, J. (2001) 'Developing robust neural network models by using both dynamic and static process operating data', *Industrial & Engineering Chemistry Research*, 40(1), pp. 234-241.
- Zhang, J. (2002) 'Improved on-line process fault diagnosis using stacked neural networks', *Proceedings of the 2002 Ieee International Conference on Control Applications, Vols 1 & 2*, pp. 689-694.
- Zhang, J. (2004) 'A reliable neural network model based optimal control strategy for a batch polymerization reactor', *Industrial & Engineering Chemistry Research*, 43(4), pp. 1030-1038.
- Zhang, J., Martin, E.B., Morris, A.J. and Kiparissides, C. (1997) 'Inferential estimation of polymer quality using stacked neural networks', *Computers & Chemical Engineering*, 21, pp. S1025-S1030.
- Zhang, J., Morris, A.J., Martin, E.B. and Kiparissides, C. (1998) 'Prediction of polymer quality in batch polymerisation reactors using robust neural networks', *Chemical Engineering Journal*, 69(2), pp. 135-143.
- Zhang, Q., Turton, R. and Bhattacharyya, D. (2016) 'Development of Model and Model-Predictive Control of an MEA-Based Postcombustion CO₂ Capture Process', *Industrial & Engineering Chemistry Research*, 55(5), pp. 1292-1308.
- Zhou, Q., Chan, C.W., Tontiwachiwuthikul, P., Idem, R. and Gelowitz, D. (2009) 'A statistical analysis of the carbon dioxide capture process', *International Journal of Greenhouse Gas Control*, 3(5), pp. 535-544.

- Zhou, Q., Chan, C.W. and Tontiwachwuthikul, P. (2010a) 'An application of neuro-fuzzy technology for analysis of the CO₂ capture process', *Fuzzy Sets and Systems*, 161(19), pp. 2597-2611.
- Zhou, Q., Wu, Y.X., Chan, C.W. and Tontiwachwuthikul, P. (2010b) 'Applications of Three Data Analysis Techniques for Modeling the Carbon Dioxide Capture Process', *2010 23rd Canadian Conference on Electrical and Computer Engineering (Ccece)*.
- Zhu, Q.Y., Qin, A.K., Suganthan, P.N. and Huang, G.B. (2005) 'Evolutionary extreme learning machine', *Pattern Recognition*, 38(10), pp. 1759-1763.
- Zhu, X.M., Hong, W.R. and Wang, S.Q. (2004) 'Implementation of Advanced Control for a Heat-Integrated Distillation Column System', *Iecon 2004: 30th Annual Conference of Ieee Industrial Electronics Society, Vol 3*, pp. 2006-2011.
- Ziaii, S., Rochelle, G.T. and Edgar, T.F. (2009) 'Dynamic Modeling to Minimize Energy Use for CO₂ Capture in Power Plants by Aqueous Monoethanolamine', *Industrial & Engineering Chemistry Research*, 48(13), pp. 6105-6111.
- Ziegler, J.G. and Nichols, N.B. (1942) 'Optimum settings for automatic controllers', *Trans. ASME*, 64, p. 759.
- Zoannou, K.S., Sapsford, D.J. and Griffiths, A.J. (2013) 'Thermal degradation of monoethanolamine and its effect on CO₂ capture capacity', *International Journal of Greenhouse Gas Control*, 17, pp. 423-430.