Modelling Activated Sludge Wastewater Treatment Plants

Using Artificial Intelligence Techniques

(Fuzzy Logic and Neural Networks)

Rabee Rustum

Submitted for the degree of Doctor of Philosophy

Heriot-Watt University

School of the Built Environment

April, 2009

ABSTRACT

Activated sludge process (ASP) is the most commonly used biological wastewater treatment system. Mathematical modelling of this process is important for improving its treatment efficiency and thus the quality of the effluent released into the receiving water body. This is because the models can help the operator to predict the performance of the plant in order to take cost-effective and timely remedial actions that would ensure consistent treatment efficiency and meeting discharge consents. However, due to the highly complex and non-linear characteristics of this biological system, traditional mathematical modelling of this treatment process has remained a challenge.

This thesis presents the applications of Artificial Intelligence (AI) techniques for modelling the ASP. These include the Kohonen Self Organising Map (KSOM), backpropagation artificial neural networks (BPANN), and adaptive network based fuzzy inference system (ANFIS). A comparison between these techniques has been made and the possibility of the hybrids between them was also investigated and tested.

The study demonstrated that AI techniques offer viable, flexible and effective modelling methodology alternative for the activated sludge system. The KSOM was found to be an attractive tool for data preparation because it can easily accommodate missing data and outliers and because of its power in extracting salient features from raw data. As a consequence of the latter, the KSOM offers an excellent tool for the visualisation of high dimensional data. In addition, the KSOM was used to develop a software sensor to predict biological oxygen demand. This soft-sensor represents a significant advance in real-time BOD operational control by offering a very fast estimation of this important wastewater parameter when compared to the traditional 5-days bio-essay BOD test procedure. Furthermore, hybrids of KSOM-ANN and KSOM-ANFIS were shown to result much more improved model performance than using the respective modelling paradigms on their own.

ACKNOWLEDGMENTS

It has been a long journey since I started the work presented in this thesis. So when it comes to the time to write this page, it would not be easy to enumerate all of the support I have received throughout my studies. However, I am sure that without many of them, the thesis would not have become a reality.

Words are not enough to express my thanks to my supervisor, Dr Adebayo Adeloye, who has been vigorously supporting my studies and guiding my thesis. Without his continuous interest and encouragement, numerous fruitful hours of discussion and support, this work would have been substantially more difficult. During these long hours of discussion, many cups of coffee were emptied whilst the trays of ideas were filled. He has a tremendous capability to put things into a broader context and to discern the important elements of the research.

I would also like to thank the editors and reviewers of my published work for their valuable comments and a number of suggestions that were incorporated into the thesis. There are other people who have shown great interest in my work. I would especially like to thank the process control staff of Thames Water, especially Mr Duncan Taylor and Mr Paul Banfield for their patience and practical assistance.

I would like to acknowledge Damascus University for their financial support during my stay in Edinburgh, UK. In addition, I would also like to express my sincere thanks to my family and friends, because without their patience and support I would not have overcome the frustrations of this working period.

I would like to express special thanks to my wife, Ola, who always reminded me on the existence of life beside modelling and simulation. I am particularly grateful to her for her patient during those long evenings and weekends I spent on the work rather than with her. I would like to thank my beloved son, Moghny, for allowing me to be absent of his life and for being quiet and understandable during the writing of this thesis.

Research Thesis Submission

Name:	Rabee Rustum		
School/PGI:	School of the Built Environment		
Version: *(i.e. First, Resubmission, Final)*	Final	Degree Sought:	PhD

Declaration

In accordance with the appropriate regulations, I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate:		Date:	

Submission

Submitted By *(name in capitals):*	Rabee Rustum
Signature of Individual Submitting:	
Date Submitted:	

For Completion in Academic Registry

Received in the Academic Registry by *(name in capitals):*			
Method of Submission *(Handed in to Academic Registry; posted through internal/external mail):*			
Signature:		Date:	

TABLE OF CONTENTS

ABSTRACT ii

ACKNOWLEDGMENTS iii

TABLE OF CONTENTS v

LIST OF PUBLICATIONS x

CHAPTER 1 ..... **11**
INTRODUCTION **11**
1.1 Problem background 11

1.2 Aim and Objectives 15

1.3 Structure of the thesis 16

CHAPTER 2 ..... **20**
ACTIVATED SLUDGE WASTEWATER TREATMENT SYSTEMS **20**
2.1 Importance of wastewater treatment 20

2.2 Brief history of wastewater treatment 21

2.3 The Structure of Wastewater Treatment Plant 23

2.3.1 Preliminary treatment 23
2.3.2 Storm Tank 25
2.3.3 Primary treatment 26
2.3.4 Secondary treatment 26
2.4 Activated Sludge Process 28

2.4.1 The Secondary Clarifier 30
2.4.2 Operation of Activated sludge system 31
2.4.3 Modelling activated sludge process 33
2.4.4 Types of Modelling of the activated sludge process 36
2.4.4.1 Fundamental models of the activated sludge wastewater treatment process 36
a. Primary clarifier models 36
b. Biological reactor models 38
c. Secondary clarifier model 45
d. Limitations of fundamental models 48
2.4.4.2 Empirical models of wastewater treatment plants 49
2.5 Summary 58

CHAPTER 3 ..... **60**
ARTIFICIAL INTELLIGENCE TECHNIQUES (AI) **60**
3.1 Introduction 60

3.2 Artificial Neural Networks 61

3.2.1 Historical overview 62

3.2.2 Inspiration from Neuroscience 63
3.2.3 The architecture of ANN 64
3.2.3.1 The neuron 64
3.2.3.2 Types of transfer Functions 67
3.2.4 Neural network topology 68
3.2.4.1 Feed-forward neural networks (FFNNs) 69
3.2.4.2 Recurrent Neural Networks (RNNs) 70
3.2.5 Modelling using ANN 71
3.2.6 Learning Algorithm 72
3.2.6.1 Supervised learning algorithms 72
3.2.6.1.1 Back-propagation algorithm 72
3.2.6.1.2 Enhancing the performance of backpropagation neural networks 75
a. Sample size 76
b. Pre-processing and post-processing 76
c. Early Stopping 77
d. Selection of the network elements 78
3.2.7 Advantages and Limitations of Neural Networks 78
3.2.6.2 Unsupervised learning algorithms 79
3.3 Kohonen Self-Organising Map (KSOM) 80
3.3.1 The structure of the KSOM 81
3.3.2 Training the KSOM 84
3.3.2.1 Sequential training algorithms 84
3.3.2.2 Batch training algorithms 87
3.3.3 The quality of the KSOM 91
3.3.4 Applications of KSOM 92
3.4 Fuzzy Logic modelling technique 93
3.4.1 Fuzzy Logic concepts 94
3.4.2 Fuzzy Sets 95
3.4.3 Fuzzy Logic System 97
3.4.3.1 Architecture of Fuzzy Logic System 98
1 The fuzzification unit 100
2 The Knowledge base 101
a. The database 101
b. The rule base 102
i) The choice of input variables 103
ii) Generation method of fuzzy rules 103
iii) Implementation method of fuzzy rules 103
3 The fuzzy inference 105
4 The defuzzification unit 105
3.4.3.2 Types of fuzzy logic systems 107
a. Mamdani fuzzy model 108
b. Takagi and Sugeno model 108
3.5 Hybrid Neuro-Fuzzy Systems 110

3.5.1 Adaptive-Network-based Fuzzy Inference System (ANFIS) 113
3.6 Features Selection, Extraction, and visualisation 118
3.6.1 Features Selection 118
3.6.2 Features Extraction 119
3.7 Summary 124

CHAPTER 4 125
METHODS AND MATERIALS 125
4.1 Case study 125
4.2 Data 130
4.2.1 Data collection system 130
4.2.2 Data Description 130
4.2.2.1 Biochemical Oxygen Demand (BOD) 131
4.2.2.2 Chemical Oxygen Demand (COD) 131
4.2.2.3 Dissolved Oxygen (DO) 132
4.2.2.4 Mixed Liquor Suspended Solids (MLSS and RAS-MLSS) 132
4.2.2.5 Quantification of settling properties (SSVI) 133
4.2.2.6 Flow 133
a. Flow to full treatment 133
b. Return activated sludge (RAS) 134
c. Waste or surplus activated sludge (WAS) 134
4.2.2.7 Sludge Age 135
4.2.2.8 Food to Micro-organisms ratio (F/M) 135
4.2.3 Data treatment 135
4.4 Computer Software 136
4.5 Selecting model structure 137
4.5.1 KSOM modelling Strategy 137
4.5.2 Artificial Neural Network modelling strategy 137
4.5.3 ANFIS modelling strategy 138
4.6 Models performance evaluation criteria 139
4.7 Visualizing the performance of the model 142
4.8 Summary 142

CHAPTER 5 143
APPLICATION I: ACTIVATED SLUDGE DATA PRE-PROCESSING USING THE KOHONEN SELF ORGANISING MAP 143
5.1 Introduction 143
5.2 Outliers and Missing Values 144
5.2.1 Outliers 144
5.2.2 Missing Values 146
5.3 Replacing outliers and missing values 147

5.4 Overview of Predicting Missing values using KSOM 148

5.5 Experimental data 150

5.6 SOM analysis 153

5.7 Results and discussion 153

5.8 Discussion of Results 161

5.9 Conclusion 163

Chapter 6 164
Application II: Applying Kohonen Self-Organizing Map as a Software Sensor to Predict the Biochemical Oxygen Demand 164
6.1 Introduction 164

6.2 Review of alternatives to bioassay determination of the BOD5 166

6.3. Methodology 169

6.3.1 Case Study 169

6.3.2 Numerical Analysis and Modelling 171

6.4 Results 173

6.5 Discussion of Results 178

6.6 Summary 183

Chapter 7 184
Application III: Modelling Activated Sludge Process using hybrid KSOM-ANN 184
7.1 Introduction 184

7.2 Methodology 185

7.3 RESULTS AND DISCUSSION 190

7.3.1 Artificial neural networks models using raw data 190

7.3.2 Artificial neural networks using the features of the data 196

7.3.3 Comparison between the models 196

7.4 Conclusion 211

CHAPTER 8 213
APPLICATION IV: MODELLING ASP USING HYBRID KSOM-ANFIS 213
8.1 Introduction 213

8.2 Methodology 214

8.3 Results and Discussion 217

8.4 Conclusion 226

CHAPTER 9 240
DISCUSSION AND CONCLUSIONS 240
9.1 Discussion 240

9.1.1 Data preparation 241
9.1.2 Features extraction and data visualization 241
9.1.3 Software sensor for fast predicting of BOD_5 242
9.1.4 Modelling ASP using AI paradigms 243
9.2 Conclusions 247
9.3 Recommendation for further work 249
REFERENCES 290
RESUME 308

LIST OF PUBLICATIONS

Journal Papers:

- Rustum R., Adeloye A.J., and Scholz M., 2008. Applying Kohonen Self-organizing Map as a Software Sensor to Predict the Biochemical Oxygen Demand. Water Environment Research, 80 (1), 32-40.

- Rustum R and Adeloye A.J., 2007. Replacing outliers and missing values from activated sludge data using Kohonen Self Organizing Map. Journal of Environmental Engineering (ASCE), 133 (9), 909-916.

- Rustum R., Adeloye A.J., and Simala A., 2007. Kohonen self-organizing map (KSOM) extracted features for enhancing MLP-ANN prediction models of BOD_5. Water Quality and Sediment Behaviour of the Future, IAHS Publ., Red Book No. 314, 181-187.

Conference Proceedings

- Rustum R. and Adeloye A.J., 2008. Activated Sludge Data Analysis using Kohonen Self-Organising Map (a case study of Seafield wastewater treatment plant). 1st Postgraduate Researchers' Conference on 'Meeting Environmental Challenges in the Coastal Region of Nigeria, 29 – 30 September 2008, Dundee, United Kingdom.

- Rustum R and Adeloye A.J., 2007. Modelling activated sludge process using adaptive network-based fuzzy inference system. Presented at the 10th IWA Specialised Conference on Design, Operation and Economics of Large Wastewater Treatment Plants, 9-13 September, Vienna, Austria: 101-104.

- Rustum R. and Adeloye A.J., 2006. Features Extraction from Primary Clarifier data using Kohonen Self Organizing Map, Unsupervised Neural Network. Presented at the 7th International Conference on Hydroinformatics, HIC, September 2006, Nice, FRANCE, 918-925.

- Rustum R. and Adeloye A.J., 2006. Knowledge Discovery from Activated Sludge Process Using Unsupervised Neural Networks (Kohonen Self-Organizing Map). 3rd international symposium on integrated water resources Management, 26-28 September, Ruhr University, Bochum, GERMANY.

CHAPTER 1

INTRODUCTION

1.1 Problem background

Increased regulations, through such as the EU urban wastewater treatment directive (EEC, 1991), to protect the environment and water bodies have led to growing demands to reduce point source pollution impacts on the quality of receiving water ecosystem. Compounding the problems are increasing plant loadings due to the growth of urban areas which means that existing facilities are now operating close to the limit of their design capacity. In addition, the privatisation of water industry, in the UK for example, has led to increased pressures for efficient design and operation of wastewater treatment plants, and other cost saving initiatives.

Achieving the desired protection and/or enhancement of receiving water quality will be either improving the performance of existing wastewater treatment plants or the construction of new facilities, as illustrated in Figure 1.1. The second approach is costly, as the capital expenditure required for the construction of new wastewater treatment facilities is very high and the required land may not be available due to planning and environmental constraints. The time scale involved is also such that this option is not feasible in the short-to-medium term. Hence the first approach, which if properly done, can improve effluent water quality, reduce the need of chemicals and save energy and operational costs (Olsson et al., 2005; Vanrolleghm, 2001; 1998; Majalli et al, 2007).

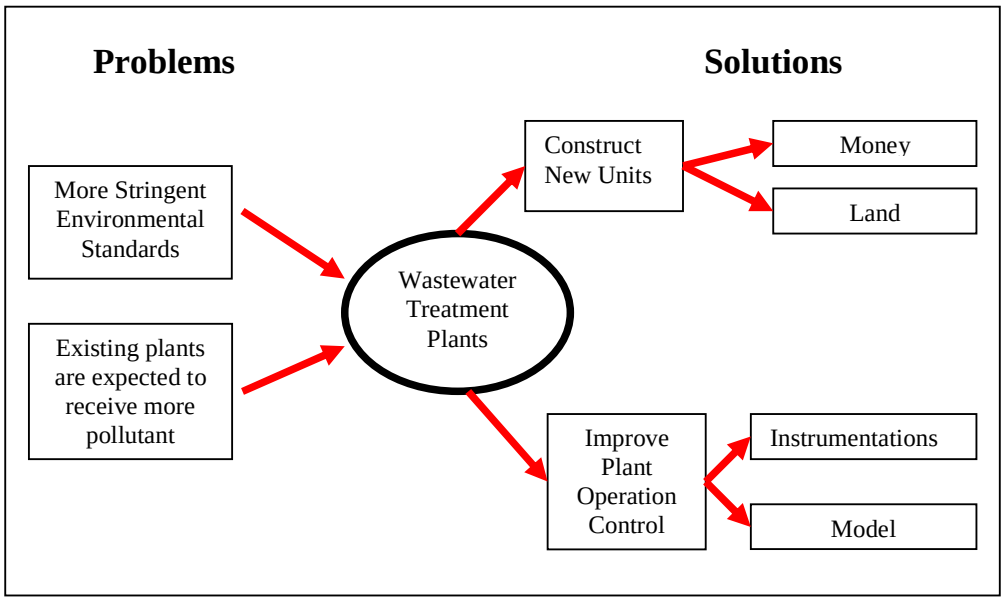


Figure 1.1 *Problems and solutions for wastewater treatment plant*

Therefore, sustainable solution to the problems of wastewater treatment will require the development of adequate information system for control and supervision of the process. However, because of variations in raw wastewater composition, as well as the changing and complex nature of the biological system of the activated sludge process, the operation and control of activated sludge wastewater treatment plants is quite complicated (Pu and Hung, 1995b). This reality has encouraged environmental engineers to use new modelling techniques to improve plant operation and control by designing operational control systems for qualitative and quantitative description of the dynamic behaviour of treatment plants. Such systems help the process engineer to convert unsatisfactory dynamic behaviour into satisfactory behaviour, thus reducing operational costs for meeting the requirement of regulatory agencies and minimising any adverse effects on the environment. Indeed, the mathematical modelling of the activated sludge process is a useful tool for the optimal control, mainly because the effects of adjusting the operating variables can be studied far more quickly on a computer than by doing experiments (Andrews, 1992; 1994). By using these models to simulate the effect

of possible correction actions, it is possible to rapidly respond to any change in the process, and devise an operational strategy, which can move the plant to new operating condition that improves its stability, the quality of the effluent and at the same time achieve reduction in the running costs.

Several efforts have been devoted to the modelling of the activated sludge process using mechanistic models as summarised by Lessard and Beck (1993) and Manfred et al. (2002); these are also discussed in more detail in Chapter 2. However, most of the models have been proposed to simulate the dynamic behaviour of the biological reactor and the secondary settler as if they were separable, independent units; very few models have looked at the interconnection between these two units (Dupont and Henze, 1992; Hamed et al, 2004). Moreover, the models were developed using data obtained under controlled laboratory conditions and are therefore more suitable for the design of treatment plants and may not be suitable for operational control (Nokyo ,2002). More importantly, very few of them have been validated with real field data (Cote, 1995; Han and Kamber, 2001; Majalli, 2007).

Thus, while models are vital for the effective control of wastewater treatment plants, the limitations of the currently available models necessitate more investigations in this field. One possibility, which has received increased attention recently, is the use of artificial intelligence (AI) modelling techniques. AI approaches are suitable for modelling the complex activated sludge process due to their learning ability to construct nonlinear relationships that can explain the complex relationships within the data without the difficult task of dealing with deterministic non-linear mathematics (Hamed et al, 2004). In addition, AI techniques can deal with complexity and uncertainty of the system in a manner similar to the human way of thinking and reasoning. Furthermore, AI models have the ability to generalize the input-output relationship to produce an output when

presented with previously unseen inputs. Although that might be possible with mechanistic models properly calibrated and validated, but mechanistic models require so much data that their effective calibration and validation are always difficult if not impossible. There is also the uncertainty associated with model identification for mechanistic models, i.e. The exact form of the functional relationship is unknown and so whatever mathematical expression is postulated for the unknown relationship, it bound to be a mere approximation. AI techniques on the other hand are data-driven techniques and there is no requirement to specify the mathematical form of the relationship being modelled. Among the commonly used AI tools and techniques are Fuzzy Logic system (FLS) and artificial neural networks (ANN), both of which have played an important role in the development of models for complex Environmental systems (Esteves, 2002; Cinar, 2005).

The fundamental and complementary characteristics of fuzzy logic and ANNs techniques have led researchers to combine them into an integrated system termed Fuzzy Neural Network (FNN). FNN combines the benefits of both NNs and FLSs by bringing together the learning and the computation powers of NNs and the high level human-like thinking and reasoning of FLSs. In addition, FNNs do have other characteristics that make them such a versatile tool in modelling applications namely, highly parallel structure implying a certain level of fault tolerance and the natural ease in dealing with multivariable systems (Jang, 1993).

This thesis does not go deep into many technical issues relating to the AI techniques as there are a plethora of excellent text books on the subject. Rather it merely utilized the most commonly used, well known, and easily understood of these techniques, namely Multi-Layered Perceptron artificial neural networks, Kohonen Self Organizing Maps (KSOM), and Fuzzy logic system to demonstrate

their feasibility for modelling complex environmental systems such as the activated sludge system (Jang, 1993).

1.2 Aim and Objectives

The aim of the work presented in this thesis is to provide a systematic and thorough approach to the development of artificial intelligence techniques in modelling and monitoring the activated sludge wastewater treatment plants and to show the potential of hybrid systems of these techniques to deal with the complexity and uncertainty in the process. Thus, the specific objectives of the study are:

1. To apply Kohonen Self Organising map (KSOM), unsupervised neural networks, to pre-process high-dimensional data of activated sludge process for the sole purpose of predicting the missing values and replacing identified outliers which are sample values that differ notably from the mean of the measurement series, taking into account the multivariate nature of the system.

2. To apply KSOM to extract the salient features of high-dimensional data, Activated sludge data, by removing the noise and redundant information in the available data and to visualise the correlation between wastewater treatment parameters in the resulting low dimensional data space.

3. To develop a software sensor for the rapid prediction of biochemical oxygen demand, based on finding its correlation with other water quality variables, in order to facilitate the use of this parameter for real time monitoring and control of the activated sludge process.

4. To develop integrated unsupervised-supervised Artificial Neural Network models that improve the performance of ANN models through the use of KSOM feature extracted data.

5. To develop integrated fuzzy-neural networks for modelling of the activated sludge process in order to absorb the advantages of both fuzzy logic and neural network in order to improve the prediction capability of the model.

1.3 Structure of the thesis

The thesis is divided into nine chapters. Going through the thesis, the reader will be confronted with such widely varying disciplines as civil engineering, software engineering, microbiology, numerical analysis, control engineering, mathematical modelling, statistics, environmental and chemical engineering, and probably some more.

Chapter 2, Activated Sludge Wastewater Treatment Systems, reviews the basics of activated sludge wastewater treatment plants. It presents the importance of wastewater treatment and its history and the structure of a conventional activated sludge wastewater treatment plant. The biochemistry and microbiology of the process are described. A review of activated sludge modelling techniques are presented. Finally the use of artificial intelligent techniques as a tool to model and control such processes is also reviewed.

Chapter 3, Artificial Intelligence Techniques (AI), covers the essential background for understanding the subsequent chapters is briefly introduced. The main context of the chapter is an overview of the basic of AI techniques used in this study, namely, backpropagation Artificial neural networks, Kohonen features

map (Kohonen self organising map) and fuzzy logic. The hybrid modelling systems are also presented and discussed.

In chapter 4, Method and materials, the sources of the data sets used to develop the models in this study and the associated treatment works are described. The chapter also descries the data treatment methods and presents how to use the MATLAB programming language to program the AI models developed in this study. Models performance evaluation criteria are also discussed in this chapter.

In essence therefore, the first four chapters have covered all that is needed to know about the theoretical basis of the methodology, the data and the various assumptions inherent in the subsequent analysis carried out. Consequently, the next chapters are presented as applications of the previously described tools to the case study data and treatment works.

Chapter 5 contains Application 1 and presents the results of applying Kohonen Self-Organising Map (KSOM), unsupervised Neural Networks, for predicting the missing values and for replacing outliers of the time series data. This task is the first step in modelling the activated sludge wastewater treatment plants using intelligent techniques such as Fuzzy Logic and Artificial Neural Networks.

Application 2 is in Chapter 6 and presents a completely novel methodology based on the use of the Kohonen self-organizing map (KSOM) models to predict biochemical oxygen demand (BOD_5) concentrations in wastewater. Extensive testing and validation of the model shows that the model is sufficiently general to predict the BOD_5 readily using variables, which can be measured within three hours or in real-time using on-line hardware sensors, thus making it possible to estimate BOD_5 very rapidly. This allows for a timely intervention and cost reduction during problem diagnosis.

The third Application is contained in Chapter 7, Modelling ASP using hybrid KSOM-ANN, which presents the use of a new methodology based on a hybrid supervised-unsupervised artificial neural network to improve the performance of the basic backpropagation neural network method in modelling the activated sludge wastewater treatment plant. The findings prove the ability of KSOM to improve the performance of modelling using basic back-propagation neural networks, by extracting salient features from available noisy data which is a common problem with the process data of wastewater treatment plants.

The final Application is in Chapter 8, Modelling ASP using hybrid KSOM-ANFIS, and demonstrates the use of ANFIS for modelling wastewater treatment plants. The ANFIS allows fuzzy rules to be extracted and the ANN enabled optimised fuzzy membership functions to be determined, a significant important over the traditional trial-and-error method of the developing such membership functions. The results indicate that the KSOM-ANFIS hybrid not only outperforms the basic ANFIS model in modelling capability with different number of inputs and different number of fuzzy membership functions, it is also unhindered by missing values or gaps in the data.

Chapter 9 documents the discussion, conclusions and recommendations for further research.

CHAPTER 2

ACTIVATED SLUDGE WASTEWATER TREATMENT SYSTEMS

2.1 Importance of wastewater treatment

Water is special. Every living thing on earth- microorganisms, plants, animals, human and even our brain consists mostly of water. In addition, water is used for numerous purposes, for example domestic consumption, industrial production, irrigation, transport of material, energy production as well as cleaning. However, although more than 70% of the earth's surface is covered by water, only 0.5% of this is suitable for all human uses (Gleick, 1996). This small fraction is diminishing as agriculture, industry, and domestic needs consume more and more of this small fraction, while the wastes generated constitute pollutants that further degrade the quality of the available water, thus rendering it unfit for purpose. That is because wastewater contains a considerable amount of organic materials, which, if discharged in large quantities to the receiving water bodies, would cause depletion of the dissolved oxygen levels and other environmental problems. This may lead to the environment becoming uninhabitable for higher life forms such as fish. In addition, toxic materials may be present due to industrial components (Metcalf & Eddy, 2003).

Therefore, in order to protect the environment and sustain life, wastewater must be adequately treated prior to being discharged. In particular, the biological treatment of wastewater helps to reduce the organic content of the wastewater, thus limiting its dissolved oxygen impacts in the receiving water body. Other benefits of biological wastewater treatment systems are (Metcalf and Eddy, 2003):

- Prevention of disease and nuisance conditions;
- Avoidance of contamination of potable water supplies;
- Maintenance of clean water for survival of fish, bathing, and recreation;
- General conservation of water, soil, and even air quality for future use;
- Reduction of ammonia concentration and hence, its toxicity to aquatic life forms, e.g. fish;
- Elimination of other problems caused by excessive nitrogen compounds, e.g. cancer, blue-baby syndrome, and increased chlorine demands during disinfection;

Although there are several methods of biological wastewater treatment, the most often used is the activated sludge biological treatment system (Spellman, 2003). This process is capable of removing soluble and particulate carbon, nitrogen and in some cases phosphorus from domestic sewage using chemical treatment. In the following sections, further details about the history of wastewater treatment in general and the activated sludge system in particular are given. The chapter concludes with the state of the art in modelling the process.

2.2 Brief history of wastewater treatment

Although the earliest sewers known in the world were the great underground drain of ancient Rome, wastewater treatment is a comparatively recent development dating from the late 1800s and early 1900 (Spellman, 2003). In England, wastewater treatment did not receive much attention until the construction of sewerage systems in the mid-1800 after the cholera plague, which claimed over 25,000 victims between 1848 and 1854 (Cooper, 2002). Because of the relatively small sizes of the British rivers, untreated wastewater discharged into them readily

became a nuisance. As the rivers became polluted and the amount of land suited for wastewater disposal by irrigation was limited, the development of intensive methods of wastewater treatment became imperative.

In the United States, wastewater treatment and disposal did not receive as much attention as in England because the extent of pollution caused by wastewater discharged into the relatively large bodies of water was little and because of the wide areas available for land treatment of wastewater (Spellman, 2003). The first septic tanks used in the US, according to Spellman (2003), were reported in 1876, and in 1887 the Lawrence Experiment station was established by the Massachusetts State Board of Health to study both water and wastewater treatment (Department of Army, 1975, reported by Spellman, 2003).

The first idea of recovery of water quality through treatment was based on physical means, such as dilution and sedimentation. Spellman (2003) reported that the first wastewater treatment plant in Germany (Frankfurt/Main based on grit removal, screens, settling tank) was put into operation as late as 1887 (see also Seeger, 1999). However, this became insufficient as cities became larger and the environmental standards become more stringent with time. This led to improvement in treating through the development of biological treatment systems such as trickling filters or biological beds, which date back to the late 19th century and further developed and improved in the early 20th century.

Another breakthrough in biological treatment of sewage was the discovery that supplemented aeration of wastewater resulted in better and faster purification. Thus, in the beginning of the 20th century, experiments were carried out on what was called the activated sludge process, which was discovered in 1913 by Ardern and Locket from laboratory experiments at Davyhulme treatment plant in Manchester, England (Cooper, 2002). It was named activated sludge because it

involves the production of an activated mass of the micro-organisms capable of stabilizing a waste aerobically. Many versions of the original process are in use today, but fundamentally, they are all similar. For example, activated sludge processes that have two or more aeration tanks may be designed to operate in a variety of modes of operation. The feed point of primary clarifier effluent or activated sludge influent and return activated sludge determine the mode of operation. Modes of operation include complete mix, plug-flow, step feed, and contact stabilisation as illustrated in Figure 2.1.

Later on, chemical additions were introduced to increase the settleability of the waste during clarification. Furthermore, in some sensitive areas, tertiary treatment such as nitrogen removal techniques and sand filtration are introduced in order to improve the overall performance of the treatment plants. During the last few decades, wastewater treatment has become an industry of high complexity.

2.3 The Structure of Wastewater Treatment Plant

In modern wastewater treatment plants, treatment is generally carried out in several steps before it is released to the receiving water body: physical, chemical and biological treatment, which is used in many different combinations (Metcalf and Eddy, 2003; Spellman, 2003). A schematic of a typical plant is presented in Figure 2.2. Each of the treatment stages is described below.

2.3.1 Preliminary treatment

In this step, mechanical treatment is applied. The purpose of preliminary treatment is to protect plant equipment from clogs or jams or excessive mechanical wear by removing large objects such as rags, cans, sanitary pads, condoms, branches, leaves, roots, and many kinds of materials commonly known as trash. In addition, it saves valuable space within the treatment plant.

Preliminary treatment operations include screening, shredding, grit removal, pre-aeration, chemical addition. More details can be found in (Spellman 2003; Metcalf and Eddy, 2003).

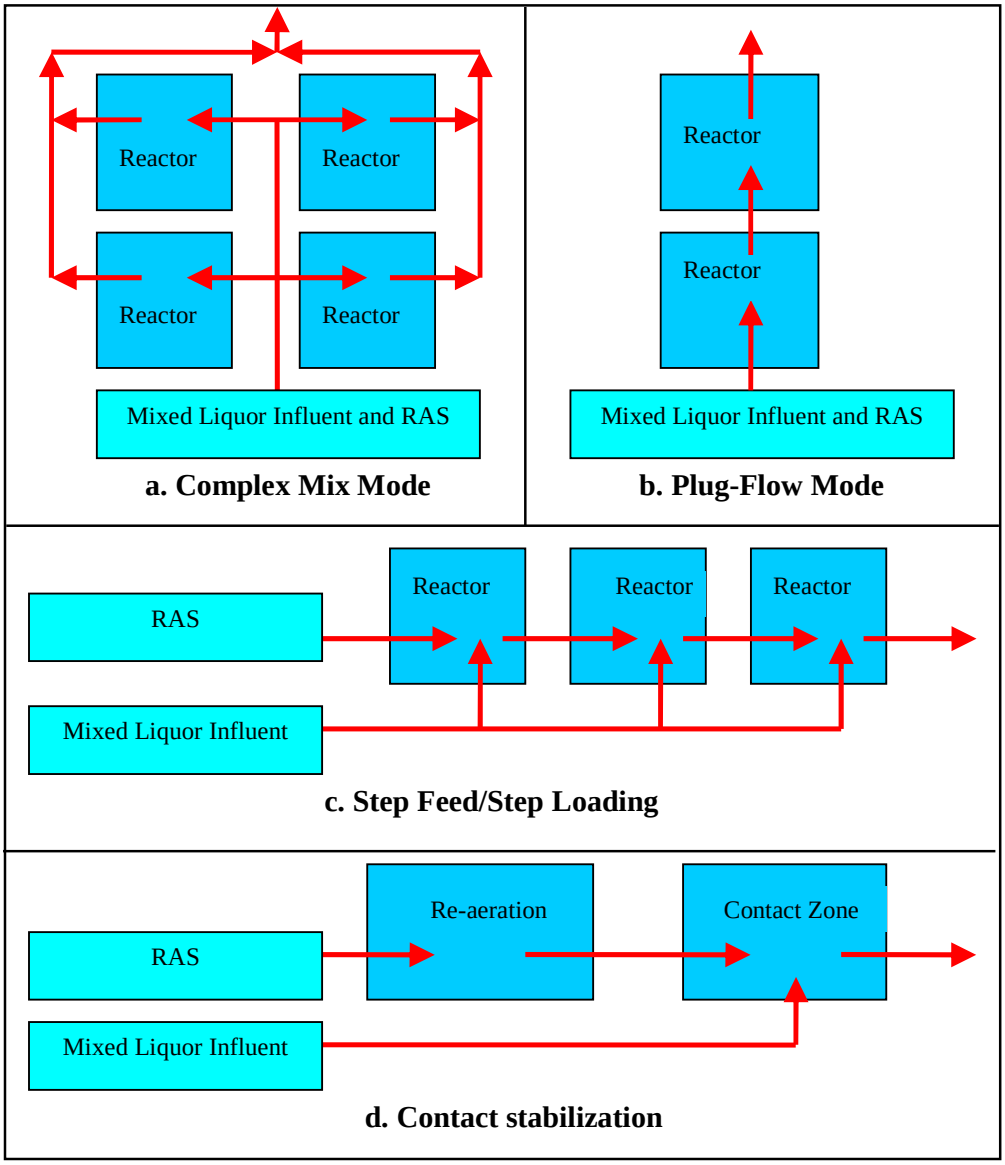


Figure 2.1 *Modes of operation commonly used in the activated sludge process (after Gerardi, 2002).*

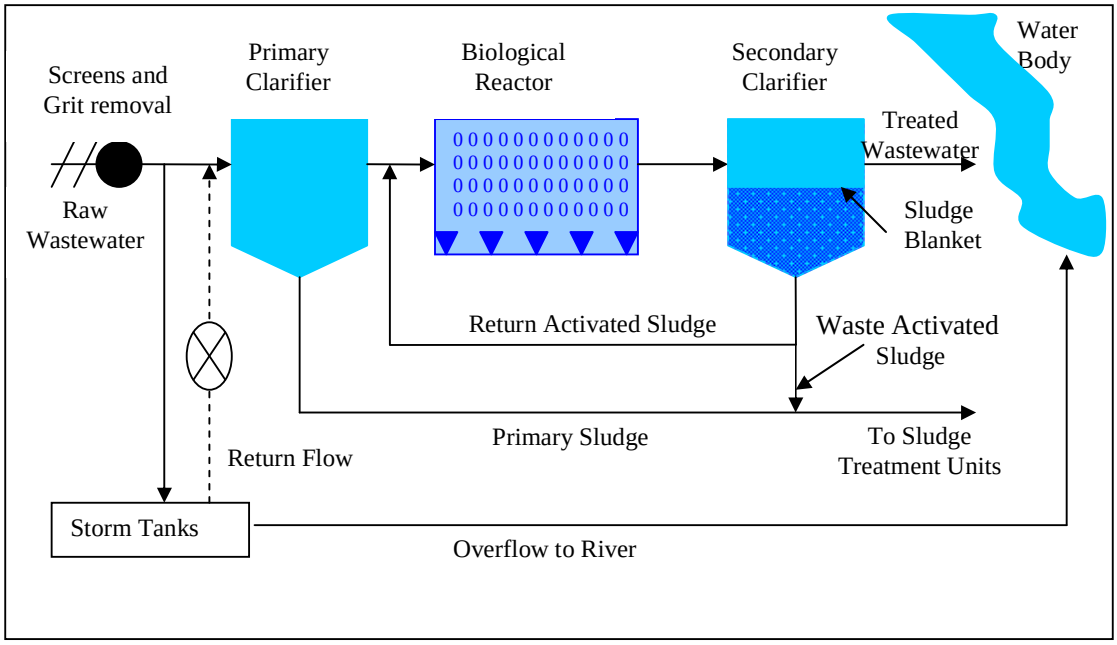


Figure 2.2 *Schematic of a wastewater treatment plant based on the conventional activated sludge biological treatment*

2.3.2 Storm Tank

Storm tanks serve the purpose of storing sewage flows caused by rainfall which are in excess of the capacity either of the treatment plant or of the sewer conveying flows to the treatment plant. Usually these tanks are located at the inlet of the plant. These tanks also allow the water stored to be partially treated by sedimentation. The importance of the storm tanks results from the fact that they can be used for control of the inflow to the treatment plant and for reduction of combined sewage overflow (CSO) discharges. Also, storm tanks storm the first foul flush.

2.3.3 Primary treatment

The purpose of primary treatment (primary sedimentation or primary clarification) is to remove settleable solids and floatable materials from wastewater. The efficiency of its operation influences directly the subsequent biological and sludge treatment units. Typically, 50 to 70 % of the total suspended solids are removed in this stage (Spellman, 2003). Since some of the solids are biodegradable, biochemical oxygen demand (BOD) is also removed, typically reduced by 30 to 40% (Metcalf and Eddy, 2003; Gray, 2004). Primary clarification uses large basins in which settling is achieved via gravity conditions. Solids that are heavier than water and have adequate settling velocity will settle to the bottom within the allowed detention time, while solids that are lighter than water, such as oil and grease, float to the top. Within these basins, mechanical scrapers collect the settled solids into a hopper where they are pumped to sludge handling and treatment facilities. The oil, grease, and other floating materials (scum) are skimmed from the surface. The effluent is discharged over weirs into a collection trough and goes on to the next step in the treatment process. Standard retention times are about 0.5-2 hours, while retention times greater than 3 hours do not significantly improve the efficiency of the primary clarifiers (Spellman, 2003). The primary clarifier also has an equalizing effect on variations in influent wastewater concentration. Peak concentrations are damped due to a time lag of the order of the retention time (Spellman 2003; Metcalf and Eddy, 2003).

2.3.4 Secondary treatment

The main purpose of the secondary treatment, sometimes referred to as biological treatment, is to provide biological oxygen demand (BOD) removal beyond what is achievable by primary treatment, and in some configurations, nutrients are also removed. Secondary treatment takes advantage of the ability of microorganisms to

convert dissolved, suspended and colloidal organic wastes (or BOD) into more stable solids that can either be removed by settling or discharged to the environment without causing harm.

Biological process is based on biological cultures. Gray (2004) reported that many different species have been observed to be present in the process namely, heterotrophic, autotrophic, yeasts, algae, fungi, filamentous bacteria, and protozoa. The organic pollutants in the wastewater serve as food and energy sources for the microbiological culture as it grows. The microbiological culture can either grow suspended in the water phase or in a fixed position on a surface of the media, as a biofilm (Metcalf and Eddy, 2003; Spellman, 2003).

Fixed film systems, or the trickling filter, are processes that use a biological growth that is attached to some form of media like trickling filter. Wastewater passes over or around the media where the organisms remove and oxidize the organic solids. The media may be stone or any other substance that is strong (capable of withstanding weather conditions for many years), and provide a large area for biomass growth and an open space for ventilation. On the other hand, suspended growth systems are processes that use a biological growth that is mixed with the wastewater. Typically, suspended growth systems consist of various modifications of the activated sludge process, and are usually more compact than the trickling filter for the same population equivalent served.

Due to the efficiency and compactness of activated sludge system for large wastewater treatment handling, this thesis only focussed on this type of treatment systems. Hence, more details about this system will be presented in the next section.

2.4 Activated Sludge Process

The activated sludge process is the most commonly used biological wastewater treatment system. It mainly consists of several biological reactors (aerated tanks), and solid-liquid separators (secondary clarifiers or settlers). It is capable of performing four critical wastewater treatment functions, namely: the degradation or oxidation of carbonaceous wastes; the degradation or oxidation of nitrogenous wastes; the removal of fine solids; and the removal of heavy metals. These functions are achieved primarily through the growth and maintenance of a large, diverse, and active population of bacteria. Hence, it transforms the biodegradable constituents (substrate) into new biomass, carbon dioxide, water, and residual organic matter using the dissolved oxygen supplied by the aerators. The clarifier functions are to separate the suspended solids and biomass from the aerated sewage and thicken the sludge before it is recycled to the reactor (Spellman 2003; Metcalf and Eddy, 2003).

As stated previously, activated sludge process is a biological process in which microorganisms oxidize and mineralize organic matter. Hence, the main requirement of the activated sludge process is to keep a high concentration of a mixed culture of microorganisms, known as the mixed liquor suspended solids (MLSS), in an artificially aerated reactor. The composition of the species of microorganisms depends not only on the influent wastewater but also on the operation of the wastewater treatment plant. The microorganisms grow slowly in the aerated tank and are kept suspended either by blowing air into the tank or by using agitators. Oxygen is used by the microorganisms to oxidize organic matter. On leaving the aeration tank (detention time typically 6 hours), the MLSS enters the secondary settling tank where it is clarified and thickened. To maintain the microbiological population in the aeration tank, part of the thickened sludge from the secondary clarifier is re-circulated back to the aeration tank; the surplus

thickened sludge is then wasted. The volume of sludge returned to the aeration basin is normally 40 to 60% of the wastewater flow. A basic schematic of the biological process in an activated sludge process is illustrated in Figure 2.3.

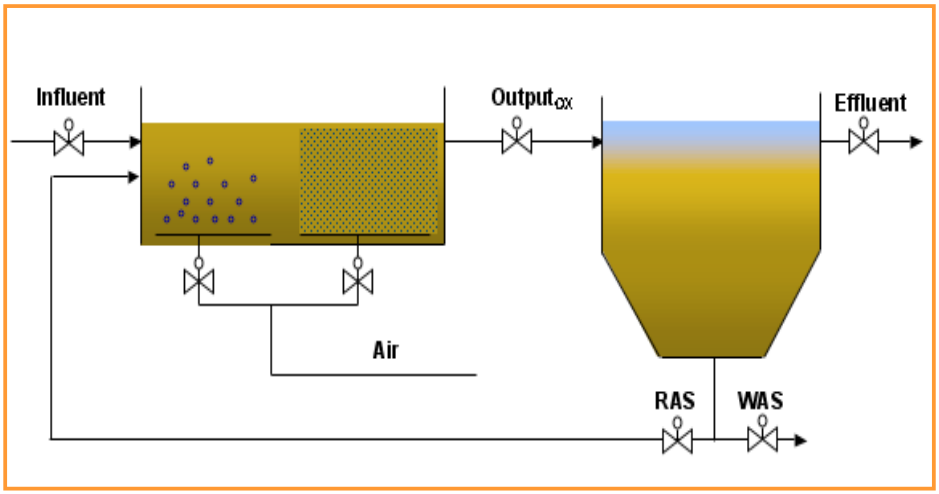


Figure 2.3 *basic schematic of the biological process in an activated sludge process.*

The biomass growth rate depends on many variables such as the amount of biomass, the substrate, temperature, pH, and the presence of toxins. The growth in number and diversity of bacteria occurs over time or increasing mean cell resident time (MCRT) or sludge age. During this time, the BOD is transformed into new less polluting wastes and more new bacterial cells or sludge. The bacteria along with ciliated protozoa and metazoan, remove fine solids and heavy metals from the bulk solution. An additional and critical role performed by the ciliated protozoan and metazoan is the consumption of the dispersed cells. The consumption of dispersed bacteria by these organisms is known as cropping action. By cropping bacteria the bacteria are removed from the waste stream (Spellman 2003; Metcalf and Eddy, 2003; Gerardi and Wiley, 2002). During bio-

reduction (decay of microorganisms), biologically inert (non-biodegradable) matters are produced. Incoming wastewater will contain some inert matter as well. This matter flows unaffected through the process and is collected and removed in the settler.

2.4.1 The Secondary Clarifier

The secondary clarifier (SC) is an integral part of the activated sludge system. It has two main functions: it separates the biomass from the water in order to produce a good quality effluent free from settleable solids and it also thickens the biomass. Part of the thickened biomass is then wasted as sludge and part of it is returned to the biological reactor to maintain an appropriate biomass concentration. The SC also removes floating foam and scum produced in the aeration tank (Gerardi and Wiley, 2002; Spellman 2003).

The operation of the secondary clarifier is crucial for the whole treatment plant (Gerardi and Wiley, 2002; Chen, 1993). As Beck (1984) puts it "it is in the secondary clarifier where adverse operational problems of bulking, rising, or dispersed sludge either develop or become critically apparent". The term "bulking sludge" refers to sludge that has poor settling characteristics and poor compactability. Causes of sludge bulking include the growth of filamentous organisms or bacterial cells swelling through the addition of water. "Rising sludge" is caused by the denitrification in the secondary clarifier. Denitrification may result in nitrogen gas becoming trapped in the sludge layer and causing the sludge to rise. Another operational problem present in the absence of filamentous organisms is "dispersed sludge" which thickens easily but gives an effluent with high concentration of fine suspended solids. Hence, the main goal in the operation of the secondary clarifier is to prevent excessive rise of the sludge blanket, which eventually may result in loss of sludge into effluent. This not only increases the

effluent concentration of solids and organic mater considerably, it also affects the performance of the activated sludge process itself, since biomass which is necessary in the aeration tank for proper functioning of the process is lost from the system (Chen, 1993; Gerardi and Wiley, 2002; Spellman 2003).

2.4.2 Operation of Activated sludge system

To obtain the desired level of performance in an activated sludge system, a proper balance must be maintained between the amount of food (Organic matter), organisms (activated sludge), and dissolved oxygen (DO). The majority of problems with the activated sludge process results from an imbalance between these three parameters (Spellman, 2003). The actual operation of an activated sludge systems is thus regulated by three factors: aeration and dissolved oxygen, the rate of activated sludge recirculation (RAS) (pumped from the secondary clarifier back to the aeration tank), and the amount of excess sludge withdrawn from the system (WAS) (usually pumped from the secondary clarifier towards sludge treatment).

Aeration and dissolved oxygen has two main objectives: keep the oxygen concentration within the appropriate limit, usually 2 mg/l, to maintain the microorganisms active, and ensure that the tank contents are sufficiently well mixed to keep the solids in suspension. Low dissolved oxygen concentration can limit the growth of microorganisms and encourage the predominance of filamentous bacteria with the subsequent deterioration of the effluent quality, as described previously in Section 2.4. On the other hand, high dissolved oxygen concentration represents a high energy waste through excess turbulence, especially with mechanical aerators, that may break up the biological floc resulting in poor settling characteristics and high concentration of solids in the effluent.

The return activated sludge (RAS) rate is a critical control variable as it redistributes the sludge between the secondary clarifier and the aeration tank, such that the healthy population of biomass is maintained in the aeration basin. Thus, the operator must maintain a continuous return of activated sludge to the aeration tank or the process will show drastic decrease in performance. If the RAS rate is too low, solids remain in the settling tank, resulting in solids loss and a septic return. If the rate is too high, the aeration tank can become hydraulically overloaded, causing reduced aeration time and poor performance. Therefore, there should be a balance between the return activated sludge and the wastage in order to achieve as desired performance.

On other hand, because the activated sludge contains living organisms that grow, and produce waste matter, the amount of activated sludge is continuously increasing. If the activated sludge is allowed to remain in the system for too long, the performance of the process will decrease. If too much activated sludge is removed from the system, the solids become very light and will not settle quickly enough to be removed in the secondary clarifier. Hence, WAS is an important operational parameter because it allows the operator to establish the desired concentration of MLSS, food to microorganisms ratio (F: M), and sludge age. Furthermore, the separation of solids and liquid in the secondary clarifier results in a blanket solid. If solids are not removed from the clarifier at the same rate they enter, the blanket will increase in depth. If this occurs, the solids may carry over into the process effluent. The sludge blanket depth may be affected by other conditions, such as temperature variation or sludge bulking, as explained in Section 2.4.1.

2.4.3 Modelling activated sludge process

Modelling the activated sludge process has an important role in implementing efficient control actions for better process performance. Models are helpful mainly because the effects of adjusting the operating variables can be studied far more quickly on a computer than by doing experiments as illustrated in Figure 2.3. Hence, many alternative designs and operational strategies can be compared without the need for physical trials of each scenario (Andrews, 1992, 1994; Olsson, 2005). By simulating these models with the possible correction actions, it is then possible to rapidly respond to any change in the process and devise an operational strategy, which can move the plant to new operating conditions that improve its stability, the quality of the effluent and at the same time achieve reduction in the running costs. Therefore, optimum process configurations, which meet given effluent quality standards at least cost, can be achieved (Olson, 2005; Rivas et al., 2008).

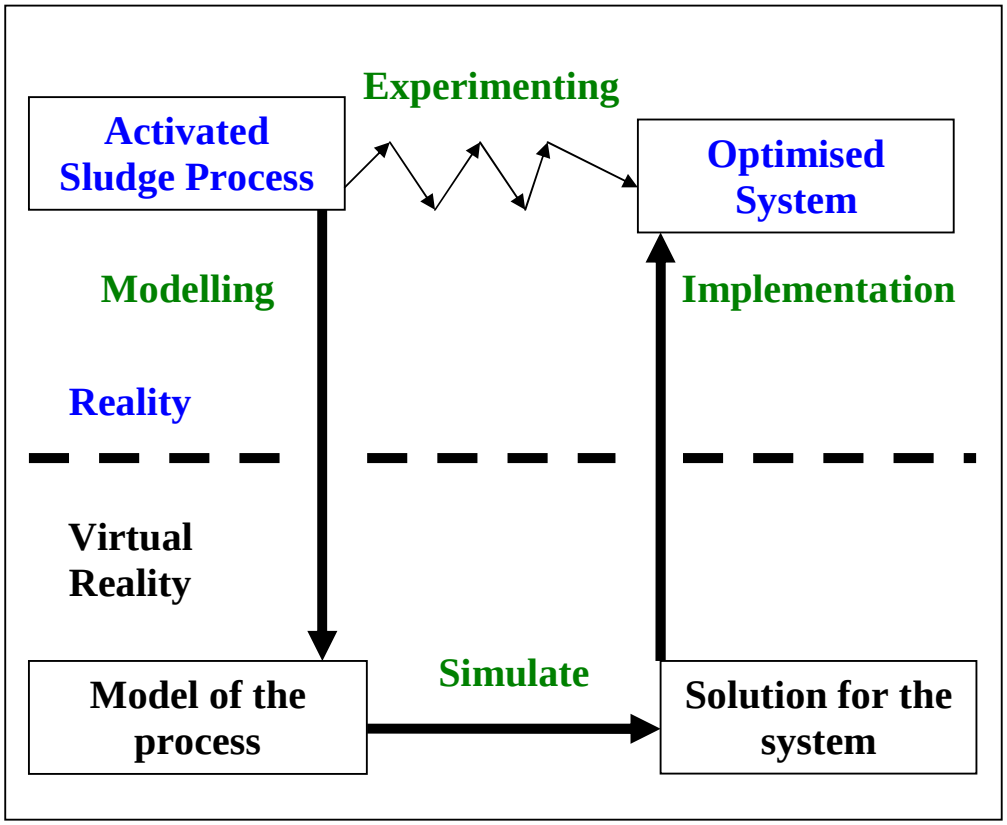


Figure 2.3 *Applying models to solve problems in activated sludge wastewater treatment plant. The Figure illustrates that instead of applying the correction action on a real system, it is better to test several correction scenarios using the model and then apply the results on the real system.*

However, modelling the wastewater treatment process is not without its problems, namely:

- The process is time varying. It consists of many sub-processes with strong dynamics of various scales. It has time constants, which range from seconds to months. Some variables are slow to change, for instance sludge dynamics (MLSS) and temperature, with time scales of days, weeks and even sometimes months. The daily variation in influent flow rate and substance concentrations is perhaps the most dominant variation in the process. However, there are even faster variations present, such as dissolved oxygen (DO).

- The process has a complex multivariable nature, with large amount of data collected by the measurement system. This high dimensionality makes it difficult to decide what should be considered as inputs and outputs. In addition, there is a significant complex interaction between variables. In other words, there is a complex cause-effect relationship caused by the biological cultures, recirculation and control actions. Furthermore, these variables typically exhibit complex nonlinear characteristics (Spellman, 2003; Fu and Poch, 1995).

- As a biological process, there is a lack of reliable on-line measurement instruments. For example, the process has variables, such as BOD, that are difficult to measure in real time. In addition, many sensors are not reliable because they are noisy, have long response time, require frequent maintenance and can drift (Schilling, 1994; Harremoës et al., 1993; Vanrolleghem et al., 1990; Steyer et al., 1999; Olsson, 2005). Consequently, most of the available data records have lots of missing and/or erroneous values (Rustum and Adeloye, 2007-a, b).

- Many factors that affect the process are not routinely monitored in most wastewater treatment plants. For example, factors such as the soluble inert matter form part of the mechanistic model's representation of the treatment process (ASM1 for example) but are rarely measured on full-scale plants (Henze et al., 1986).

- The process is subject to large unpredictable conditions such as the effect of toxic industrial materials, mechanical breakdowns (Manfred et al., 2002), some of which are quite difficult to formulate mathematically.

- No two wastewater treatment works are the same. Variations exist in the size and circumstances; the nature of the industrial waste inputs is site

specific. Climate differences are also to be expected. Therefore, there is a need to consider the specific features and circumstances of any wastewater treatment works that is to be simulated using mathematical models.

Despite all of the above, some advances have been reported in modelling wastewater treatment process, especially the activated sludge system.

2.4.4 Types of Modelling of the activated sludge process

Mathematical modelling of the activated sludge processes has received considerable attention in the last three decades. A variety of model structures has been proposed to represent the processes occurring within the activated sludge system such as carbon oxidation, nitrification and denitrification (Manfred et al., 2002).

Activated sludge models can be classified into two classes, first according to the part of the plant to be modelled, for example primary clarifier, aeration reactor and secondary clarifier, and second according to the modelling strategy to be employed, i.e. fundamental or empirical. Models derived from mechanistic equations are called fundamental models. In contrast, empirical models fit the data but do not reflect physical consideration in the system. Empirical models are identified from system input-output data where the model coefficients are fitted to input-output map using statistical methods. The following subsections present the efforts devoted to modelling the activated sludge wastewater treatment plant based on the different modelling strategies.

2.4.4.1 Fundamental models of the activated sludge wastewater treatment process

a. Primary clarifier models

Due to the importance of primary clarifiers (primary sedimentation tanks), not least because their performance affects the performance of subsequent units and

the sludge treatment, numerous efforts have been devoted to the development of primary clarifier models (Lessard and Beck, 1991). Primary clarification is often considered as being not very “sensitive”, resulting in the use of simplified models to represent its dynamic behaviour such as steady-state approach (Otterpohl and Freund, 1992). Most of the primary clarifier models do not consider any biological reactions to occur in the reactor, simulating only the suspended solids (SS) behaviour. However, in certain cases some biological phenomena take place in the primary settler as modelled by Lessard and Beck (1988).

There are several problems associated with precise modelling of primary clarifiers. These problems are caused by the complexity of the dynamic behaviour of the sedimentation process and include:

- Variability of influent characteristics;
- Variability of particle size and corresponding velocities;
- Presence of complex flow patterns and density currents in the tank;
- Scouring phenomena and the effects of temperature.

Although the above difficulties exist and are well documented, a variety of primary clarification models have been proposed ranging from simple steady-state models to lumped and distributed parameter models (Leassard and Beck, 1991). Simple models usually relate the removal efficiency to influent suspended solids concentration and overflow rate. Examples of primary clarifier models can be found in Alarie et al. (1980), Otterpohl and Freund (1992), Paraskevas et al. (2003). However, most of these models depend on parameters that are not measured in the real operation of the treatment works. For example, the model developed by Takacs et al. (1991) is frequently used to describe the dynamic behaviour of settlers. In this model, the clarifier is divided into a number of layers

(usually ten) and a mass balance is made over each layer to evaluate the SS profile in the settler. The particularity of the model is the use of a settling velocity model describing both clarification and thickening as described in Equation 2.2.

$$v_{sj} = v_0 e^{-r_k X_j} - v_0 e^{-r_p X_j} \tag{2.1}$$

where:

v_{sj} is a settling velocity of the solids in the layer j (m/d);

v_0 is a maximum settling velocity (m/d);

r_k is hindered settling parameter (m^3/g);

r_p is flocculent settling parameter (m^3/g);

X_j is $X_i - X_{min}$ (X_i = SS concentration in layer j (g/m^3);

$X_{min} = fns X_{in}$;

fns = non-settleable fraction of X_{in} and

X_{in} =influent SS concentration (g/m^3).

Therefore, according the characteristics of the primary clarifier influent and the design characteristics, its performance can be calculated.

b. Biological reactor models

Deterministic mathematical models rely on differential equations and kinetic parameters and coefficients to describe the process. These models take into account changes of flow rate, composition and concentration of the influent wastewater. Manfred et al.(2002) provide excellent reviews of the progress in modelling wastewater treatment plants using fundamental model from where it is

clear that the most widely known deterministic model is the Activated Sludge Model Number 1 (ASM1) (Henze et al., 1987).

In 1986, the International Association of Water Pollution Research and Control (IAWPRC), later renamed International Association of Water Quality (IAWQ), and then International Water Association (IWA), formed a task group on "Mathematical Modelling for Design and Operation of Biological Wastewater Treatment" in order to promote the development of practical models for aiding the design of biological wastewater treatment plants. This group contained one representative from each of five countries (Denmark, USA, Switzerland, Republic of South Africa, and Japan). Each representative had experience in activated sludge modelling. Their assignment was to review existing models and to develop from the literature a consensus model able to realistically predict the performance of activated sludge process that perform carbon oxidation, nitrification and denitrification. This led to Activated Sludge Model No. 1 (ASM1) (Henze et al., 1987).

ASM1 represents the state of the art in modelling the activated sludge process and has been proven as a successful model in many applications (Béline et al., 2007; Chen and Ribarova, 1999; Maryns and Bauwens, 1997). Consequently, it has been introduced in a number of computer programs; e.g. Single Sludge Simulation Program (SSSP) developed by Bistrup and Grady (1988), ASIM (Gujer and Henze, 1991) and GPS-X (Patry and Takacs, 1990).

ASM1 contains 18 parameters, made up of 5 stoichiometric and 13 kinetic coefficients. Table 2.1 lists the ASM1, whilst Tables 2.2 defines the state variables. Table 2.3 and 2.4 list the kinetic and stoichiometric coefficients, respectively. ASM1 models eight processes involving 13 state variables, including various fractions of organic matter, biomass, nitrogen components, particulates

and alkalinity. Model representation is usually done by listing the processes as rows of a matrix with the state variables as column headings. Each matrix entry indicates the appropriate stoichiometric coefficient for the relationship between the variables in the individual processes (see Table 2.1). Process equations for each state variable are easily read by summing up the products of each entry of the related column with the kinetic coefficients of the process, which are given in the right-most column of the matrix. For example,

$$X_S = (1 - f_p)(b_M X_{BH}) + (1 - f_p)(B_M X_{BA}) - 1\times$$

$$(K_h \frac{X_S / X_{BH}}{K_X + K_S / X_{BH}} (\left(\frac{S_O}{K_{OH} + S_O} \right) + \eta_h \left(\frac{K_{OH}}{K_{OH} + S_O} \right) \left(\frac{S_{NO}}{K_{NO} + S_{NO}} \right)) X_{BH} \ ) \qquad (2.2)$$

Some useful hints for determination of some of the model's parameters for a given treatment plant are given by Nowak et al. (1999).

As remarked previously, the ASM1 has seen several practical applications worldwide. For example, Chen and Ribarova (1999) used the ASM1 to investigate the feasibility of upgrading Parada, a conventional WWTP in Portugal, for biological nitrogen removal. Very good correlations between measured data and simulation results were achieved using the parameter values proposed by the authors of ASM1. Computer simulations were done to study whether the capacity of the treatment plant is enough for biological nitrogen removal. They found that very high effluent quality can be achieved if the first tank volume is used for denitrification, the second and third tank volumes are used for BOD removal and nitrification, the recirculation flow ratio is 1.5 and the sludge age is 15 days.

Further development of modelling the activated sludge wastewater treatment plants resulted in the ASM2, ASM2D, and ASM3 models. The ASM2 provided detailed biological kinetics and reflected the state of art on the understanding of

nitrification, denitrification, and biological phosphorus removal (Henze et al., 1995). The ASM2D, an extension of ASM1 and ASM2, is a model for biological phosphorous removal with simultaneous nitrification-denitrification in activated sludge systems (Gujer et al., 1999). The ASM3 can predict oxygen consumption, sludge production, nitrification and denitrification for activated sludge system (Gujer et al., 1999). The ASM3, (Henze et al., 2000), also includes storage of organic substrates as a new process and the decay process was exchanged for an endogenous respiration process. Additionally, ASM3 is provided in a form that can be implemented in a PC code without further adjustments but it does not include biological phosphorous removal as it is contained in the ASM2.

However, the complexity of these models and the detailed microorganism's growth and decay data required by them mean that they are not appealing to process designers and operators as reported by Lessard and Beck (1991), and Weijers and Vanrolleghem (1997). For example, ASM1 requires the determination of about 31 parameters, coefficients and variables, most of which are not routinely measured by large wastewater treatment plants. Furthermore, a certain number of simplifications and assumptions must be made in order to make a model of a WWT system practically useful. Some of these are associated with the physical system itself, while others concern the mathematical model (Jeppsson, 1996).

Table 2.1 *Activated Sludge Model Number 1 ASM1 (Manfred et al, 2002)*

Component →		*i*	1	2	3	4	5	6	7	8	9	10	11	12	13	Process Rate, ρ_j [ML^{-3}T^{-1}]
j	Process	↓	S_I	S_S	X_I	X_S	$X_{B,H}$	$X_{B,A}$	X_P	S_O	S_{NO}	S_{NH}	S_{ND}	X_{ND}	S_{ALK}	
1	Aerobic growth of heterotrophs			$-\frac{1}{Y_H}$			1			$-\frac{1-Y_H}{Y_H}$		$-i_{XB}$			$-\frac{i_{XB}}{14}$	$\hat{\mu}_H\left(\frac{S_S}{K_S+S_S}\right)\left(\frac{S_O}{K_{O,H}+S_O}\right)X_{B,H}$
2	Anoxic growth of heterotrophs			$-\frac{1}{Y_H}$			1				$-\frac{1-Y_H}{2.86Y_H}$	$-i_{XB}$			$\frac{1-Y_H}{14\cdot 2.86Y_H}-\frac{i_{XB}}{14}$	$\hat{\mu}_H\left(\frac{S_S}{K_S+S_S}\right)\left(\frac{K_{O,H}}{K_{O,H}+S_O}\right)\left(\frac{S_{NO}}{K_{NO}+S_{NO}}\right)\eta_g X_{B,H}$
3	Aerobic growth of autotrophs							1		$-\frac{4.57}{Y_A}+1$	$\frac{1}{Y_A}$	$-i_{XB}-\frac{1}{Y_A}$			$-\frac{i_{XB}}{14}-\frac{1}{7Y_A}$	$\hat{\mu}_A\left(\frac{S_{NH}}{K_{NH}+S_{NH}}\right)\left(\frac{S_O}{K_{O,A}+S_O}\right)X_{B,A}$
4	'Decay' of heterotrophs					$1-f_P$	-1		f_P					$i_{XB}-f_P i_{XP}$		$b_H X_{B,H}$
5	'Decay' of autotrophs					$1-f_P$		-1	f_P					$i_{XB}-f_P i_{XP}$		$b_A X_{B,A}$
6	Ammonification of soluble organic nitrogen											1	-1		$\frac{1}{14}$	$k_a S_{ND} X_{B,H}$
7	'Hydrolysis' of entrapped organics		1			-1										$k_h\frac{X_S/X_{B,H}}{K_X+(X_S/X_{B,H})}\left[\left(\frac{S_O}{K_{O,H}+S_O}\right)+\eta_h\left(\frac{K_{O,H}}{K_{O,H}+S_O}\right)\left(\frac{S_{NO}}{K_{NO}+S_{NO}}\right)\right]X_{B,H}$
8	'Hydrolysis' of entrapped organic nitrogen												1	-1		$\rho_7(X_{ND}/X_S)$

Table 2.2 *Definition of state variables in the ASM1*

Component Number	Component Symbol	Definition
1	S_I	Soluble inert organic matter $M(COD)L^{-3}$
2	S_S	Readily Biodegradable Matter $M(COD)L^{-3}$
3	X_I	Particulate inert organic matter $M(COD)L^{-3}$
4	X_S	Slowly biodegradable substrate $M(COD)L^{-3}$
5	X_{BH}	Active heterotrophic biomass $M(COD)L^{-3}$
6	X_{BA}	Active autotrophic biomass $M(COD)L^{-3}$
7	X_P	Products from biomass decay $M(COD)L^{-3}$
8	S_O	Dissolved Oxygen $M(-COD)L^{-3}$
9	S_{NO}	Nitrate and nitrite nitrogen $M(N)L^{-3}$
10	S_{NH}	Ammonia nitrogen $M(N)L^{-3}$
11	S_{ND}	Soluble biodegradable organic nitrogen $M(N)L^{-3}$
12	X_{ND}	Particulate biodegradable organic nitrogen $M(N)L^{-3}$
13	S_{ALK}	Alkalinity- Molar Unit

Table 2.3 *Kinetic coefficients of the ASM1*

Kinetic event	Symbols	Unit
Heterotrophic max. specific growth rate	$\hat{\mu}_H$	Day $^{-1}$
Heterotrophic decay rate	b_H	Day $^{-1}$
Half-saturated coefficient (hsc) for heterotrophs	K_S	g COD m^{-3}
Oxygen hsc for heterotrophs	K_{NO}	g NO_3-N m^{-3}
Autotrophic max. specific growth rate	$\hat{\mu}_A$	Day $^{-1}$
Autotrophic decay rate	b_A	Day $^{-1}$
Oxygen hsc for autotrophs	$K_{O,A}$	g O_2 m^{-3}
Ammonia hsc for autotrophs	K_{NH}	g NH_3-N m_{-3}
Correction factor for anoxic growth of heterotrophs	μ_g	dimensionless
Ammonification rate	k*a*	m^3 (g COD day)$^{-1}$
Max specific hydrolysis rate	k_h	g slowly biodeg. COD (g cell COD day)$^{-1}$
Hsc for hydrolysis of slowly biodeg. substrate	K_x	g slowly biodeg. COD (g cell COD)$^{-1}$
Correction factor for anoxic hydrolysis	μ_h	dimensionless

Table 2.4 *Stoichiometric coefficient of the ASM1*

Stoichiometric coefficient	Symbol	Unit
Heterotrophic yield	Y_H	G cell COD formed (g COD oxidized)$^{-1}$
Autotrophic yield	Y_A	G cell COD formed (g N oxidized)$^{-1}$
Fraction of biomass yielding decay products	f_p	Dimensionless
Mass N/Mass COD in biomass	i_{XB}	G N (g COD)$^{-1}$ in biomass
Mass N/Mass COD in decay products	i_{XP}	G N (g COD)$^{-1}$ in endogenous mass

c. Secondary clarifier model

Generally, the final settling tanks are represented as two different processes: clarification and thickening. However, it is difficult to determine what the state of the art for dynamic models for the clarification and thickening is. That might be because relationships among operational variables of the reactor, the components of the sludge floc and its settling velocity are too complex to be quantitatively described (Chen, 1993). According to Lessard et al. (1993), final clarifier models are mostly based on empirical relationships and are largely related to the particular plant used to determine the parameters, and most models have not been evaluated thoroughly against experimental work at plant scale, especially for the thickener models.

Existing secondary clarifier models range in complexity from very simple empirical models for clarification of the settler (Plósz et al., 2007) to some very complicated two and three dimensional models which considered hydro dynamic effects (Giokas et al., 2002). However, the most famous model for secondary clarifier is based on the work by Lessard and Beck (1993). Its clarification part used an empirical approach derived from data from the treatment plant of Celle/ Germany. This model consists of four layers as shown in Figure 2.4:

1. Clarification zone of fixed volume where the water is assumed clean.

2. Dead Zone which occupies the volume left by thickening zone where no reaction occurs.

3. Thickening zone of variable volume where the water is separated from the sludge.

4. Compression zone of fixed volume where the sludge has already settled.

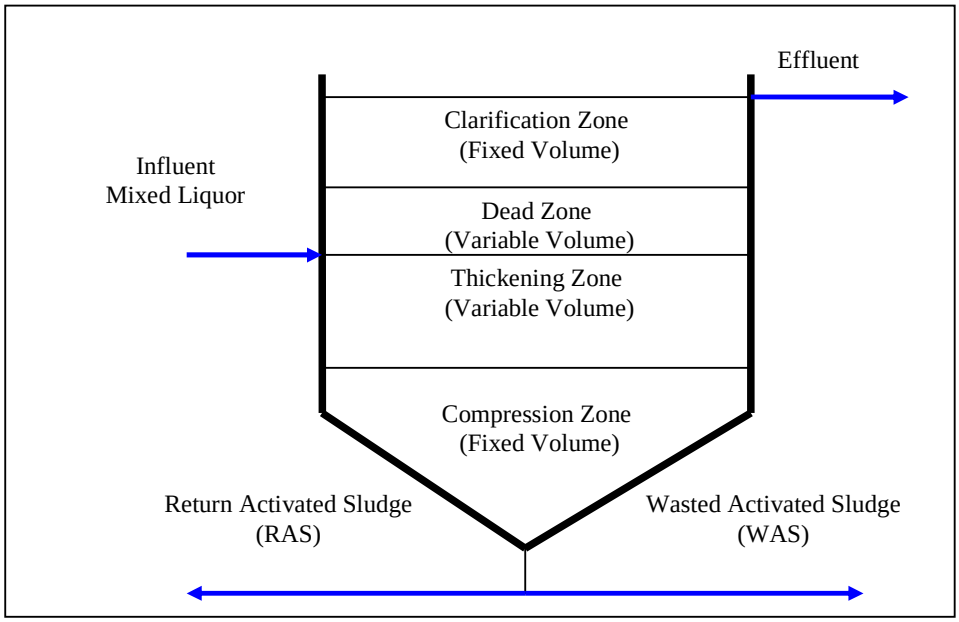


Figure 2.4 *Schematic representation of the zones in the final settling tank model (Lessard and Beck, 1993)*

In this model, the effluent suspended solids concentration is based on an empirical relationship determined by:

$$SS_{eff} = a_1 + a_2 \times (Q + Q_{RAS}) \quad (2.3)$$

Where

SS_{eff} : effluent suspended solids concentration (g/m^3)

a_1: minimum effluent suspended solids concentration in the secondary clarifier effluent (g/m^3), the minimum that can be achieved.

a_2: proportionality constant for the effluent of flow on effluent SS

Q: Influent Flow (m^3/s)

Q_{RAS}: return activated sludge rate (m^3/s)

The thickening part of the model is based on the flux theory as presented by Dick and Young (1972) in which the relationship between sludge settling velocity and concentration is represented by:

$$v = V_0 X_0^{-n} \quad (2.4)$$

Where

v: settling velocity (m/h)

V_0 : the start velocity (m/h)

X_0: the MLSS concentration (g/m^3)

n: is constant describe the sludge characteristics (l/g)

However, the limitations of the Lessard and Beck (1993) and most of the other secondary clarifier modes can be summarised as:

- These models can not describe dynamic process within the sludge blanket and there is hardly any potential for the prediction of the effluent quality (Krebs, 1995).
- Most of these models are developed based on pilot-scale data sets and they were not evaluated against field data.
- Most of these models are based on few number of layers to represent the thickening process. However, according to Jeppsen and Diehl (1996), at least 30 layers should be used in order to obtain reliable results under normal operating conditions.

- According to Dupont and Dahl (1995), most of these models suffer from at least one of the following two problems when applied to full-scale wastewater treatment plants:
    - Incorrect calculation of sludge concentration profile near the effluent weirs;
    - Incorrect calculation of the return sludge concentration.
- These models are dependent on specific site parameters, and would accordingly require extensive calibration efforts when applied to another site (Lessard and Beck, 1993).

d. Limitations of fundamental models

Although fundamental (deterministic) models are preferred because the models have the ability to predict beyond the range of existing operating data, their success is limited due to several factors; these factors can be summarised as:

- Fundamental models are often developed using data obtained under controlled laboratory conditions and are ,therefore, more suitable for the design of treatment plants and may not be suitable for operational control (Nokyoo, 2002; Leassard and Beck, 1991).
- Most fundamental models have been proposed to simulate the dynamic behaviour of the biological reactor and the secondary settler as independent units; very few models have looked at the interconnection between these two units (Dupont and Henze, 1992).

- Mechanistic models require recalibration, and they may be hard to reconfigure if the physical system is modified. That is because they are developed based on a specific physical system.

- Mechanistic models require high number of parameters and coefficients that limits the accuracy of the model. Besides many of these variables and parameters are typically not parts of the routine monitoring of plant performance. For example, soluble inert, and particulate inert, form part of the ASM1 model's representation of the treatment process but are rarely measured on full scale plants.

- There are significant costs associated with the collection of data to support a modelling exercise using these deterministic models, as well as the cost of the time devoted to the development of these models (Stokes, 1998).

Consequently, empirical models have been seeing increasing applications in the modelling of activated sludge wastewater treatment plants. In the next section, examples of empirical models available are presented with emphasis on those based on artificial intelligence paradigms.

2.4.4.2 Empirical models of wastewater treatment plants

The complexity of wastewater treatment plant necessitates more sophisticated control systems capable of delivering better and more flexible performance. Increasingly, control systems are required to have high dynamical performance and robust behaviour, and yet be able to cope with complex, uncertain and highly nonlinear process relationship over a wide operating environment. This requires the use of models of the system as argued previously.

However, due to the factors that limit the practicality of mechanistic models in modelling the wastewater treatment plant presented earlier, researchers are motivated to use new techniques to deal with the complexity and uncertainty in wastewater treatment plant operations. One of these techniques is the intelligent models that deal with complexity and uncertainty of the system in a manner similar to the human way of thinking and reasoning, without the difficult task of dealing with deterministic non-linear mathematics (Shankir, 2000). Among these tools and techniques are stochastic models, expert systems, Fuzzy Logic, data mining, and neural networks.

Stochastic modelling (or time series analysis) is the methodology that deals with the study of a set of observations generated sequentially in time. It includes combinations of the Box-Jenkins models such as autoregressive (AR), autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) among others (Box and Jenkins, 1976). Several applications of stochastic models to forecasting of treatment process time series have been reported in the literature. For example, uni-variate and multivariate process models were applied by Capodaglio (1994) to make one day ahead predictions of the water flow and SS based on measurements of rainfall with good results. Stochastic models have also been incorporated into a prototype Real Time Control (RTC) system for the control of an activated sludge process in Denmark by Kristensen et al. (2004). The data originated from Fnjoskadalur wastewater treatment plant. The goal was to predict the flow and the predictions were to be used for on-line automatic control in the wastewater treatment plant. The input data were precipitation, measured at the wastewater treatment plant, and the output data were flow data from the last pumping station before the treatment plant.

Expert systems (ES) models are implemented using expert knowledge and database (Capodaglio, 1994). Since the 1980s, several demonstration and research projects using ESs for control of WWTPs have been used (Andrew, 1992; Paraskevas et al, 2003). However, these models may not be generally applicable to every system as it is generally difficult to collect the expert's knowledge (Esteves, 2002).

Fuzzy logic models are a compromise between the vague statements which humans often use and the strict logic of expert systems. No complex mathematical relationships are required in the construction of fuzzy logic applications. Beside, it is believed to be conceptually easy to understand, flexible and tolerant of imprecise data allowing the modelling of complex non-linear functions. However, the drawback of fuzzy models is that tuning the parameters of the fuzzy membership functions is difficult and time consuming. Furthermore, the main difficulty is to define the number of necessary fuzzy rules.

Because Fuzzy logic models can handle highly non-linear, imprecise and uncertain systems that are poorly understood mathematically and depend not only on black box concepts, such as ANN, but also use a combination of knowledge of the system and operational experience, they have been suggested and applied to model wastewater treatment plants. Most of these applications are to simulate or control pilot plants (Tsai et al., 1993;1994; 1996; Muller et al. 1998; Ferrer et al., 1998; Manesis et al. 1998; Steyer et al.,1999; Carrasco et al.,002). Only few studies have used fuzzy modelling of the whole process including biological reactors and secondary settler (Tong et al.,1980; Geselbracht et al., 1988; Czogala and Rawilk, 1989; Yi et al., 1990; Fu and Poch, 1998; Watanabe et al., 1993; Fu and Poch, 1995a,b; Marsili-Libelli, 1996; Cohen et al., 1997; Huang and Wang, 1999; Kalker et al.,1999 ; Tomiello et al., 1999; Meyer et al., 2003; Sanchez et al., 2001; 2003; Traore et al., 2006). Some applications have been applied to

anaerobic digester as well (Boscolo et al., 1993; Estaben et al., 1997; Giraldo-Gomez and Dugue, 1998; Steyer et al., 1999; Polit et al., 2001; Punal et al., 2002; Murnleitner et al., 2002).

Artificial neural networks (ANNs) are non-linear mathematical structure that are capable of representing the arbitrary, complex non-linear functional relationship between the input and outputs of any system. ANN models have been used successfully to model complex non-linear input-output time series relationships in a wide variety of fields, including water resources, as for example, in predicting reservoir storage-yield-rehability relationship (Adeloye and Demnnari, 2006). Additionally, neural networks model was used for predicting the monthly values of water quality parameters in rivers (Diamantopoulou et.al, 2005). The success with which ANNs have been used to model dynamic systems suggests that the approach may prove to be an effective and efficient way to simulate complex wastewater treatment systems. Hence, the literature exhibits a wide range of applications of ANNs applied to wastewater treatment systems for example Mjalli et al. (2007), Hamed et al. (2004), Du et al. (1999), Raduly et al. (2007), Chen et al. (2003), Gamal-El-Din and Smith (2002). Further details about these data driven models will be given in the next chapter. In the mean time, some examples of their applications to wastewater treatment plants modelling and their limitations will be discussed.

Mjalli et al. (2007) used the ANN to model the Doha West wastewater treatment plant, Qatar. They used crude sewage quality characteristics inputs namely BOD, COD and TSS to predict the effluent stream BOD, COD and TSS. The authors used data over one year that were sampled every 5 days. They used smoothed data series instead of raw data to develop the model. The smoothing technique is used to reduce short-term volatility in the data by consolidating the available data points into longer units of time, namely an average of four historical data points.

For example, the data were performed every 5 days then every 4 points were smoothed, then every 20 days were represented by just one value which is not reasonable for wastewater treatment plant data. Thus, even though the performance of the model was relatively good, the peaks of the data were damped through the smoothing process. In addition, the authors did not mention the quality of the data and they just ignored the missing values in the measurement system. Furthermore, they used BOD as input to their model, although BOD takes 5 days to be measured in the bioassay method. Moreover, the authors used single input single output with just 40 neurons in the hidden layer, but there was no sensitivity test of the impact of changing the number of hidden neurons. They got over 70% correlation coefficient, but they did not distinguish if these results were from the training, validation, or testing data set.

Hamed et al. (2004) apply ANN to predict the performance of WWTP in a major conventional treatment plant in Great Cairo district, Egypt. Daily records of BOD and SS concentrations through various stages of the treatment process over 10 months were used. The ANN model was found to provide an efficient and a robust tool in predicting WWTP performance in terms of BOD and SS. The data records used in this study, BOD and SS, contain many missing values, for example, from 300 days, 247 were missing; they were merely ignored by the authors. The results obtained from this study indicated that R^2 values ranged from 0.63 to 0.81 for BOD and from 0.45 to 0.65 for SS. However, much more could have been achieved by their modelling if the missing values were in-filled. Moreover, Hamed et al. (2004) concluded that the model was hindered by the limitation of the data, the noisiness of the data, and the restriction of just two parameters, BOD and SS. There was also no mention of any independent validation of the model.

Du et al. (1999) develop a fuzzy- neural network hybrid model to predict the sludge age of the activated sludge process. It was found that fuzzy-neural network model is able to extract fuzzy rules from a set of numerical data that can be used to carryout heuristic reasoning. The model has three input variables; feed flow rate of activated sludge process (m^3/day); feed substrate concentration of activated sludge process (g COD/m^3) and sludge recycle rate (m^3/day). The model has one output variable, the sludge age. The data record were obtained from rigorous ASM1 simulation investigation because real data were unavailable. They assumed that the feed stream of the activated sludge process under study is described in terms of the feed flow rate and the feed substrate concentration and they ignored other parameters or assumed them to be time invariant. The recycle rate is assumed to be the only manipulated variable.

Choi and Park (2001) apply principal component analysis (PCA) to extract features of industrial wastewater treatment plant data. The extracted features are then used to predict the total Kjeldahl nitrogen (TKN) of influent industrial wastewater using the ANN as illustrated in Figure 2.5. The hybrid system shows an enhancement of prediction and reduces the over-fitting problem of ANN. Eleven industrial wastewater quality parameters are reduced to just five principal components, PR1-PR5 shown in Figure 2.5. These principal components became inputs to the ANN model. It is found that the PCA enhanced the performance of ANN. However, even if the PCA reduces the dimensionality of the data, the first two principal components could not extract more than 38% of the variance of the data due to the high nonlinearity of the system. Moreover, the developed model cannot deal with the presence of missing values in the measurement system.

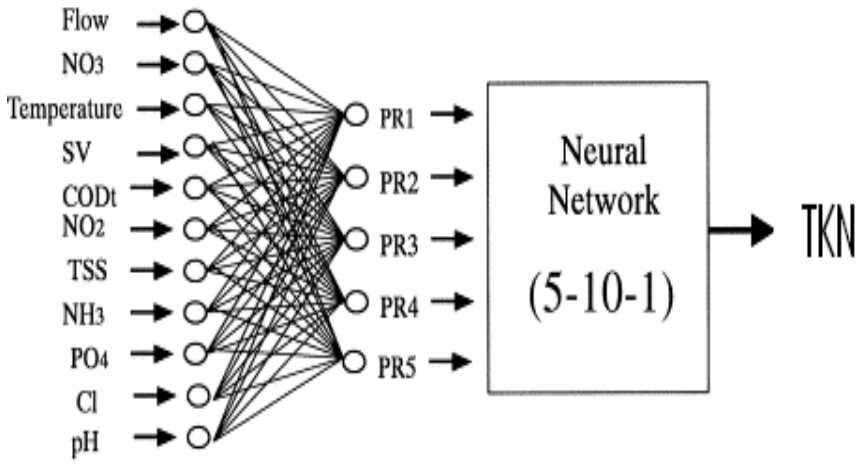


Figure 2.5 *The structure of hybrid neural network combined with principal component analysis (after Choi and Park, 2001)*

Raduly et al. (2007) apply ANN for rapid WWTP performance evaluation. However, the data used for training the ANN were generated using ASM3. In addition the input variables for the ANN were those not routinely measured at WWTP, for example, soluble inert material, readily biodegradable substrate, ammonium nitrogen, particulate inert material, and heterotrophy and autotrophy bacteria concentrations. The outputs of the model were the BOD, total COD (COD_{tot}), TKN, Soluble ammonium, total nitrogen, and the TSS. The results obtained from the work confirmed that ANNs can readily be used in simulation work for WWTP. In particular, it emerged that simulation with ANN was 36 times faster than simulation with mechanistic model such as ASM1.

Chen et al. (2003) develop a recurrent neural network model to predict the nitrogen contents in treated effluents to be used for ground water recharge. The model uses three online parameters, pH, Oxidation-reduction potential (ORP), and dissolved oxygen (DO) in conjunction with three off-line nutrient tests, BOD, Ammonium nitrogen (NH_4-N) and nitrate nitrogen (NO_3-N). The outputs were

total nitrogen, NO_3-N and NH_4-N. The accuracy of this model was over 90% in this work. The quality of the data has not been mentioned in term of missing values and outliers.

Gamal – El-Din and Smith (2002) develop an ANN model to predict wastewater inflow rate that enters the Gold Bar WWTP, Alberta, Canada. The neural model uses rainfall data observed in the collection system discharging to the plants as inputs to the ANN model. Eight rain gauges were selected to be used by the neural network inputs. An index to represent the day of the week and another index to represent the hour of the day were used as inputs to the neural network model as well. The raw data flow records had some negative values, which indicated segments of faulty data. These segments of data were eliminated and considered as data gaps.

Traoré et al. (2006) developed a fuzzy model of the sludge height in the secondary settler. The modelling strategy was based on simple on-line data (influent, removal and recycle flows) and daily analytical values of the sludge volume index (SVI) allowing the fuzzy algorithm to reduce sludge height variations and thus to increase the settling process efficiency. The developed model has then been adapted and applied to the Cassà de la Selva activated sludge WWTP (Spain). The influent flow and the sludge volume index have been used as inputs to estimate the sludge recycle and removal flows. The use of fuzzy logic as control tool made it possible to combine two kinds of knowledge. The process qualitative knowledge or fuzzy rules were obtained from experts and operators working on the plant. The results obtained showed the fuzzy controller efficiency for both increasing and decreasing SVI values. The application of the controller to the Cassà-WWTP data also allowed an important reduction of suspended solids concentration fluctuations.

Polit et al. (2001) developed fuzzy estimators for the concentration of the substrate at the input, the total and partial alkalinity in the influent and in the effluent and the volatile fatty acid (VFA) concentration at the output of the reactor of a pilot-scale anaerobic digestion reactor for the treatment of raw industrial wine. Those observers were built on available on-line measurements like pH, temperature, input flow rate and output gaseous flow rate. The fuzzy observers follow quite well the measured values of the effluent partial and total alkalinity and of the VFA at the output of the reactor. Those quantities are difficult to measure on-line and nevertheless those values are very useful in the process knowledge.

It is clear from the above discussion that a variety of neural networks models have been used in WWTP research. However, attention has focused on supervised neural networks and there is less directed at exploiting the Kohonen self organising map or (KSOM) in the field of wastewater treatment. Few exemptions of such applications are discussed in the following paragraphs.

Garcia and Gonzalez (2004) apply KSOM for clustering data obtained from a steelworks wastewater treatment plant in order to estimate, monitor and visualise the process states. They use a combination of the KSOM and other clustering techniques to estimate and monitor the diverse state of the wastewater treatment, and they establish the correlation among process variables which is necessary in order to obtain a knowledge based system.

Hong et al. (2003) applied the KSOM to analyse the multi-dimensional process data, and to diagnose the inter-relationship of process variable in a real activated sludge WWTP. By using the component planes, they discovered some local relationship between process variables under different operating conditions. They found that the KSOM provides an effective analysing and diagnosing tool to

understand the system behaviour and to extract knowledge contained in multi-dimensional data of a large-scale WWTP.

Cinar (2005) used KSOM to classify operational data of the Pelham WWTP in Western Carolina and to determine the reasons of high effluent concentration of BOD, TSS, and faecal Coliform. He found that the reasons causing high effluent concentration of these parameters were the high pH in the biological reactor and the high solid retention time (SRT).

Gonzalez and Garcia (2006) proposed a self organising map and clustering algorithms to achieve the aerobic end point detection of a sequencing batch reactor in a coke wastewater treatment plant. They outlined the validation methods for KSOM training and testing the predefined criterion to determine the KSOM size.

The above merely introduces the plausibility of AI modelling techniques for wastewater treatment plants system. Further details about the nature, structure, strength and limitations of AI models in treatment plant applications are given in the next chapter.

2.5 Summary

This chapter reviewed the background of activated sludge wastewater treatment plants. It presented the importance of wastewater treatment, its history and the structure of a conventional activated sludge wastewater treatment plant. The biochemistry and microbiology of the process are described. A review of the state of the art in modelling activated sludge wastewater treatment plants is also presented and discussed.

The chapter concluded that the complexity of ASM1 and the detailed microorganism's growth and decay data required by it mean that it is not

appealing to process operators. The ASM1 requires the determination of about 31 parameters, coefficients and variables, most of which are not routinely measured by large wastewater treatment plants. Furthermore, a certain number of simplifications and assumptions must be made in order to make a model of a WWT system practically useful. Some of these are associated with the physical system itself, while others concern the mathematical model.

The next Chapter will introduce the essential background of Artificial Intelligence Techniques used in this study. These techniques are backpropagation Artificial Neural Networks, Kohonen Self organising map and Fuzzy Logic. The adaptive neural networks based on fuzzy inference system will also be presented and discussed.

CHAPTER 3

ARTIFICIAL INTELLIGENCE TECHNIQUES (AI)

3.1 Introduction

Artificial intelligence (AI) techniques concern several areas relating to the simulation of human intelligence in a computing machine. However, research in AI techniques suffered in the past due largely to a lack of understanding of how the human brain actually works, but also due to the disbelief about the thinking machine, and the fact that the computational power required to apply these techniques was unavailable at that time. In 1973, the Lighthill Report (Lighthill, 1973), in the United Kingdom concluded that there was no future in AI research and recommended that all research funding in the area be terminated. Despite this, however, work on AI has continued due to advances in the computing field.

The primary attraction of AI techniques is that they are able to represent systems with non-linear characteristics, without the difficult task of dealing with deterministic non-linear mathematics. The existing modelling strategies of AI can be divided into three categories, namely, ‘white-box’, ‘black-box’, and ‘grey-box’ based on the type of knowledge used for the model development. In ‘white-box’ modelling strategies, also called deterministic models, the model development is mainly driven by the knowledge of the relevant mechanisms and balances. In other words the model equations are developed from general balance equations applied to mass and other conserved quantities, resulting in a set of differential equations. A ‘black-box’ or input-output model is mainly driven by the measured data obtained from the process. However, black-box model is only as good as the

data that were used to calibrate them (Gernaey et al., 2004). Indeed, 'black box' models are not believed to have any extrapolation properties; consequently, one has to obtain a large body of data that cover the possible range of fluctuations in the relevant input variables for process modelling. Finally, 'grey-box' model may be defined as a suitable combination of 'black-box' and 'white-box' model, in such a way that the model is developed using data driven techniques and at the same time can extract some useful knowledge from the data (The Mathworks, System Identification Toolbox http://www.mathworks.com/products/sysid/).

Among commonly used AI tools and techniques are Artificial Neural Networks (Black Box), Fuzzy logic (Grey Box), Expert Systems (white box), and a wide variety of search techniques such as genetic algorithms. However, the rest of this chapter will present just ANN and Fuzzy logic, as they are the core of this work.

3.2 Artificial Neural Networks

An artificial neural network (ANN), often just called "neural network" (NN) is a mathematical model or a form of computing algorithms inspired by the functioning of the biological nervous system (biological neural networks). In practical terms, neural networks are non-linear statistical data modelling tools used to model complex relationships between inputs and outputs or to find patterns in data. In most cases, an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. In other words, knowledge is acquired by the network through a learning process and the inter connections between the elements of the network store the knowledge (Arbib, 2003).

In the past decades, interest in neural networks has increased dramatically. The number of papers published relating to ANNs and their applications is an

indication of the success of ANNs. Broadly, ANNs have been used to solve complex problems in various fields of application such as, pattern recognition, identification, classification, speech, vision, acoustics, robotics, image processing, financing, control systems, aerospace, banking, defence, electronics, manufacturing, medical, oil and gas exploration, securities, telecommunications and transportation. In recent years, ANNs have also been applied to address real life problems in neuroscience, biological science, earth science, physical science, chemical engineering, civil engineering, structural engineering, translation engineering, and others.

In addition, ANNs have been used successfully to solve water resources problems such as rainfall-runoff modelling (Chiang et al., 2004), rainfall forecasting (Olsson et al., 2004), Water Demand modelling (Pulido-Calvo, 2007), generalised storage-yield- reliability planning (Adeloye and DE Munari , 2006). Moreover, ANNs have been used in forecasting water consumption (Pulido-Calvo and Gutiérrez-Estrada, 2008), water treatment process (Liu and Kim, 2008). The success with which ANNs have been used to model dynamic systems suggests that the approach may prove to be effective and efficient way to simulate the complex wastewater treatment systems. Hence, the literature contains some applications of ANNs applied to model wastewater treatment systems as discussed in section 2.4.4.2.

3.2.1 Historical overview

McCulloch and Pitts (1943) are recognised as the founding fathers of neural networks. They developed simple models of neural networks based on their understanding of neurology. However, since their seminal work, other groups have made further attempts to develop neural networks. For example, Rosenblatt (1958) made considerable improvement in the neural field when he designed and

developed the Perceptron. The Perceptron had three layers and could learn to connect or associate a given input to an output unit. Werbos (1974) developed and used the well-known back-propagation learning method, which is the most widely applied neural networks to-date. In the early 1980s, Hopfield (1982) introduced the network architecture. In the same year, Kohonen (1982) introduced the self-organising features map (KSOM) to capture the essential features of the data. During the late 1980's and early 1990's, different combinations of network topologies were investigated.

3.2.2 Inspiration from Neuroscience

NN is inspired by knowledge from neuroscience but it draws its methods from statistical physics (Arbib, 2003). Therefore, ANNs are based on the parallel architecture of the human brain. The brain is composed of about 1011 neurons (nerve cells). Figure 3.1 is a schematic drawing of a single biological neuron. Tree-like networks of nerve fiber called dendrites are connected to the cell body. Extended from the cell body is a single long fiber called the axon that branches into strands or parts. At the end of these are the transmitting ends of the synaptic junctions to other neurons. The transmission of a signal from one cell to another is a complex chemical process. The ease of transmission of signals is altered by activity of the nervous system. The ability to adjust signals constitutes the mechanism for learning (Poznanski, 2001).

The fundamental aspects of ANNs are the use of simple processing elements, which are models of the neurons in the brain. These elements (neurons) are then connected together in a well-structured way. The network then is taught, to achieve a particular task or function of interest, by patterns of data presented, such that it can subsequently not only recognise such patterns when they occur again, but also recognise similar patterns by generalisation (Abrahart et al, 2004).

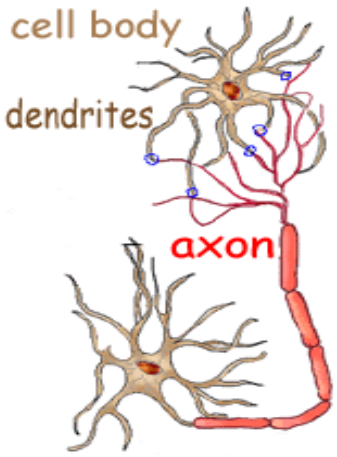


Figure 3.1 *A schematic drawing of a biological neuron (www.utexas.edu/reearch/asrec)*

3.2.3 The architecture of ANN

3.2.3.1 The neuron

A neuron is the basic processing unit of a neural network, taken from the Greek meaning nerve cell. Figure 3.2 illustrates a single input neuron with scalar input *p* transmitted through the connection which is multiplied by the strength of the scalar connection weight *w* to form the product *w*p*. This product is summed along with the scalar bias *b* to form an output scalar *n*. The bias *b* provides an additional variable that can be adjusted to obtain the desired network performance. This sum is the argument of the transfer function *f*, which takes the argument and produces the output *a*. The transfer function limits the permissible amplitude range of the output signal to some finite values in the range [-1, 1]. Both *w* and *b* are adjustable parameters that enable the network to exhibit the

desired behaviour. Thus, the output *a* can be calculated as in Equation 3.1 (Demuth and Beale, 1998).

$$a = f(n) = f(wp + b) \tag{3.1}$$

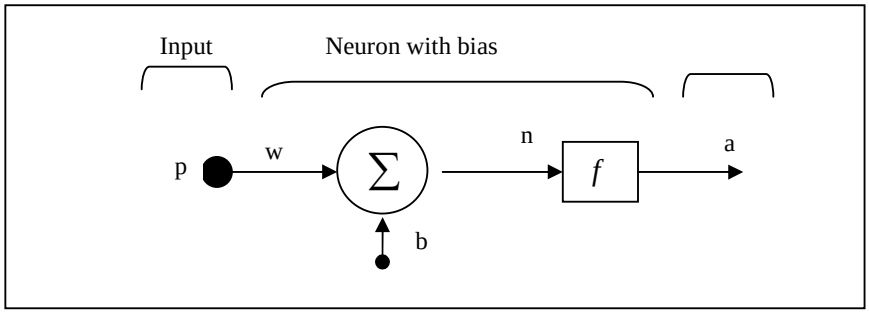


Figure 3.2 *Schematic of Single input neuron*

In a multiple input neuron , illustrated in Figure 3.3., the individual inputs $p_{(1)}$, $p_{(2)}$, …, $p_{(r)}$ are weighted with elements $w_{(1,1)}$, $w_{(1,2)}$,…., $w_{(1,r)}$. These weighted values are inputs to the summing junction. The notation $w_{(i,j)}$ indicates that the scalar weighting has destination neuron *i* from source input/neuron *j*. As with the single input neuron, the bias *b* is summed with the weighted inputs to form the net input *n* and this is the argument presented to the transfer function *f*, as is in Equation 3.2.

$$a = f(n) = f(WP + b) \tag{3.2}$$

Where W is the weight Matrix and P is the input vector

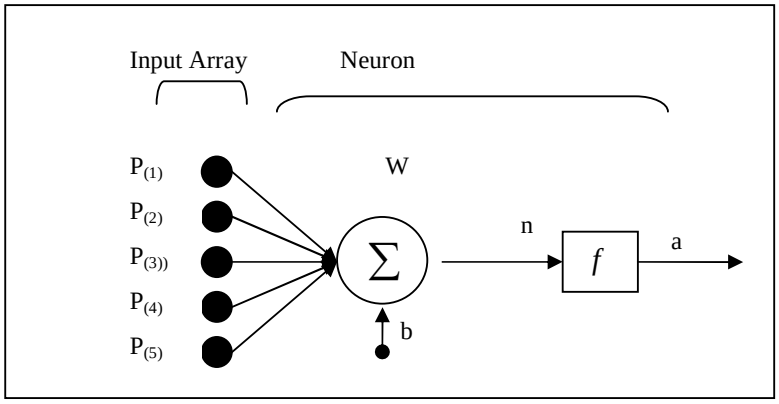


Figure 3.3 *Multiple input neuron*

Two or more neurons can be combined together in a layer. A layer includes the combination of weights, the multiplication and summing operation, the bias *b* and the transfer function *f*. It should be noted that, the number of neurons in a layer do not need to equal the number of inputs to that layer. A typical network consists of a sequence of layers with a connection weights between successive layers. Usually, these layers are called input layer, hidden layers and output layer as illustrated in Figure 3.4 (Demuth and Beale, 1998).

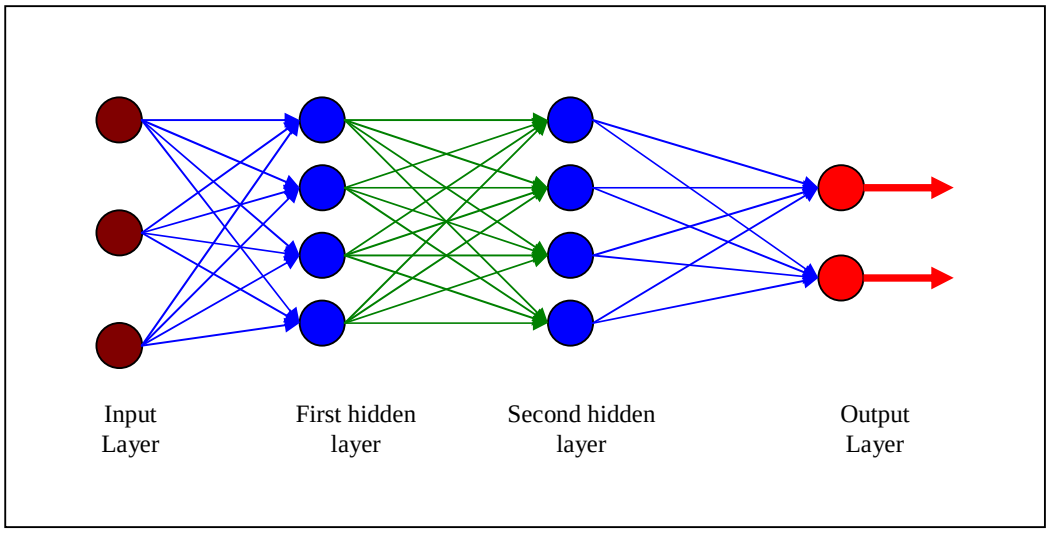


Figure 3.4 *Multilayer networks with two hidden layers*

3.2.3.2 Types of transfer Functions

Each hidden or output unit in a neural network receives values through the connections from the units of a previous layer and they are combined to a single value. After the combination, the scalar value is passed through a transfer function, also known activation function, which gives the power to the neural network to handle non-linearities. The choice of transfer functions depends on the complexity of the application. Many transfer functions are available in the literature (Figure 3.5); however, the most widely used transfer functions are: hard limit transfer function, linear transfer function, and sigmoid transfer function.

The hard-limit transfer function limits the output of the neuron to either 0, if the input argument *n* (wp+b), is less than 0, or 1, if *n* is greater than or equal to 0 as can be seen from Equation 3.3.

$$f(n) = \begin{cases} 1 & n \geq 0 \\ 0 & n < 0 \end{cases} \tag{3.3}$$

In the linear transfer function, when *n* is the input to the neuron and *a* is the output after passing through the transfer function, the linear transfer function can be written as in Equation 3.4. The linear function is a popular choice for the output layer function as it allows the output to take any value.

$$a = f(n) = n \tag{3.4}$$

Sigmoid transfer function is a simple activation function that can introduce non-linearity to the network. Sigmoid function produces output in the shape of 'S'; hence the term sigmoid function. There are many variations of sigmoid functions in the literature; however, the most widely used is the logistic (ie. Log-Sigmoid) that takes the form in Equation 3.5.

$$a = f(n) = \frac{1}{1+e^{-n}} \tag{3.5}$$

The output of logistic sigmoid function ranges between zero and one. 'Sigmoid' functions are the most popular choice function for the hidden layer of a network due to its easily calculated derivative. The difference between hard limit transfer function and logistic sigmoid transfer function is that whereas a hard limit assumes the value of zero or one, a logistic function assumes a continuous range of values from zero to one.

3.2.4 Neural network topology

As stated previously, ANNs generally consist of a number of interconnected processing elements or neurons. How the inter-neuron connections are arranged determines the structure of a network. How the strengths or the weights of the connections are adjusted or trained to achieve a desired overall behaviour for the network is governed by its learning algorithm. Both the structure and learning algorithm constitute the architecture of the network. Therefore, the architecture defines the network structure, i.e. the number of the neurons in the network and their interconnectivity. Based on the architecture, ANNs can be grouped into two categories, namely feed-forward neural networks (FFNNs) and recurrent neural networks (RNs) (Bishop, 1995; Demuth and Beale, 1998).

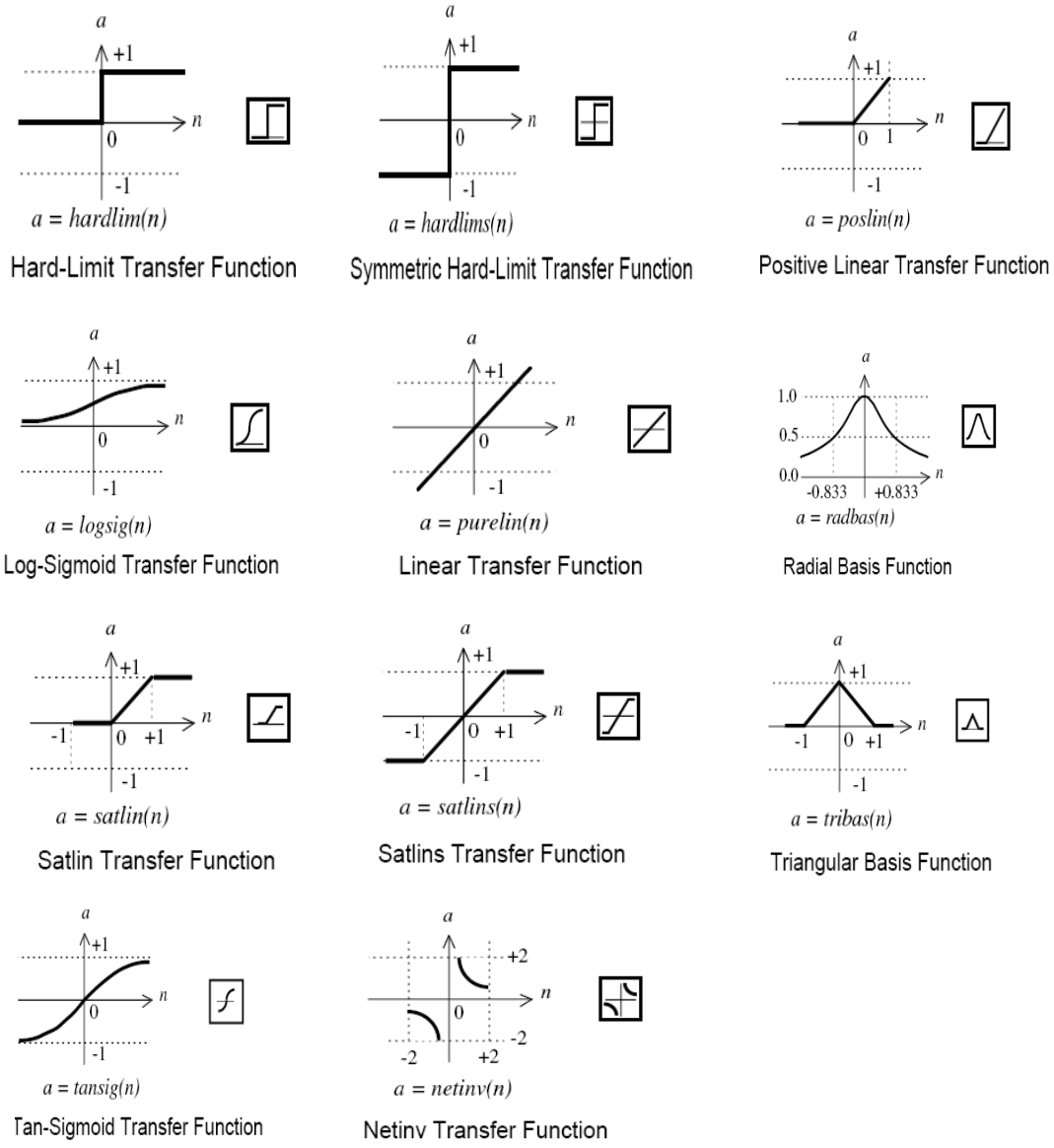


Figure 3.5 *Transfer functions Graphs (Demuth and Beale, 1998).*

3.2.4.1 Feed-forward neural networks (FFNNs)

In the FFNNs, neurons are organised into layers where information is passed from the input layer to the final output layer in a unidirectional manner as illustrated in Figure 3.4. This type of ANNs is capable of mapping the given set of inputs to their corresponding outputs. Therefore, the FFNN is only capable of statically

mapping the input vectors to their corresponding targets. However, they are still widely used in dynamic system mapping by feeding the past input and current output values of the system to be modelled as inputs to the network (Bishop, 1995; Gurney, 1997, Demuth and Beale, 1998). The widespread use of such a network is due to its ability to model complex functional relationship between the given input and output data sets by learning from examples.

3.2.4.2 Recurrent Neural Networks (RNNs)

In the recurrent neural networks (RNNs), feed-back connections within the network either between layers and/or between neurons can be found as illustrated in Figure 3.6. Therefore, RNNs are dynamic, meaning that the output at time *t* is dependent on the previous output or state of the neurons within the network as the result of the feed-back paths. The internal feedback paths allow the network to exhibit temporal behaviour and the greater the number of feedback interconnections, the richer the dynamic representation. It can be said that the introduction of feedback makes RNN a nonlinear dynamic system (the matheworks,Neural-networks-toolbox,
http://www.mathworks.com/products/neuralnet/). The feedback loops involve the use of particular branches composed of unit-delay elements (denoted by D) which results in a nonlinear dynamic behaviour. As with FFNN, RNN can be multi-layered (MRNN) or single-layered (SRNN) and fully or partially connected.

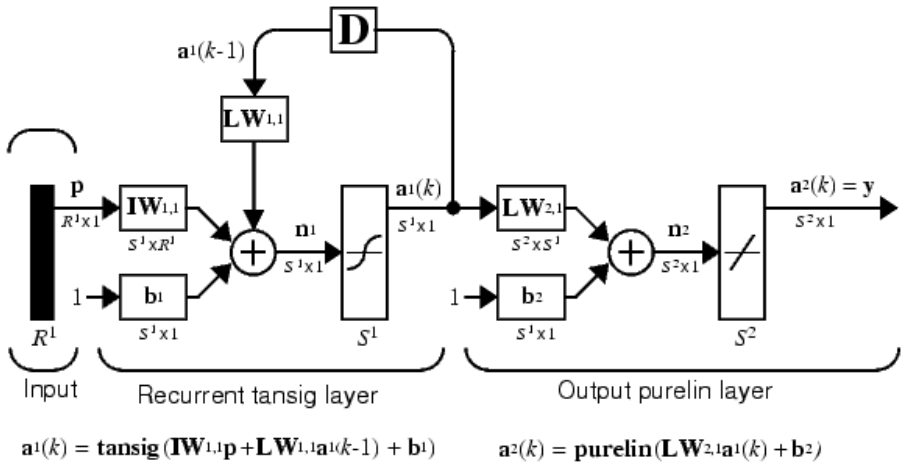


Figure 3.6 *Recurrent neural networks*

3.2.5 Modelling using ANN

Modelling is the process of finding a model that best regenerated the original output signals when subjected to the same input signals. ANNs could be utilised to model a process without having to take into account the complex physical laws that govern the system. Therefore, the task of modelling using ANNs essentially involves finding a suitable model structure and subsequently finding good numerical values for its parameters (weights and biases of the network), i.e. establishing the architecture. The architecture of a neural network should be optimised in order to achieve a better interaction with the system of interest. Unfortunately, there is no well theory for choosing the architecture of the ANN. Therefore, the number of neurons in the hidden layer(s) is generally determined through trial and error (Bishop, 1995; Demuth and Beale, 1998). However, the theoretical basis of non-linear modelling by using ANNs has been well established in the last decades. It has been shown that one hidden layer, having sufficient number of neurons, can approximate any complex relationships.

3.2.6 Learning Algorithm

Learning algorithm is the process in which the weights and biases in NN are adjusted in response to input-output training data set. The learning process enables the network to find a set of weights that will produce the best possible input/output mapping. Normally, the best mapping is achieved by minimising of the error between network output and the desired output. Learning algorithm for neural networks can be classified into two major algorithms: supervised learning and unsupervised learning.

3.2.6.1 Supervised learning algorithms

Supervised learning, as the name implies, requires an external reference (teacher) to match each input vector with a desired output. When an input vector is introduced, the network proceeds to calculate the output of this input vector. Then the error between the network output and the desired output is calculated. This error is often used to modify the weights according to an adopted learning algorithm. The weights are then adjusted with each training iteration or epoch until the error for the entire set of training vectors reaches an acceptable level. A schematic representation of the supervised training is illustrated in Figure 3.7. An example of supervised learning algorithm is back propagation (Bishop, 1995; Demuth and Beale, 1998).

3.2.6.1.1 Back-propagation algorithm

The feed foreword multi layer perceptron artificial neural networks (FFMLP) are trained with the popular backpropagation algorithm. This training algorithm is used in perhaps 80 to 90% of practical modelling applications. Back propagation neural networks models are very effective in capturing the non-linear relationships that exist between input-output variables in complex systems. In other words,

back propagation ANN can be viewed as a form of regression model between input and output variables.

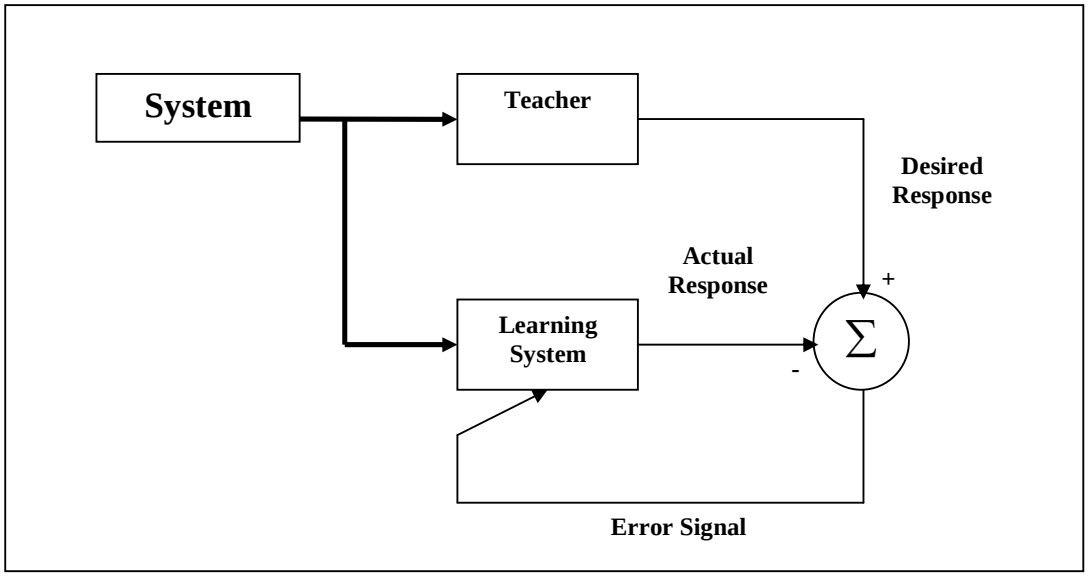


Figure 3.7 *Schematic representation of supervised learning*

The learning paradigm of backpropagation neural networks utilises a gradient descent optimisation method for the learning process. In this method, the network typically starts with randomly generated weights. Then it is exposed to a training set of input-output data. At each iteration or epoch, the network weights and biases are updated in the direction in which the performance function decreases. This means that it learns by making changes in its weights in a direction to minimise the objective function, e.g. mean square errors between its computed output and target output. As the training proceeds, the network's weights are adjusted until it is responding within the required limits of accuracy (Demuth and Beale, 1998).

The Backpropagation learning rules consists of two passes of the different layers of the ANN, a feedforward pass and backward pass. In the feedforward pass, the

input is applied to the ANN and is passed through the different layers, at this stage, the weights and biases are fixed and do not change. At the ANN output, an error is recorded which is the difference between the ANN output and the desired output. During the backward pass, the weights and biases are adjusted in accordance with the error-correction rules. The error signal is then propagated backward through the ANN and the parameters are adjusted layer by layer until the whole layers are covered.

To teach the neural network training data set is needed. The training data set consists of input signals (p_1 and p_2) assigned with corresponding target (desired output) d. During training, the output predicted by the network $a(t)$ is compared with the actual (desired, target) output $d(t)$ and the mean square error (MSE) between the two is calculated. The error function (or instantaneous value) at time t, E(t), is given by Equation 3.6.

$$E(t) = \frac{1}{2}\sum_{j \in C} \delta^2(t) = \frac{1}{2}\sum_{J \in C} (a(t) - d(t))^2 \tag{3.6}$$

where the set C includes all the neurons in the output layer.

Then the error is propagated back to adjust the weights using Equation 3.7:

$$W_{ji}(t+1) = W_{ji}(t) + \Delta W_{ji}(t) \tag{3.7}$$

The weight increment ΔW_{ji} is calculated using Equation 3.8 in which the gradient descent method is applied. This results in weights being changed in the direction of steepest descent down the error surface. The size of the step taken down the error surface is determined by the learning rate η.

$$\Delta W_{ji} = \eta(\partial E / \partial W) \tag{3.8}$$

The learning rate η affects network-teaching speed. There are a few techniques to select this parameter. The first method is to start teaching process with large value of the parameter. While weights coefficients are being established the parameter is being decreased gradually. The second, more complicated, method starts teaching with small parameter value. During the teaching process, the parameter is being increased when the teaching is advanced and then decreased again in the final stage (Demuth and Beale, 1998).

3.2.6.1.2 Enhancing the performance of backpropagation neural networks

Several problems are associated with the performance of the back propagation artificial neural networks. These are namely, over-fitting, local minima, and convergence. To avoid such problematic conditions and be successful in building the best-suited back propagation network for a particular application, it is important to be aware of these technical aspects. There is, however, no universal rule for avoiding their problems completely, but there are some rules given by practitioners, who have worked in the field. Some of these rules will be summarised in the next subsections. For now, the essential features of these problems are briefly explained.

In the case of over-fitting, the mapping ability of neural networks can lead to a very accurate fit of the training data but result in a poor generalisation of unseen data (Bishop, 1995; Demuth and Beale, 1998). In such a case, the model is not only modelling the essential dynamic of the system but also undesirable features such as noise. One method of solving the problem of overfitting is to ensure the neural network is just large enough to provide an adequate data fit. However, it is difficult if not impossible to know beforehand how large the network should be. The most widely used method for improving the generalisation of neural networks is the early stopping (Demuth and Beale, 1998).

The problem of local minima is associated with the gradient descent procedure that may lead the training to be trapped in local minima of the error surface. Although it sometimes provides an acceptable solution, a network trapped into a local minimum during learning is likely to exhibit poor performance in terms of learning and generalisation capabilities (Rocha et al., 2007).

The third problem is that of convergence. In this case, the learning process in back propagation is unpredictable and lengthy, and this is perhaps the most serious problem with the algorithm. Whilst some complex problems may require hours or days to train a network, under some circumstances, there is a possibility that the network will be unable to improve or achieve an acceptable performance at all (Kamarthi and Pittner, 1999).

The above mentioned problems can be solved using several methods. These methods are summarised as:

a. Sample size

A simple way to alleviate the above-mentioned problems is to increase the number of examples in the database (Sahiner et al., 2008). Large sample size decreases the noise effects and improves generalisation of the network.

b. Pre-processing and post-processing

This method is used to scale all signals, both input and output, to the same variance. Hence, signals of differing magnitudes are equalised to ensure that signals of larger magnitudes will not become too dominant. This would certainly ensure that all the input signals apply the same influence throughout the training process. Moreover, scaling makes for a numerically more robust training algorithm and leads to faster convergence in neural network learning (Demuth and Beale, 1998). After the network has been trained, the outputs of the network have

to be post-processed to give the required output values using the inverse of the pre-processing transformation. When new data are to be presented to the network, they must first be pre-processed. They are then post processed to return to the original variable as output.

A procedure for scaling network data set is to normalise it in which it will have zero mean and unit variance. The input and target variables are treated independently and for each variable x_i; its mean $\bar{x}_i$ and standard deviation σ_i and a standardised variable, $x_{i\text{-}nor}$, is then obtained using:

$$x_{i-nor.} = \frac{x_i - \bar{x}_i}{\sigma_i} \tag{3.9}$$

c. Early Stopping

Since the goal of the network training is not to learn the exact representation of the training data itself but to build a model of the process that generates the data, it is important that the network exhibits good generalisation. Early stopping is the most widely used technique to overcome the overfitting problem and to find the network having the best performance on new data. Early stopping involves the splitting of the available data into 3 subsets: training set, validation set and testing. During network training, the error on the validation set is monitored as well as the error on the training set. Training will continue until the error on the validation set increases implying overfitting (Bishop, 1995; Demuth and Beale, 1998). Training can therefore be stopped at the point of the smallest error with respect to the validation data set, since this gives a network that is expected to have the best generalization performance. After the model has been trained, another data set, testing set, is used to verify the effectiveness of the stopping criterion and to estimate the expected performance in the future (Adeloye and De Munari, 2006).

d. Selection of the network elements

The choice of the number of layers and the number of hidden neurons in the hidden layers is a major concern in constructing a network, as they do not follow simple rules, but are a result of a process of trial and error. In theory, more hidden neurons produce better mapping ability. However, in practice, the networks may use the extra nodes to fit the noise in the data (Beltratti et al., 1996; Pulido-Calvo et al., 2007). Although having no formal mathematical basis and being the subject of a lot of research, some suggestions from empirical research indicate that one-hidden-layer network with different number of hidden neurons is capable of accurate approximation to any complex system over a reasonably sample set (Demuth and Beale, 1998). Moreover, the number of inputs is often unknown and different models with different inputs can be trained in order to select the optimal model. The decision of the final model is usually determined by evaluating the trained model using several evaluation criteria on new data.

3.2.7 Advantages and Limitations of Neural Networks

The above are a clear demonstration of the usefulness of modelling complex environmental systems, particularly those exhibiting strong non-linearities which are difficult or impossible to specify in closed mathematical forms. ANNs offer other advantages that can be summarised as follows (Bishop, 1995):

1. Although neural networks have the potential to solve complex problems, they are inherently simple to understand and develop.

2. ANNs require no explicit knowledge of the system under study, which makes them well suited to applications where knowledge extraction is difficult or in cases where the interrelationship between process parameters are hard to model.

3. Compared to deterministic or mechanistic models, ANN internally encodes knowledge on a network of nodes and connections. The knowledge in the network is developed by using sufficient amount of actual historical observations from past experience. Therefore, there is no need to acquire rules or algorithms from human expert or from the mechanism of the process.

4. ANNs learn by examples, and as long as examples are available and an appropriate design is adopted, effective solutions can be constructed far more quickly than is possible using traditional mechanistic models, which are entirely reliant on experience in a particular field.

5. With careful design, ANN can be trained to give the correct response to data that have not been previously encountered or used during training. This aspect is often described as the ability to generalise on test data.

6. The use of neural networks have the potential to reduce time spent on modelling complex processes and can give better performance than other mechanistic models.

However, a major drawback of using ANNs is that they are not able to simulate outputs outside the range of those they were trained with, i.e. they are poor extrapolators (*Vos and Rientjes, 2005*).

3.2.6.2 Unsupervised learning algorithms

As stated in subsection 3.2, there are two ways to train a network: supervised and unsupervised. In supervised learning, the network is presented with examples of known input-output data pairs, after which it starts to mimic the presented input-

output behaviour or pattern. The network is then tested to see whether it is able to produce correct output, when new input is presented to it. In unsupervised learning, the network learns on their own, in a kind of self-study without teacher. In such a case, a data set is presented to such a network and they learn to recognize patterns in the data set. By so doing, the input data are categorised into groups or clusters. Of course, the net does not understand the meaning of the groups. It is up to human users to interpret or label the groups in some meaningful way (Back et al., 1998; Kalteh et al., 2008). The most widely used unsupervised neural network is the Kohonen Self-Organising Map, KSOM. Therefore, and due to the importance of this method in this research it will be discussed details in the next section.

3.3 Kohonen Self-Organising Map (KSOM)

The KSOM (also called feature map or Kohonen map) is one of the most widely used unsupervised artificial neural networks algorithms (Kohonen et al., 1996). It is usually presented as a dimensional grid or map whose units (nodes or neurons) become tuned to different input data patterns. Its algorithms are based on unsupervised competitive learning, which means that training is entirely data driven and the neurons or nodes on the map compete with each other. In contrast to supervised neural networks, which require that target values corresponding to input vectors are known, KSOM dose not require the desired output to be known, hence, no comparisons are done to predetermine the ideal responses. During training, only input patterns are presented to the network which automatically adapts the weights of its connections to cluster the input patterns into groups with similar features (Alhoniemi, 1997; 1998, Obu-Can, 2001; Astel et al., 2007).

The principal goal of the KSOM is to transform an incoming signal pattern of arbitrary dimension into a two-dimensional discrete map. It involves clustering

the input patterns in such a way that similar patterns are represented by the same output neurons, or by one of its neighbours (Back et al., 1998). In this way, the KSOM can be viewed as a tool for reducing the amount of data by clustering, thus converting complex, nonlinear statistical relationship between high dimensional data into simple relationship on low dimensional display (Kangas, 1995; Kohonen et al., 1996; Zhang, 2009). This mapping roughly preserves the most important topological and metric relationship of the original data elements, implying that the KSOM translates the statistical dependences between the data into geometric relationships, whilst maintaining the most important topological and metric information contained in the original data. Hence, not much information is lost during the mapping. Hence, similarities relationship within the data and clusters can be visualised in a way that enables the user to explore and interpret the complex relationship within the data set.

3.3.1 The structure of the KSOM

The KSOM consists of two layers: the multi-dimensional input layer and the competitive or output layer; both of these layers are fully interconnected as illustrated in Figure 3.8. The output layer consists of M neurons arranged in a two-dimensional grid of nodes. Each node or neuron i (i = 1,2,…,M) is represented by an n-dimensional weight or reference vector $Mi= [mi_1,….,mi_n]$. The weight vectors of the KSOM form a codebook. The M nodes can be ordered so that similar neurons are located together and dissimilar neurons are remotely located on the map.

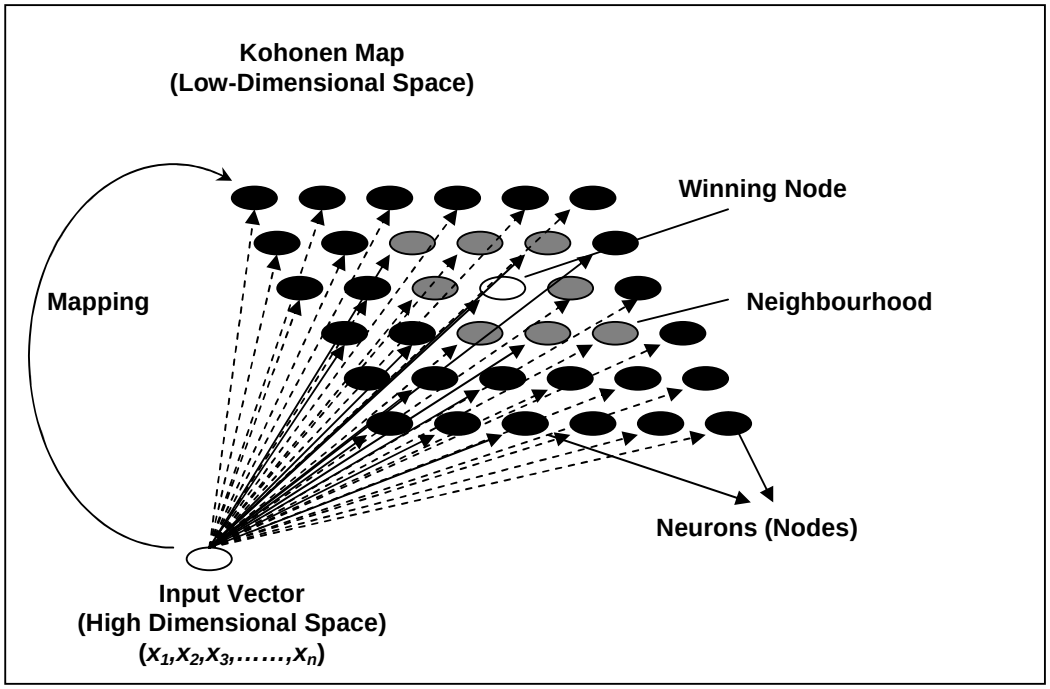


Figure 3.8 *Illustration of the winning node and its neighbourhood in the Kohonen Self-organizing Map*

The network topology is described by the number of output neurons presented in the network and by the way in which they are interconnected. Usually neurons in the output layer are arranged in either a rectangular or a hexagonal grid as seen from Figure 3.9. In a rectangular grid, each neuron is connected to four neighbours, except for the ones at the edge of the grid. However, in a hexagonal structure, every neuron is connected to exactly six neighbours, except for the ones at the edge of the grid (Back et al., 1998; Vesanto et al., 2000). There is a need to point out that while the rows and the columns on the output layer are interpreted as co-ordinate axes to locate units and upon which the output of the KSOM can be interpreted, they do not have explicit meaning or relations to the variables of the input data set.

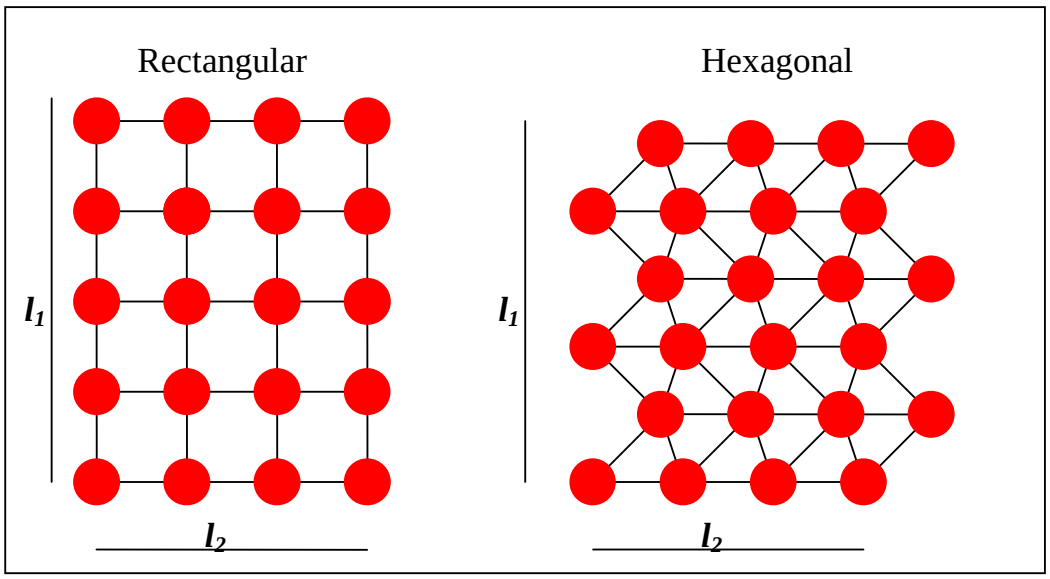


Figure 3.9 *Examples of map topologies in the KSOM*

The number of neurons (Map size), M, may vary from a few dozen up to several thousands, where the number of neurons affects accuracy and generalisation capability of the KSOM (Alhoniemi, 1997; 1998) as will be seen later on in this subsection. The self-organising map team at the Helsinki University of Technology offers guidance for determining the optimum number of neurons using the heuristic formula presented in Equation 3.10 (Vesanto et al. 2000; Garcia and Conzalis, 2004).

$$M = 5\sqrt{N} \tag{3.10}$$

where M is the number of map units or neurons and N is the total number of data samples or records. Once M is known, the number of rows and columns in the KSOM can be determined. A guideline by the self-organising map team is in Equation 3.11:

$$\frac{l_1}{l_2} = \sqrt{\frac{e_1}{e_2}} \tag{3.11}$$

where l_1 and l_2 are the number of rows and columns respectively, e_1 is the biggest eigenvalue of the training data set and e_2 is the second biggest eigenvalue.

3.3.2 Training the KSOM

Upon start of training, the initial values of the elements of the weight vectors in the grid are randomly assigned, usually numbers between zero and one. Then the weight vectors are updated using two types of training algorithms, sequential training algorithms and batch training algorithms. Both of these algorithms follow the same general procedures depicted in Figure 3.10.

3.3.2.1 Sequential training algorithms

The multi-dimensional input data is first standardised by deducting the mean and then dividing the result by the standard deviation (see Equation 3.9). This procedure ensures that every variable has equal importance in training the KSOM, so that no components will have excessive influence or control of the training results by virtue of its higher absolute value (Alhoniemi, 1998). Then a standardised input vector is chosen at random and presented to each of the individual neurons in the output layer or map for comparison with their code vectors in order to identify the code vector most similar to the presented input vector. The identification uses the Euclidian distance, which is defined in Equation 3.12.

$$D_i = \sqrt{\sum_{j=1}^{n} (x_{ij} - m_{ij})^2}; i = 1,2,...,M \tag{3.12}$$

Where Di is the Euclidian distance between the input vector and the weight (or code) vector i; x_{ij} is the j^{th} element of the current input vector; m_{ij} is the j^{th} element of the weight vector i, n is the dimensionality of both the input and the code vector; and M is the number of neurons in the KSOM (or the size of the

map). The neuron whose vector most closely matches the input data vector (i.e. for which the Di is minimum) is chosen as a winning node or the best matching unit (BMU) as indicated in Figure 3.10. The weight vectors of this winning node and those of its adjacent neurons are then adjusted to match the input data using Equation 3.13, thus bringing the weight vectors further into agreement with the input vector as seen from Figure 3.11 (Vesanto et al.,2000).

$$m_i(t+1) = m_i(t) + \alpha(t)h_{ci}(t)[x(t) - m_i(t)] \tag{3.13}$$

where *t* denotes time, $\alpha(t)$ is the learning rate at *t*, $h_{ci}(t)$ is the neighbourhood function centred in the winner unit *c* at time *t*, ($h_{ci}(t)$ defines the region of the influence that the input sample has on the KSOM as will be seen later on in this subsection) and all the other variables are as defined previously.

The nodes surrounding the winning node, its neighbourhood, are also updated so that they are made to look less like the input vector as seen from Figure 3.10. The size of adjustment in the weight vector of the neighbouring neurons is dependent on the distance of those neurons from the winner in the output array (Back et al., 1998; Kalteh et al., 2008). This adaptation procedure stretches the BMU and its topological neighbours towards the sample vector.

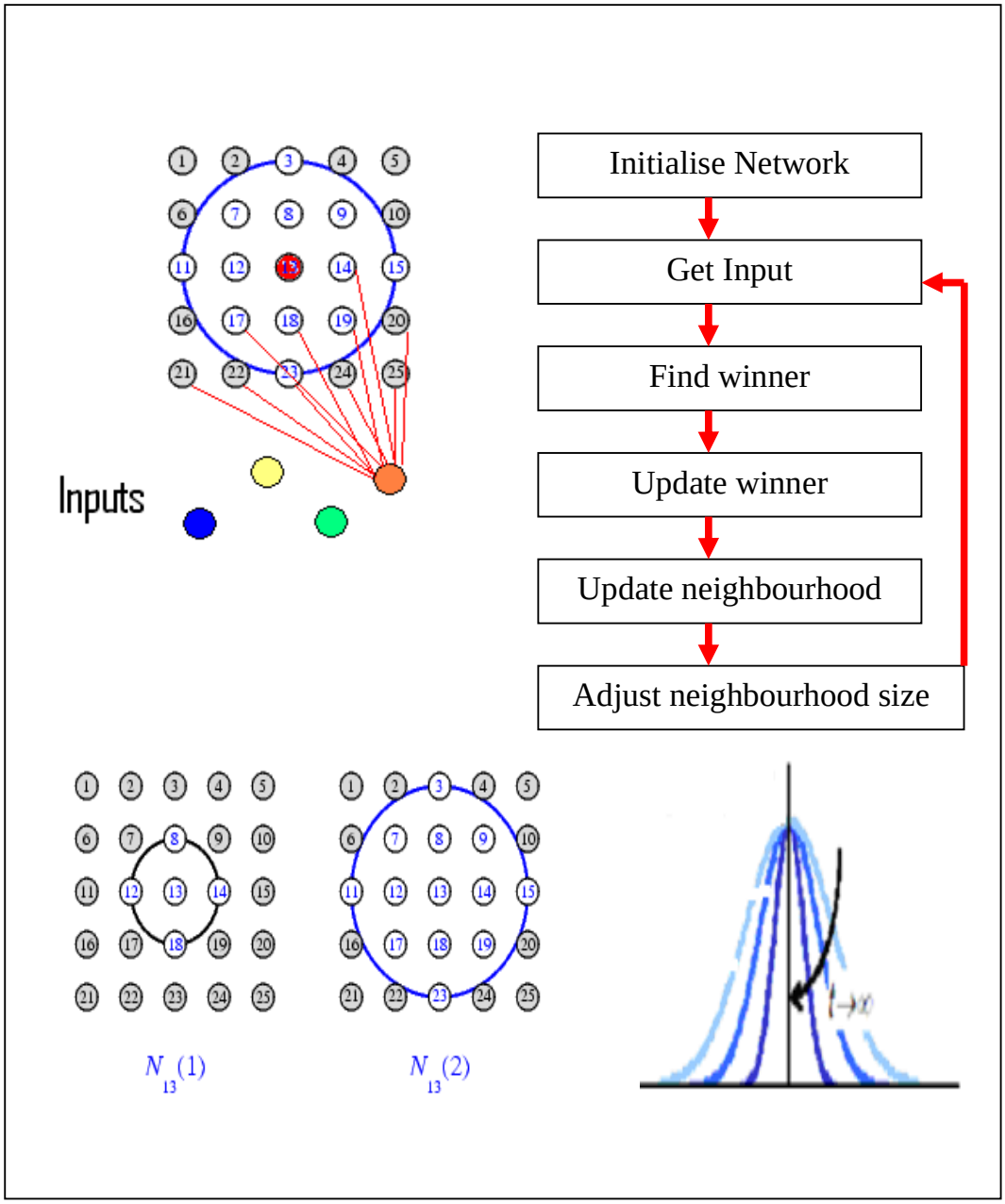


Figure 3.10 *the training procedures of the KSOM. Node 13 is the wining node, it can be seen how the neighbourhood of the BMU (node 13) moves toward the BMU with each iteration.*

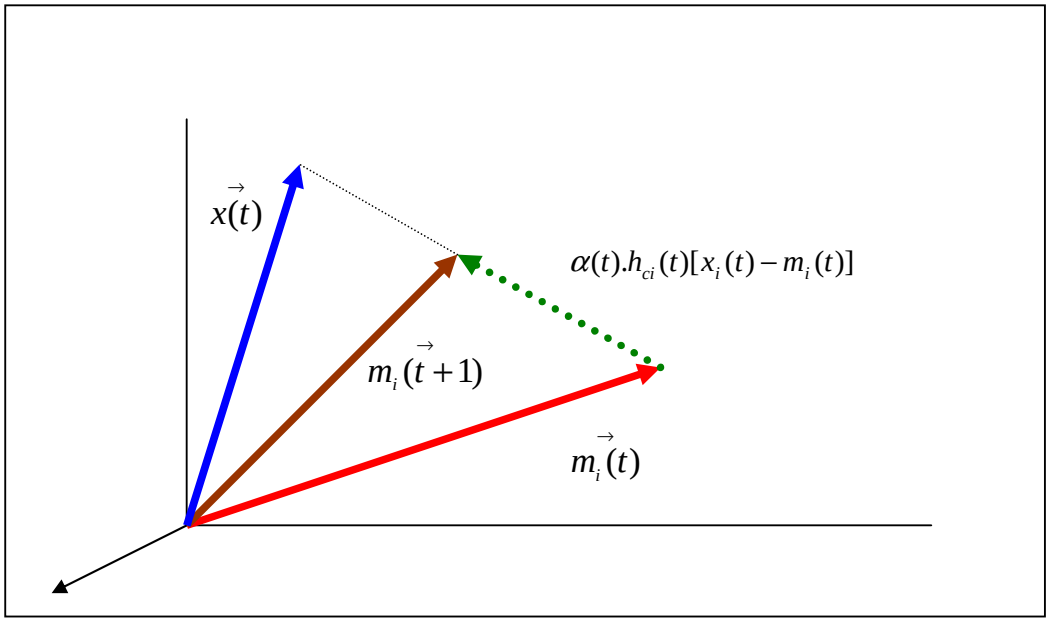


Figure 3.11 *Prototype vector m_i (t) of the neuron i is updated close to data vector x(t) to be $m_i(t+1)$.*

This process of random selection of a data record, competition of the winning node, and updating the node and its neighbourhood repeats many times-from 1,000 to 25,000 or more. As the process continues, the area of the neighbourhood decreases, as does the magnitude of the adjustment to the winning and neighbourhood nodes. In this manner, each node in the map internally develops the ability to recognize input vectors similar to itself. This characteristic is referred to as Self-Organising, because no external information is supplied to lead to a classification (Penn, 2005).

3.3.2.2 Batch training algorithms

Batch training algorithm is also an iterative process, but instead of using a single data vector at a time, the whole data set is presented to the map before any

adjustments are made-hence the name "batch." In each training step, the data set is partitioned between the map units according to their distance from that unit (Kohonen, 2001), i.e. each data vector belongs to the map unit to which it is closest. After this, the new weight vectors are calculated as:

$$m_i(t+1) = \frac{\sum_{j=1}^{n} h_{ic}(t) x_j}{\sum_{j=1}^{n} h_{ic}(t)} \tag{3.14}$$

where c is the index of the BMU of data sample x_j

In other words, the new weight vector is a weighted average of the data samples, where the weight of each data sample is the neighbourhood function value $h_{ic}(t)$ at its best BMU unit c.

Two parameters are used for training the KSOM: the learning rate ($\alpha(t)$) and the neighbourhood width parameter (h_c). The learning rate influences the size of the weight vector adjustments after each training step, whereas the neighbourhood width parameter determines to what extent the surrounded neurons are affected by the winner. An additional parameter is the training length, which measures the processing time, the number of iterations through the training data (Back et al., 1998; Vesanto et al., 2000). Both the learning rate and the neighbourhood width are time dependent and are typically changed from large to small in order to provide the best performance with the smallest training time.

There are different learning rate functions as shown in Figure 3.12. Linear function as in Equation 3.15 (solid line), power function as in Equation 3.16 (dot-dashed), and other decreasing function such as Equation 3.17 (dashed):

$$\alpha(t) = \alpha_0 (1 - t/T) \tag{3.15}$$

$$\alpha(t) = \alpha_0 (0.005/\alpha_0)^{t/T} \tag{3.16}$$

$$\alpha(t) = \alpha_0 /(1+100t/T) \tag{3.17}$$

where T is the training length or the number of iterations and α_0 is the initial learning rate. In the KSOM toolbox of Matlab, α_0 is specified as 0.5 (Vesanto, 2000).

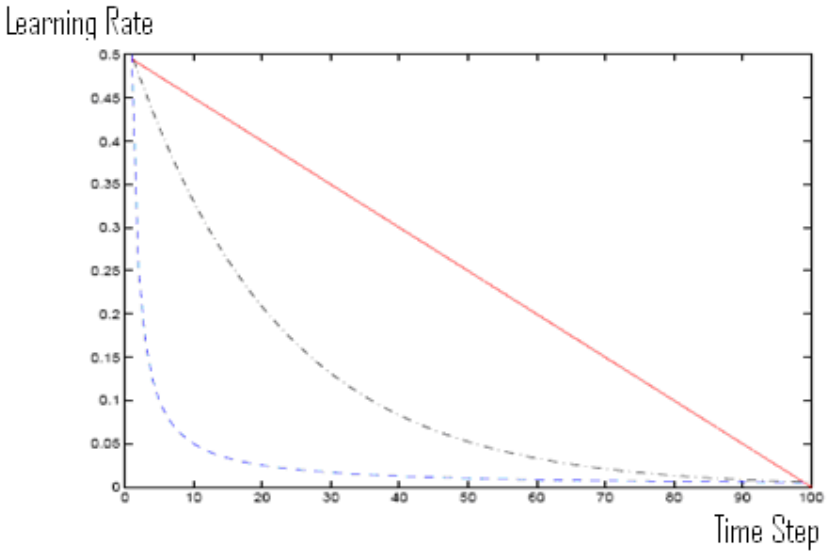


Figure 3.12 *Different types of learning rate decreasing functions. Linear function (solid line), power function (dot-dashed), and other decreasing function such as Equation (3.17) (dashed).*

Also there are different neighbourhood functions, however, the neighbourhood function is normally chosen to be Gaussian centred in the winner unit c, such that:

$$h_{ci}(t) = \exp^{-(d_{ci}(t))^2/(2\sigma^2(t))} = \exp^{-(\|r_c - r_i\|))^2/(2\sigma^2(t))} \tag{3.18}$$

In other words, all neurons *i* located in a topological neighbourhood of the winning neurons c will have their weights updated usually with a strength related

to their distance d_{ci} from the winning neuron, where d_{ci} can be calculated using formula 3.19.

$$d_{ci} = \|r_c - r_i\| \tag{3.19}$$

where r_c and r_i are the positions of nodes c and i on the KSOM grid known as the norm city-block distance.

σ^2 is the variance parameter specifying the spread of the Gaussian function.

For this neighbourhood function the distance is determined considering the distance in the lattice in each dimension, and the one having the maximum value is chosen as d_{ci} . For example, if d_{ci} =2 corresponds to a square around BMU having side length of 3 . The weights of all neurons within this square are updated with, while the others remaining unchanged. As the training progresses, this neighbourhood gets smaller and smaller, resulting in a situation that only the neurons very close to the winner are updated towards the end of the training.

To illustrate the concept of neighbourhoods, considering Figure 3.10, the diagram illustrates a two-dimensional neighbourhood of radious d=1 around nodes 13 and another diagram shows a neighbourhood of radious d=2. These neighbourhoods could be written as:

N13 (1) = {8,12,13,14,18} and

N13(2)={3,7,8,9,11,12,13,14,15,17,18,19,23}

Like the learning rate $\alpha(t)$, $\sigma(t)$ also decreases linearly as the number of iterations increases. In the early stages of training, the radius of the neighbourhood is large and most of the KSOM neurons belong to any node's neighbourhood. As the training progresses, the radius is reduced to allow good

local ordering as seen from Figure 3.10. In the KSOM tool box in Matlab, the initial radius σ_0 is specified as max (1, M/4), where M is the size of the map.

The batch training algorithm was used because its implementation in Matlab is considerably more efficient than that of the sequential training algorithm as it requires less time for training and produce less quantisation and topographic errors (see section 3.3.3 below).

3.3.3 The quality of the KSOM

The quality of the trained KSOM is measured by the total average quantisation error and total topographic error (Garcia and Gonzalez, 2004). The quantisation error measures the quality of the map fitting to the data, i.e. the average distance between each data vector and its BMU at convergence. This error is calculated using:

$$q_e = \frac{1}{N}\sum_{i=1}^{N}\left\|x_i - m_c\right\| \tag{3.20}$$

where q_e is the quantisation error, x_i is the i-th data sample or vector m_c is the prototype vector of the best matching unit for x_i and $\left\|.\right\|$ denotes Euclidian distance (Equation 3.12).

The topographic error, t_e, is an indication of the degree of preservation of the topology of the data when fitting the map to the original data set. In other words, it measures the similarity between the neighbour on the model and on the input space. It is calculated as the proportion of sample vectors for which two best and the next best matching units for a given input vector are not adjacent, i.e.,

$$t_e = \frac{1}{N}\sum_{i=1}^{N} u(x_i) \tag{3.21}$$

where N is the number of samples, x_i is the i-th data sample and $u(x_i)$ is a binary integer such that it is equal to 1 if the first and second best matching units of the map are not adjacent units; otherwise it is zero. The results of this error measure are very easy to interpret and are also directly comparable between different models.

The number of map units determines the accuracy and generalization capability of the KSOM. The bigger the map size the lower the quantization error but the higher the topographic error. Moreover, the bigger the map size the higher the computational cost. Therefore, there is compromise between the increases of the topographic error and the reduction of the quantization error. A reasonable optimum solution of the compromise among the quantization error and the topographic error to determine the side lengths from the map is the heuristic formula presented in Equation 3.11 as stated previously.

3.3.4 Applications of KSOM

Over 5000 publications relating to the KSOM were documented in the last twenty years (Kaski et al. 1998; Oja et al. 2003). The KSOM can be used for many practical tasks, such as the reduction of the amount of training data for model identification, nonlinear interpolation and extrapolation, generalisation and compression of information for easy transmission (Kohonen et al.,1996; Kangas and Simula,1995). Indeed, the KSOM has been used for a wide variety of applications, mostly for engineering problems but also for data analysis (Tananaki et al., 2007; Badekas and Papamarkos, 2007). However, the most important applications of the KSOM have been in the visualisation of high-dimensional

systems and process data and the discovery of categories and features in raw data. This application is called the exploratory data analysis or data mining (Kohonen et al., 1996; Kangas and Simula, 1995).

KSOM could be used to pre-process the incoming information in order to improve the performance of ANN or Fuzzy Logic models. This pre-processing ability would become quite significant when there was a combination of faulty information or uncertainty in the measurement system. The application of the KSOM for data pre-processing is described in more detail in Chapter 5 and and by Rustum and Adeloye (2007).

3.4 Fuzzy Logic modelling technique

Fuzzy logic models (also called linguistic models or fuzzy if-then rules) were first introduced by Mamdani (Mamdani and Assilian, 1975) based on Zadeh's theory of fuzzy sets. These models have the capability to deal with systems that are highly uncertain (Klir et al., 1988; Firat and Güngör, 2007). The theory of fuzzy sets, firstly published by Zadeh (1965), presents a useful way of representing the uncertainty and imprecision in data without the need of complex mathematical relationships. These models have the advantages of being able to model non-linear functions in an easy and understandable way by explaining the reasoning linguistically rather than with numerical quantities. They provide a useful way of representing human knowledge in a readable way in the form of fuzzy rules (Nguyen and Walker, 2006). In the following subsections, the fuzzy logic concepts are presented. Furthermore, details about the ANFIS, which utilise the learning power of ANN to optimise the fuzzy rules and the parameters of fuzzy membership functions in the fuzzy logic system, are also described.

However, before this, it is pertinent to mention the difference between multiple regression or neural networks and fuzzy logic. Whilst the first two belong to black-box models, fuzzy logic models fall in the category of grey-box models (see section 3.1). This is because artificial neural networks regression are able to learn complex non-linear relationships between inputs and outputs of the process, but are not able to help improve the heuristic understanding of the operational problem or causal relationships of the process parameters. Fuzzy logic models are grey-box because the causal relationships between the conditions (inputs) and the fact (output) are clear in such fuzzy if-then rules. Thus, Fuzzy models depend not only on black box approach, but also on a combination of knowledge of the system and human expert in which fuzzy rules are built using physical knowledge of the process, while the parameters are tuned in a black-box manner (Sadiq et al., 2004).

3.4.1 Fuzzy Logic concepts

Zadeh stated that "*as complexity rises, precise statements lose meaning and meaningful statements lose precision*". From this statement, Zadeh (1965) introduced the concept of fuzzy logic where the truth of any statement becomes a matter of degree. This theory is an extension of conventional Boolean logic that was introduced to handle the concept of partial truth between completely true and completely false (Zadeh, 2008). Zadeh used this concept as a mean to model the uncertainty of natural language. This technique is used for modelling of processes that are complex and ill-defined. A fuzzy logic model consists of linguistic *if-then* rules that depend on fuzzy set theory for representation and evaluation using computers.

3.4.2 Fuzzy Sets

Fuzzy set theory is a mathematical way to represent and deal with vagueness in everyday life. Zadeh (1965) provided a theory that one of the reasons human are better at control than machines is that they are able to make effective decision on the basis of imprecise linguistic information. Fuzzy theory generalises classical sets theory in which the membership degree of any object to a set is not restricted to the integers [0 1], but may take any value between zero and one. By this definition, a fuzzy set is a set with imprecise boundaries in which the transition from one set to another is gradual rather than abrupt (Ross, 2004; Zadah, 2008; Fuzzy Logic Toolbox for use with Matlab, 2004).

In classical mathematics, a classical set is a set with crisp boundary. For example, let X be a certain universe of discourse, where $X = \{x_1, x_2, ..., x_n) : x_n \in X\}$, and its elements x_n denote all the possible tall values (cm) of an adult person. A classical crisp set C_{tall} of X is defined as a function μ called characteristic function of C_{tall} as in Equation 3.22. For any element x of the universe X, the characteristic function μ is equal to 1 if x is an element of set A, and is equal to 0 if x is not an element of A. Just as tall, another two similar crisp sets $C_{average}$ and C_{short} can be defined as in Equations 3.23 and 3.24 respectively:

$$\mu_{tall}(x) = \begin{cases} 1 & if \quad x \quad is \quad larger \quad than \quad 180cm \\ 0 & otherwise \end{cases} \tag{3.22}$$

$$\mu_{aver.}(x) = \begin{cases} 1 & if \quad x \quad is \quad between \quad 160cm \quad and \quad 180cm \\ 0 & otherwise \end{cases} \tag{3.23}$$

$$\mu_{short}(x) = \begin{cases} 1 & if \quad x \quad is \quad less \quad than \quad 160cm \\ 0 & otherwise \end{cases} \tag{3.24}$$

One problem arises if we have to define a linguistic term "*tall*" for example. The use of any crisp set above results in a stiff situation, when a person of 180 cm is considered to be a "*tall man*", but a 179.99 cm-person is said to be "*not tall*". In contrast to a classical set above, a fuzzy set is a set with fuzzy boundaries. The membership function of a fuzzy set is allowed to have values between 0 and 1, and it expresses the degree in which an element belongs to a given fuzzy set. This transition makes fuzzy sets more flexible and intelligent for the interaction between humans and machine.

Using the same example as above, new fuzzy sets F_{tall}, $F_{average}$, and F_{Short} of X can be defined as in Equations 3.25, 3.26, and 3.27 respectively.

$$F_{tall} = \{(x, \mu F_{tall}(x)), x \in X \quad (3.25)$$

$$F_{aver} = \{(x, \mu F_{aver}(x)), x \in X \quad (3.26)$$

$$F_{short} = \{(x, \mu F_{short}(x)), x \in X \quad (3.27)$$

where μF is called the membership function (MF), and it gives the degree to which x is an element of set F. This degree, a value between 0 and 1, denotes the degree of membership, also called membership value. The difference between fuzzy and crisp definition of tall is better illustrated using Figure 3.13. For example, if a person is 170 cm tall then the membership degree for the fuzzy subset tall is about 0.6. At the same time, the membership degree for the fuzzy subset short equals to 0, and the membership degree for fuzzy subset average is equal to 1.

These membership functions map each element of X to a membership degree between 0 and 1. They are often Triangular, Trapezoidal, Gaussian, and Bell-shaped functions as illustrated in Table 3.1. When the membership function takes only two values 0 and 1, F is identical to a crisp set that is defined by a characteristic function. In this instance, crisp sets can be looked at as special cases of fuzzy sets. In practice, the input space, universe of discourse, is usually partitioned into several fuzzy sets whose MFs cover X. These fuzzy sets carry names that conform to adjectives appearing in our daily linguistic usage, such as low, medium and high. These adjectives are called linguistic values or linguistic labels, and the universe of discourse is called the linguistic variable.

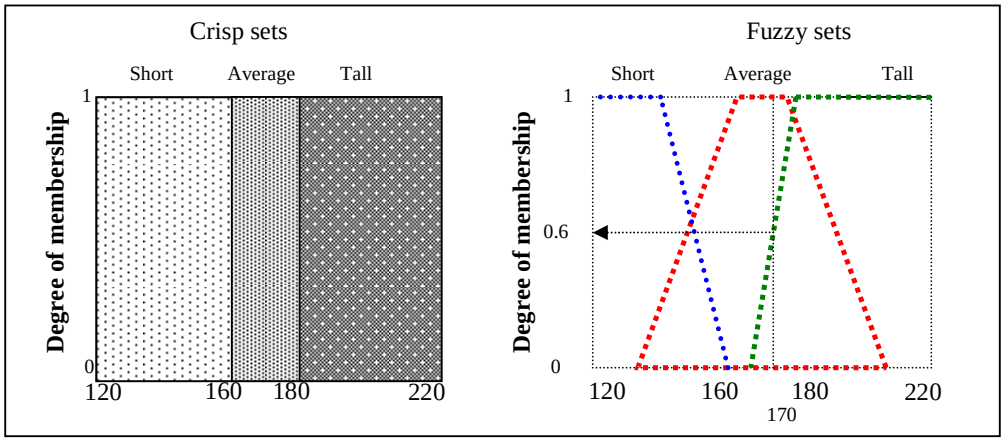


Figure 3.13 *Typical crisp sets (left) and typical fuzzy sets (right) characterising the human tall values.*

3.4.3 Fuzzy Logic System

Fuzzy logic system (FLS) is a rule based system in which an input is first fuzzified (converted from crisp number to a fuzzy value) and subsequently processed by an inference engine that retrieves knowledge in the form of fuzzy rules contained in a rule-base (Ross, 2004). The fuzzy sets computed by the fuzzy inference as the output of each rule are then composed and defuzzifide (converted

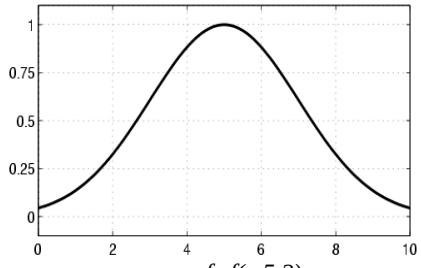
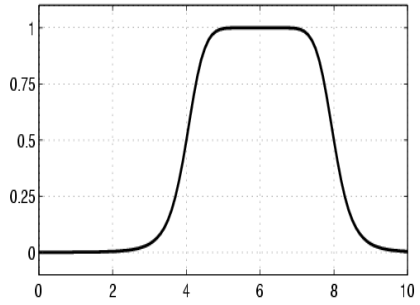
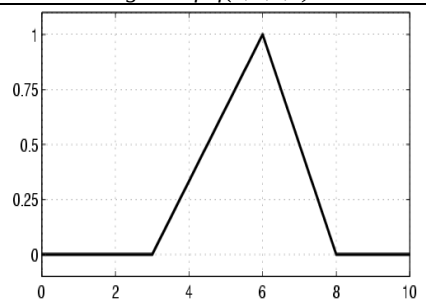
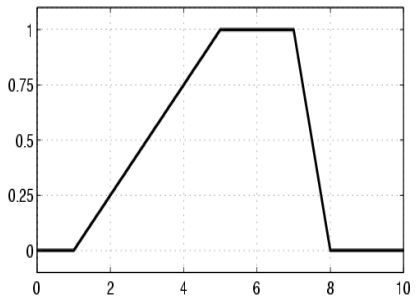
from fuzzy set to a crisp number). By this, fuzzy logic systems are a nonlinear mapping from the input to the output space. The basics of fuzzy logic systems are based on three concepts, fuzzy sets, linguistic variables, and fuzzy if-then rules.

The first implementation of the FLS idea was described by Mamdani and Assilian (Mamdani et al., 1974; 1975; 1977) which demonstrated its viability for the control of a small steam engine. Other early implementations of the FLS were a sludge controller for a wastewater treatment plant (Tong et al., 1980).

3.4.3.1 Architecture of Fuzzy Logic System

As noted previously, Fuzzy Logic Systems (FLS), also known as fuzzy-rule-based systems, fuzzy expert systems, fuzzy associative memory, or fuzzy controllers when it is used in control areas (Jang & Sun, 1995; Ross, 2004), consists of four basic components: the fuzzification inference, knowledge base (rule base or database), decision making (inference engine, or inference mechanisms) and defuzzification (Fuzzy Logic Toolbox for use with Matlab, 2004). These various components are related as illustrated in Figure 3.14. They involve membership functions, fuzzy logic operators, and fuzzy rules. The membership functions allow the representation of a degree of membership to a fuzzy set for a given numerical value, associated to a linguistic label. The fuzzy IF-THEN rules introduce the expert knowledge in a computable way.

Table 3.1 *Schematic of different membership functions*

Name	Description	Example
Gaussian curve built-in membership function	The symmetric Gaussian function depends on two parameters b and a as given by: $f(x;a,b) = e^{\frac{-(x-a)^2}{2b^2}}$	 *gaussmf=f(x,5,2)*
Generalised bell-shaped built-in membership function	The generalised membership function depends on three parameters, a,b, and c as given bye: $f(x;a,b,c) = \frac{1}{1+\left\lvert\frac{x-c}{a}\right\rvert^{2b}}$ where the parameter b is usually positive. The parameter c locates the centre of the curve	 *gbellmf=f(x;2,4,6)*
Triangular-shaped built-in membership function	The triangular curve is a function of three scalar parameters a, b, and c, as given by: $f(x;a,b,c) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ \frac{c-x}{c-b}, & b \le x \le c \\ 0, & c \le x \end{cases}$ The parameters *a* and *c* locate the "feet" of the triangle and the parameter c locates the peak.	 *trimf=f(x;3,6,8)*
Trapezoidal-shaped built-in membership function	The trapezoidal curve is a function of a vector, x, and depends on four scalar parameters a, b, c, and d, as given by $f(x;a,b,c,d) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \frac{d-x}{d-c}, & c \le x \le d \\ 0, & d \le x \end{cases}$ The parameters a and d locate the "feet" of the trapezoid and the parameters b and c locate the "shoulders."	 *trapmf=f(x;1,5,7,8)*

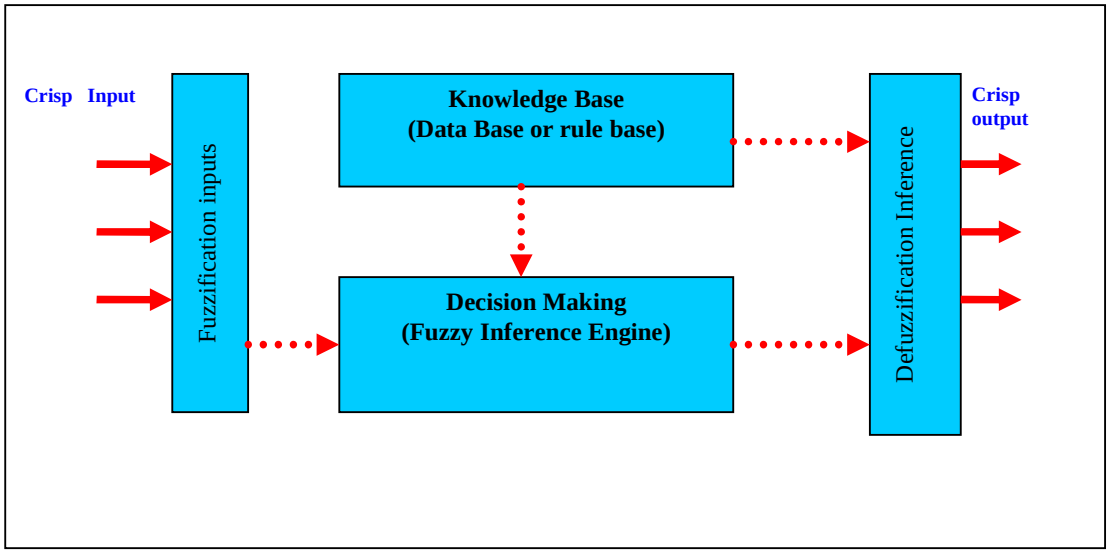


Figure 3.14 *The basic structure of a fuzzy Logic inference system and its components*

1 The fuzzification unit

The aim of the fuzzification unit is to obtain the membership degree of each input, i.e. fuzzifying the inputs. The data are processed in this unit and converted into linguistic variables by means of membership functions. The determiners of the membership functions are linguistic values (e.g. very low, low, normal, high, and very high). The outputs of this layer are fuzzy membership degree of the inputs. The relation between input and output using Gaussian membership function as a fuzzification process is as in Equation 3.28.

$$y_{ij} = \exp(-\frac{(x_i - a_{ij})^2}{2.b_{ij}^2}) \qquad (3.28)$$

where y_{ij} is the membership degree for the input x_i in membership function *ij*, a_{ij} is the centre and b_{ij} is the width of the membership function *ij* as seen from Figure 3.15. These parameters can be used to define the region or the position of

the fuzzy sets. Essentially, for this particularly type of MF, a_{ij} is the center of membership function *ij* and b_{ij} is its width as described in Table 3.1.

2 The Knowledge base

The knowledge base consists of database and linguistic rule base (or fuzzy rule base).

a. The database

The database includes all the definitions used for membership functions, fuzzy partitioning and definition of fuzzy sets. Partitioning of a universe of discourse is used to determine the initial number of fuzzy subsets required to represent that universe. These fuzzy sets can be represented as a function form such as Gaussian membership function, bell-shaped, triangle-shaped, trapezoid function, etc as stated previously.

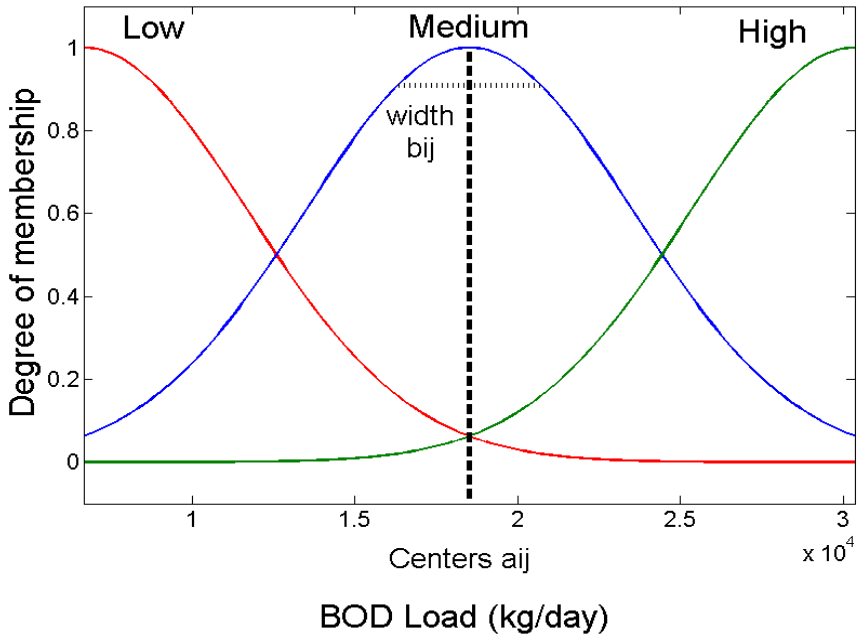


Figure 3.15 *Parameters of Gaussian membership function for BOD Load as developed in the study. Low=gaussmf(BOD Load; 6670,5030),*

Medium=gaussmf(BOD Load; 18510,5030) and High=gaussmf(BOD Load;30360,5030).

b. The rule base

The rule base contains the fuzzy rules obtained from the physical system of the process. These rules associate a condition using linguistic variables and fuzzy sets to make a conclusion. These rules can be extracted from different sources such as expert knowledge, operators control actions or knowledge extracted from the data. These fuzzy rules take the form:

IF (a set of conditions occur) **THEN** (a set of conclusions can be inferred)

Thus fuzzy rule has two parts: the *IF* or antecedent part that describes a condition and the *THEN* or consequent part that describes the conclusion. The antecedent part of each rule classifies the behaviour of measured variables by fuzzy membership functions, whereas the consequent part expresses the essential actions or conclusion. The collection of fuzzy rules forms the rule base of a FLS. An example that describes a simple fact is:

IF Pressure is high, **THEN** volume is small

where, as in the above example, pressure and volume are linguistic variables, high and small are linguistic values or labels that are characterised by membership functions.

There are several factors that can influence the design and the implementation of fuzzy rules. These factors are the choice of input and output variables, the generation methods of fuzzy rules and the implementation method on fuzzy rules.

i) The choice of input variables

The choice of input variables influences the number of rules and the performance of the FLS. The selection of these variables relies on experience on one hand and the relation between these inputs and the desired output on the other. These relations can be determined using correlation matrix or the visualisation of the component planes of the KSOM.

ii) Generation method of fuzzy rules

There are two methods to derive fuzzy rules. The first one is to generate fuzzy rules based on expert experience and control engineering knowledge. In this method, the expert put his experience as a linguistic relation between input and output variables of the FLS. This method is very difficult to be used if the process is very complex. The second method is based on the observed input-output data (Fuzzy Logic Toolbox for use with Matlab, 2004). As remarked earlier, the first approach was commonly used until the introduction of ANFIS (Tong et al, 1980; Mamdani et al., 1974; 1975; 1977).

iii) Implementation method of fuzzy rules

Fuzzy implication rule describes how several logic formulas involving linguistic variables are combined together. The combination can be achieved in many ways, all of which are derived from three fundamental operations, conjunction (AND), disjunction (OR), negation (NOT), in addition to the implication (Production rule).

- The conjunction (intersection) of two fuzzy sets A and B is a fuzzy set C written as $C = (A \cap B)$ or $C = A \quad AND \quad B$. This fuzzy set C is a collection of objects that belong to both A and B and whose MF is related to those A and B by Equation 3.29.

$$\forall x \in U : \mu_C(x) = \mu_{A \cap B}(x) = \mu_A(x) \cap \mu_B(x) = \min\{\mu_A(x), \mu_B(x)\} \quad (3.29)$$

- The disjunction (union) of two fuzzy sets A and B is a fuzzy set C, written as $C = (A \cup B)$ or $C = A \quad OR \quad B$, which is a collection of objects that belongs to either A or B and whose MF is related to those A and B by Equation 3.30.

$$\forall x \in U : \mu_C(x) = \mu_{A \cup B}(x) = \mu_A(x) \cup \mu_B(x) = \max\{\mu_A(x), \mu_B(x)\} \quad (3.30)$$

- The complement of set A denotes by A^C or NOT A is a collection of objects not belonging to the set A and whose MF is related to A and B by Equation 3.31.

$$\forall x \in U : \mu_A^C(x) = 1 - \mu_A(x) \quad (3.31)$$

The operations of union, intersection, and complement introduced in the previous definitions are graphically illustrated in Figure 3.16. Note that these operations perform exactly as the corresponding operations for ordinary sets if the values of the membership functions are restricted to either 0 or 1.

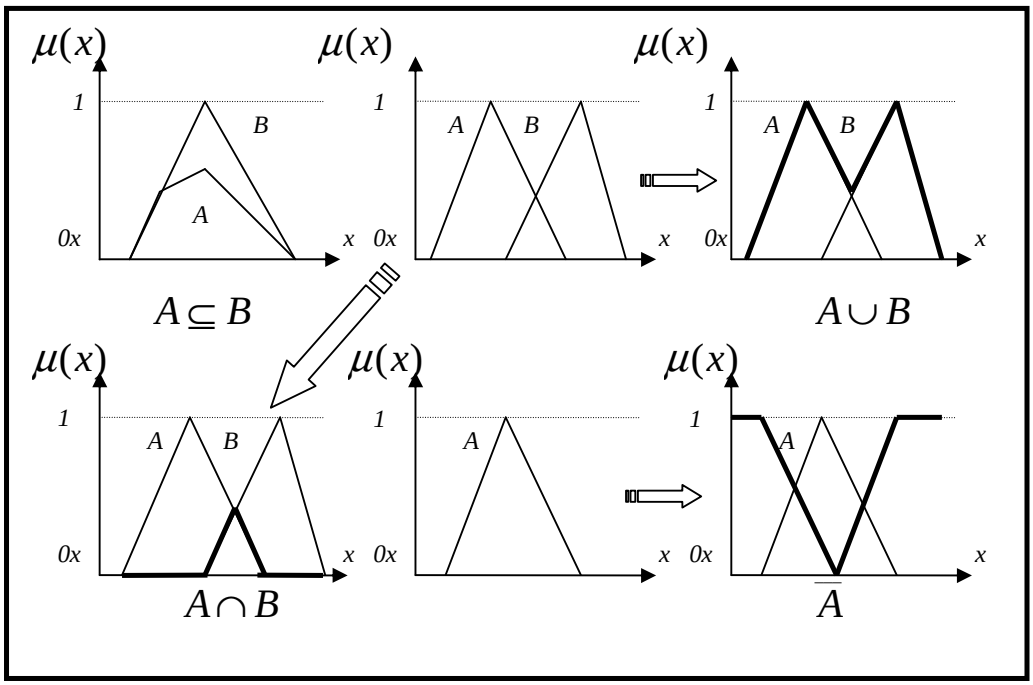


Figure 3.16 *Graphical examples of containment, union, intersection and complement*

3 The fuzzy inference

The fuzzy inference (or inference mechanism or decision-making) unit applies a fuzzy reasoning mechanism to obtain a fuzzy output; in other words, combine the results of fuzzification process in a single fuzzy output for each rule. There are several types of fuzzy inference systems. However, the most commonly used are Mamdani inference system, type 1 in Figure 3.17 and Sugeno inference system, type 2 in Figure 3.17 (Ross, 2004).

4 The defuzzification unit

The defuzzification unit aggregates the outputs of all of the rules that have been fired for a particular input to produce a crisp output. In other words, the fuzzy output is transferred back to crisp value. This crisp value can be expressed as in Equation 3.32.

$$z = Defuzzify(C) \tag{3.32}$$

where z is the crisp output, C is the fuzzy set that represent the distribution of the results of the fuzzy inference as illustrated in Figure 3.17.

The most common means of defuzzification is called the centre of gravity method in which the centre of gravity of the fuzzy set is measured and projected to the z-axis to get the crisp result as illustrated by Figure 3.18. The output of this defuzzifier is a number z given by Equation 3.33.

$$z = \frac{\int z_i \mu_{zi}(z_i) dz}{\int \mu_{zi}(z_i) dz} \tag{3.33}$$

where z is the crisp output, μ_{zi} is the fuzzy membership value at z_i

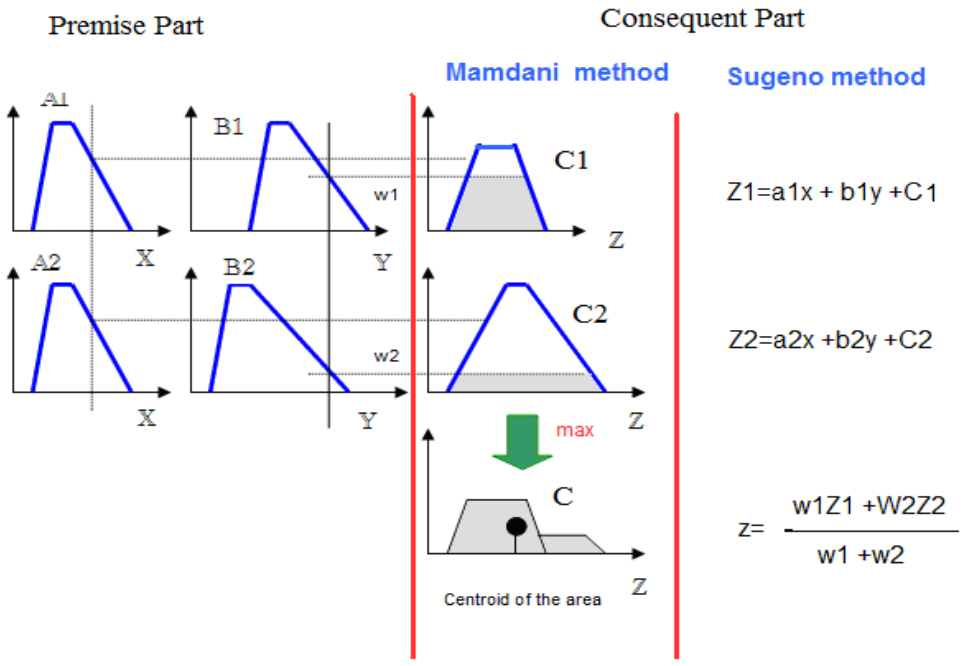


Figure 3.17 *Different types of fuzzy inference systems (modified from Jang & Sun, 1995).*

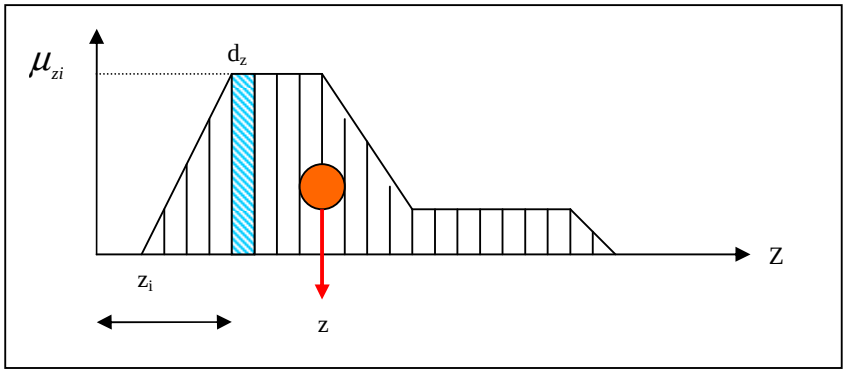


Figure 3.18 *Illustration of the defuzzification using the centre of the gravity*

3.4.3.2 Types of fuzzy logic systems

The operation of a FLS is based on the rules contained in the rule base. Fuzzy rules, or Fuzzy if -then rules, are defined as a conditional statement in the form presented in Equation 3.34.

$R_{(l)}$:**IF** x_1 is A_1 and x_2 is A_2 and ….and x_n is A_n **THEN** z is $B_{(l)}$ (3.34)

where

$R_{(l)}$ is the l rule number

$x_1, x_2, \ldots x_n$ are the input variables

$A_1, A_2, \ldots A_n$ are linguistic fuzzy membership functions in the premise part

$B_{(l)}$ is the membership function in the consequent part

As illustrated in Figure 3.17, there exist two basic forms of fuzzy rules that have been developed to date: Mamdani's fuzzy rules (Mamdani & Assilina ,1975) and Takagi-Sugeno-Kang's (TSK) fuzzy rules (Sugeno & Kang, 1988; Takagi & Sugeno, 1985). Both types of fuzzy rules are widely used in system modelling and control areas. The differences between these two types of fuzzy rules appear in the

consequence part of the rule. For the Mamdani fuzzy model, both antecedent and consequent are fuzzy propositions; however, in the Takagi-Sugeno fuzzy model, the antecedent is a fuzzy proposition but the consequent is a crisp function.

a. Mamdani fuzzy model

In this model, the inputs and outputs are partitioned into fuzzy regions. The size, the shape, and the parameters of these fuzzy regions are decided by experience and the fuzzy rules are generated based on human expert. This model uses the minimum operation as a fuzzy implication function, and then the output variables are combined together to form a single fuzzy subset for each output variable. This combination is constructed by taking maximum overall of the fuzzy subsets assigned to output variable by the inference rule. The purpose here is to aggregate all individual rule outputs to obtain overall system output. Finally, the defuzzification step is implemented as presented previously (Keshwani et al., 2008).

The disadvantages of the Mamdani model are that the number and the shape of the membership functions are difficult to determine. In addition, the number of fuzzy rules increases dramatically as the number of input variable increase. In such a strategy, if we have no a priori knowledge about the system, the structure of the model becomes a difficult task and we have to select the structure by a trial and error process.

b. Takagi and Sugeno model

Fuzzy modelling using this approach was proposed by Takagi and Sugeno (1985). The Takagi-Sugeno fuzzy system is a specific case of Mamdani fuzzy system in that the premise or antecedent of the fuzzy rules are defined with linguistic terms like very low, low, medium, high, very high but the consequence part is described by a non-fuzzy equation of the input variables (linear combination of the inputs),

instead of the fuzzy membership functions (see Ffigure 3.17). An example of the Takagi-Sugeno model is presented in Equation 3.35.

$R_{(l)}$: **IF** x_1 is A_1 and x_2 is A_2 and ….and x_n is A_n **THEN** $z_i = c_i + a_i x + b_i y$ (3.35)

where

$R_{(l)}$ is the l rule number

$x_1, x_2, \ldots x_n$ are the input variables

$A_1, A_2, \ldots A_n$ are linguistic fuzzy membership functions in the premise part

z_i is the crisp output of the rule l

a_i, b_i, c_i are constants

Takagi-Sugeno Model uses the weighted mean criterion to combine all local representations for the defuzzification (Equation 3.36).

$$z = \frac{\sum_{i=1}^{l} w_i z_i}{\sum w_i} \tag{3.36}$$

where w_i is the degree of fulfilment of the i^{th} rule and l is the number of rules in the rule base.

The performance of the fuzzy logic system depends on the rule base, the inference mechanisms and the defuzzification method and can be improved by tuning the membership functions. Normally, Mamdani's method has seen more widespread use because it is easy to be understood by human experts. However, the advantage of Takagi-Sugeno's method is that it has better computational efficiency, which makes it very good in modelling non-linear systems (Jang & Gulley, 1997; Jang &

Sun, 1995). Unfortunately, how to define the rules and membership functions requires a lot of prior knowledge (Ross, 2004). This is why the hybrid Neuro-Fuzzy system, introduced in the next section, becomes important.

3.5 Hybrid Neuro-Fuzzy Systems

The major difficulty of fuzzy models is how to acquire the knowledge for building the fuzzy rules and how to tune the parameters in the membership functions. The proper selection of the number, the type and the parameters of the fuzzy membership functions and rules is crucial for achieving the desired performance. Tuning the parameters is difficult and time consuming since there are many tuning parameters. In addition, attention must be taken to select the most important features that describe the system under study (Jang, 1993).

In the early applications of fuzzy logic models, the generation of the fuzzy rules and the adjustment of its membership functions were performed manually by trial and error and the best combination can be found by simulation test, which is a challenge (Ross, 2004). Subsequently, it has been found that the manual adjustments of membership functions sometimes lead to wrong conclusion (Jang, 1993). Hence, there is a need to formalise a systematic approach to generate fuzzy rules from an input-output data set.

To solve this problem, Jang (1993) provides a useful formalised method for tuning the parameters of fuzzy logic system (FLS) based on training data set of input values and their desired target outputs using the training algorithms of artificial neural networks. The method is called adaptive networks based on fuzzy inference system (ANFIS). The method takes advantages from both fuzzy logic systems and neural networks; it also avoids their individual shortcomings. For example neural network has an implicit knowledge representation; it is neither

easy to understand nor easy to explain its decision process. Fuzzy logic system is a subjective and heuristic system and it is time consuming to tune the fuzzy membership functions and develop the fuzzy rules. In other words, The ANFIS is fuzzy inference system implemented in the framework of adaptive network in order to map an input-output relation based on both human knowledge (in the form of fuzzy IF-THEN rules) and input-output data pairs, and hence obtain the best model from the given data.

To validate his approach, Jang (1993) provided multiple examples of the ANFIS and the results were reported to be comparable with traditional backpropagation neural networks. Jang (1993) also found that with the supervised learning capabilities of neural networks and the heuristic reasoning capability of fuzzy rules, the ANFIS model is able to learn a complex functional relations and at the same time to generate fuzzy rules. Jang et al. (1997) showed that the ANFIS has unlimited approximation power for matching any non-linear function arbitrarily well; thus the ANFIS can be considered as a universal approximation (Jang, et al., 1997).

However, the tests by Jang et al. (1997) were done using noiseless data sets that were generated by functional equations; so the application of ANFIS on noisy field data was not proven. For example, when ANFIS was applied by Miller (2006) to predict the rainfall (precipitation) from noisy weather data (temperature and humidity), it was found that the model predicted negative values of rainfall on some occasions. To solve this problem, Miller (2006) suggested the pre-processing of the data, e.g. by replacing the missing values and omitting the outliers, to improve the performance of ANFIS.

Hence, Fuzzy logic modelling is a useful nonlinear mapping of an input data vector into a scalar output, if the problem associated with the developing and

parameterising the membership functions can be tackled effectively. This is because no complex mathematical relationships are required in the construction of fuzzy logic applications. Besides, it is conceptually easy to understand, flexible and tolerant of imprecise data.

The usual design approach for a FLS is based on understanding the human expert approach and then implementing the strategy by direct translation of the linguistic rules and testing the developed FLS. The parameters can then be adjusted by trial and error and eventually it will yield good results, but offers no guarantee of optimality. The subjective strategy of such a system means that it cannot be readily applied. Therefore, there is a need for a more objective approach for converting heuristic rules, stated by a human operator, into an automatic strategy to tune the model parameters (Fuzzy Logic Toolbox for use with Matlab, 2004).

The basic idea behind neuro-fuzzy combination is to design a system that uses a fuzzy system to represent knowledge in an interpretable manner and have the learning ability of NN to adjust its membership functions and parameters in order to enhance the system performance. The main drawbacks of both individual systems could therefore be avoided, i.e., the black box behaviour of NNs, and the problem of selecting suitable membership values for FLSs (Jang, 1993).

Consequently, hybrid models of NN and FLS have been developed. The hybrid system can combine the advantages of two systems and avoid their drawbacks. This combination can constitute an interpretable model that is capable of learning, as NNs, and reasoning, as FLSs (Jange, 1993). Using this technique makes it possible to adjust the membership functions automatically from data by using NN learning algorithms.

In 1993, Jang and his colleagues started looking at the FLS as adaptive network (Adaptive Network Based on Fuzzy Inference System, ANFIS (Jang, 1993; Jang and Sun, 1995; Jang et al., 1997). This technique joins the linguistic interpretation of FLS with the computational power of neural networks that can be trained through gradient algorithms such as Back-propagation. A first foreword pass is performed to determine the network output and a second backward pass is performed to adjust the parameters for better approximation. They found that ANFIS could be easily implemented for a given input/output modelling technique. In the next sub-section, more detailed information about ANFIS is discussed.

3.5.1 Adaptive-Network-based Fuzzy Inference System (ANFIS)

Adaptive-Network-based Fuzzy Inference System (ANFIS) is a Sugeno-type (Sugeno & Kang ,1988; Takagi & Sugeno, 1985) fuzzy system in a five-layered network structure (Figure 3.19). The ANFIS has an interesting property: if the number of rules is not restricted, a Sugeno model can map any non-linear function (Jang, 1993). ANFIS is a multi-layer feedforward network in which each node performs a particular function on incoming signals. The parameters associated with these nodes are updated according to a given training data and a gradient based learning procedure in order to achieve a desired input-output mapping (Jang, 1993; Jang and Gulley, 1995, Jang et al, 1997; Chang and Change, 2006; Güngör, 2007).

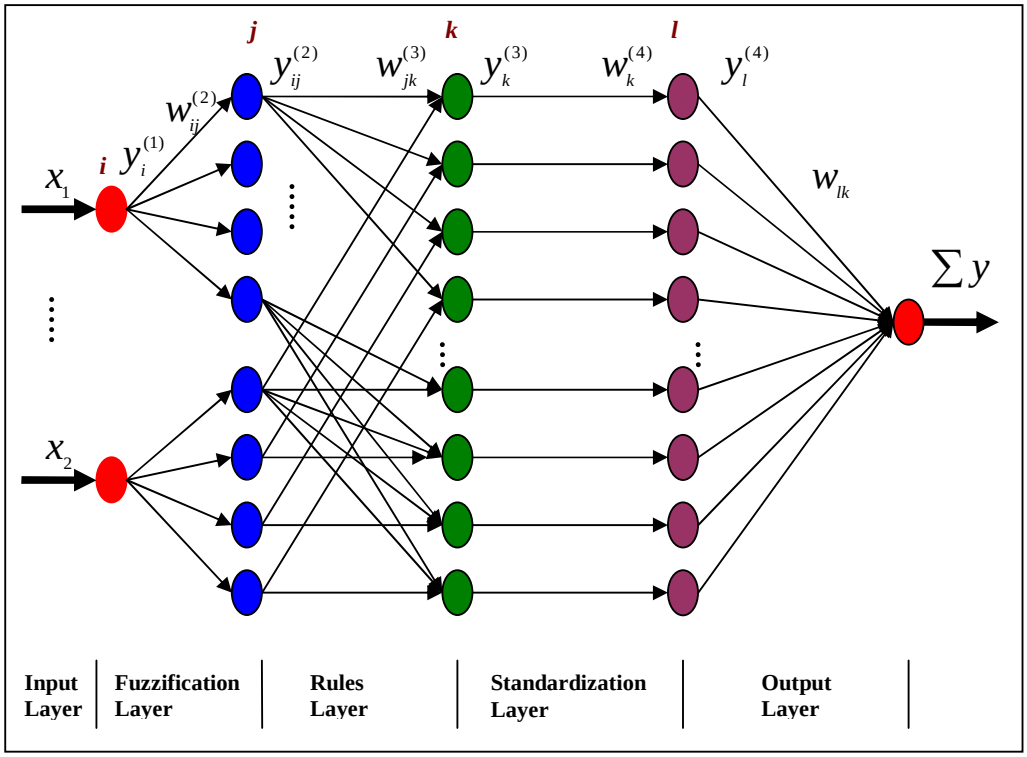


Figure 3.19 *Structure of fuzzy neural network (modified after Jang & Sun, 1993).*

ANFIS can be used to optimize membership functions and has the advantage of being able to construct fuzzy IF-THEN rules representing these optimized membership functions. The five layers in the ANFIS are: the input layer, the fuzzification layer, the rules layer, the standardization layer and the output layer shown in Figure 3.19.

1- ***The input layer*****:** The input layer contains the input variables to the model. The relation between the input and output of this layer is given by Equation 3.37.

$$y_i^{(1)} = x_i \tag{3.37}$$

where

x_i is the input to this layer

$y_i^{(1)}$ is the output from the input layer

2. The fuzzification layer: The aim of the fuzzification layer, as discussed previously, is to obtain the membership degree of each input, or fuzzifying the inputs. The data are processed in this layer and converted into linguistic variables by means of membership functions. The outputs of this layer are fuzzy membership degree of the inputs. The relation between input and output of this layer is as given by Equation 3.28. The parameters in Equation 3.28 can be used to define the region or the position of the fuzzy sets.

3- The rules layer: The inputs to this layer are the fuzzy membership degree output from the fuzzification layer. The output from this layer is also a membership degree and can be calculated from product fuzzy operation method for each rule using the formula in Equation 3.38.

$$y_k^{(3)} = y_{i,j(1)}^{(2)} \wedge y_{i,j(2)}^{(2)} \wedge ..... \wedge y_{i,j(n)}^{(2)} \tag{3.38}$$

where i is the input number, j is the number of membership function in the input i, k is the number of rules, and $\wedge$ denotes an AND operation . The value $y_k^{(3)}$ obtained is called firing strength of the rules.

4- The standardization layer: The main aim of this layer is to standardize the outputs from the third layer. In addition, the consequence parameters of the rule are determined in this layer. The output of this layer can be calculated as in Equation 3.39 and is called the normalized firing strength:

$$y_{lk}^{(4)} = w_{lk}(a_{0k} + a_{1k}x_1 + a_{2k}x_2 + ... + a_{ik}x_i) \tag{3.39}$$

where:

$$w_{lk} = \frac{y_k^{(3)}}{\sum_{k=1}^{k} y_k^{(3)}} \tag{3.40}$$

5- The output layer or Defuzzification layer: This layer computes the overall output as the summation of all incoming signals. There is only one node in this layer. The output can be calculated as in Equation 3.41

$$y = \sum_{k=1}^{l} y_{lk} \tag{3.41}$$

where l is the rule number

Therefore, there are two adaptive layers in the ANFIS architecture, the second, and the fourth layer. There are two-modified parameters in the second layer, a_{ij} which is the centre and b_{ij} which is the width of the membership function (See Equation 3.28). These parameters are related to the shape and positions of the membership function. In the forth layer, there are i+1 modifiable parameters, $a_{0k}, a_{1k}, a_{2k}, \cdots, a_{ik}$ which are related to the output of each rule.

Jang (1993) introduces two pass algorithms for adjusting the parameters using backpropagation optimization algorithms. In the forward pass, the premise parameters (a_{ij} and b_{ij}) are held fixed and the consequent parameters ($a_{0k}, a_{1k}, a_{2k}, \cdots, a_{ik}$) are adjusted by least square error (LSE). In the backward pass, the network error is back- propagated through the network and the premise parameters are adjusted by gradient descent while the consequent parameters are fixed. A training process of ANFIS is to tune all these parameters, so that the model can give a satisfying output.

However, although ANFIS is fast, gives excellent performance in modelling and has good generalization capability that make it popular and ideal for modelling complex problems. It does require strong computational power. The number of parameters to be estimated can be very large. For example, depending on the number of inputs and the number of membership functions for each input, and the shape of membership function chosen, the total number of parameters is estimated by Equation 3.42.

$$N_{total} = (N_{input} \times N_{mf} \times N_{pp}) + (l \times N_{cp}) \tag{3.42}$$

$$l = (N_{mf})^{N_{input}} \tag{3.43}$$

$$N_{cp} = N_{input} + 1 \tag{3.44}$$

where,

N_{total} is the total number of modified parameters, i.e. to be estimated;

N_{input} is the number of inputs;

N_{mf} is the number of membership functions associated with each input;

N_{pp} is the number of modified parameters per membership function, i.e. 2 in case of Gaussian membership function;

l is the number of rules;

N_{cp} is the number of modified parameters in the sequence part of each rule;

A comprehensive example of these equations is presented in Table 8.1.

3.6 Features Selection, Extraction, and visualisation

When performing analysis of complex data, one of the major problems stems from the number of variables involved. Therefore, collected data must be processed further in order to get better results. Reasons for doing this may be easier subsequent analyses, improved classification, or prediction performance through more stable representation, removal of redundant or irrelevant information or an attempt to discover underlying structure or knowledge by obtaining a graphical representation. That is because analysis with a large number of variables generally requires a large amount of memory and computation power. Therefore, the number of variables used in the model must be kept as small as possible. Hence, the objective of features extraction and visualisation is to represent the data in a reduced number of dimensions. Given a set of measurements, dimensionality reduction can be achieved in two different ways, Features selection and features extraction. These two methods can be considered as part of the data preparation phase or data transformation phase.

3.6.1 Features Selection

Features selection in the measurement space is to identify the variables that contribute to the modelling task and omit others. Features selection is important for model learning as it supports the dimensionality reduction that supports modelling strategies in which it selects the relevant variables stored in the plant's database. Using too many features is inefficient due to the curse of dimensionality. The curse of dimensionality refers to the fact that the number of data samples required to estimate some arbitrary multivariate distribution increases exponentially as the number of variables increases (Powell et al., 2007).

Algorithmic features selection is usually based on combinational optimization. The algorithm looks for the set of variables that minimize the cost function (the accuracy of the model) (Langley, 1994; Powell et al., 2007). Optimal feature selection requires an exhaustive search of all possible subsets of features. However, this requires a huge number of evaluations of the cost function, which is unfeasible with high-dimensional data set, hence, this is impractical. Instead of that, two common approaches are available for features selection. The first one is to start with minimum number of variables and then add one variable every time; the second one is to start with all variables and then gradually reduce the number eliminating. The popular method of evaluation of subsets can be done using the correlation between a candidate features and the desired output category (Guyon and Elisseeff, 2003; Powell, 2007). The last approach has been applied in this work as will be seen in Chapter 7 and 8.

3.6.2 Features Extraction

In general, the data set contains a collection of individual input sequences. So it is not feasible to estimate the model parameters for the whole data set, with its noise and outliers. Therefore, there is a need to extract a subset that is a sufficient representation of the entire data set but sufficient in size to be practical for model identification. Feature extraction is a general term for methods of constructing combinations of variables to get around the dimensionality problems while still describing the data with sufficient accuracy. These methods involve simplifying the amount of resources required to describe a large set of data through the transformation of the measurement space to a lower dimensional feature space that has higher manageable level. This transformation generates fewer, higher level of variables than the raw data itself for the modelling purpose. Usually, this process involves some form of aggregation in which variables which contribute

essentially the same information must be combined in a logical fashion in order to reduce the effective number of variables. It is found that by using this approach, the size of the feature sets can be effectively reduced and the accuracy of the classifiers can be increased (Torkkola , 2003; Wange et al., 2003).

Features extraction may be linear or nonlinear combination of the original (May and Jain, 1995; Laine, 2003). A well known linear features extraction technique is Principal Component Analysis (PCA) while Kohonen Self-Organising Map is probably the most well known non-linear technique (Kohonen, 1996). In the case of PCA method, the original features space is rotated before projecting the feature vector onto a limited amount of axes. Thus, principal component analysis (PCA) is commonly applied in features extraction phase to transform the feature vectors to orthogonal coordinate system, and to select those that are used in the model (Torkkola, 2003). PCA produces an orthogonal coordinate system in which the axes are ordered in terms of the amount of variance in the original data for which the corresponding principal components account. Dimension reduction is gained when only those axes that account for the most of the variation are selected. New features vectors are then gained by transforming the original feature vector onto this lower dimensionality space. Hence, PCA is a statistical modelling tool that attempts to extract relevant information from the data and defines a feature space of principal components that aims to capture the variance of the data.

There are several steps to find the PCA, these steps are (Jolliff, 2002; Keg et al., 2007):

1. Standardising the data: Sometimes, it makes sense to compute principal components for raw data. However, this is appropriate when all the variables are in the same units. Standardising the data is reasonable when the variables are in different units or when the variance of the different columns is substantial.

2. Calculating the covariance matrix.

3. Calculate the eigenvectors and eigenvalues of the covariance matrix.

4. Choosing components and forming a feature vector: once eigenvectors are found from the covariance matrix, the next step is to order them by eigenvalue, highest to lowest. This gives the components in order of significance. Now, the components of lesser significance can be ignored. In this case, some information is lost but if the eigenvalues are small, not too much information is lost. Hence, if some components are left out, the final data set will have fewer dimensions than the original. To be precise, if we originally have *n* dimensions in the data and we calculate *n* eigenvectors and eigenvalues, then we choose only the first desired values. What needs to be done now is to form a feature vector, which is matrix of the chosen eigenvectors. Therefore, the eigenvectors with the highest eigenvalues are the principle components of the data set.

$$feature \quad vector = (eig_1 \quad eig_2 \quad eig_3 \quad \dots \quad eig_n)$$

Hence, In the PCA, the directions are found which accounts for most of the variance in the data. This is done by calculating the eigenvectors e_1, e_2, …e_n and corresponding eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ of the covariance matrix of the data and ordering them by decreasing eigenvalues. The first direction e_1 accounts for most of the variance in the data, the second for the second largest amount, and so on. By projecting the data to the space spanned by the first two eigenvectors as much of the variance is preserved as possible. The sum of the corresponding eigenvalues gives the amount of variance preserved in the projection and thus indicates the error made in the low-dimensional projection.

5. Deriving the new data set: this is the final step in PCA, and is also the easiest. Once we have chosen the components (eigenvectors) that we wish to keep in our data and formed a feature vector, we simply take the transpose of the vector and multiply it on the left of the original data set transposed.

$$Final \quad data = (feature \quad vector)^T \times (Row \quad Data)^T$$

The first principal component is a single axis in space. When we project each observation on that axis, the resulting values form a new variable, the variance of this variable is the maximum among all possible choices of the first axis. The second principal component is another axis in space, perpendicular to the first. Projecting the observations on this axis generates another new variable. The variance of this variable is the maximum among all possible choices of this second axis. The full set of principal components is as large as the original set of variables. However, it is common place for the sum of the variances of the first few principal components to exceed 80% of the total variance of the original data (the mathworks, statistics toolbox). By examining plots of these few new variables, it is possible to develop a deeper understanding of the driving forces that generated the original data.

From the projection plot, the clusters in the data can be seen. However, relations between individual variables are lost since the new coordinates are complex linear compensations of the original variables. Hence, projection visualisations are only applicable for detecting of similar groups of the objects.

Nonetheless, PCA has many drawbacks (Jolliffe, 2002). For example, PCA reduces a linear combination of orthogonal variables, thus, PCA may not always be appropriate for data with highly nonlinear characteristics. Further PCs do not

necessarily have physical explanation. Each component is just a statistical construct representing variables in the data.

However, PCA method suffers from several limitations. Its algorithm depends on strong assumptions of the properties of the data such as the linearity and normal distribution. Therefore, the use of these algorithms requires the user to pre-process the data to fit these assumptions. Hence, it may fail to capture non-linear phenomena, or the algorithm may produce poor models. Furthermore, it is difficult to understand these components, or it is just understandable only to mathematically literate persons (May and Jain, 1995; Laine, 2003).

It should be noticed that a training data set has to contain a representative state of the system to be modelled using data driven techniques. Also, it is not of importance to repeat similar patterns in training data sets. Repeating patterns do not bring new particular learning benefit to give better performance but the training process is significantly slow. To solve this problem, the KSOM, which recently have become widely used tool for clustering multivariate data as they allow overcoming limitations of statistical methods and allowing the analysis of data containing complex non-linear relationships (Kohonen, 1996), can construct optimal code abstract feature space as seen previously. Individual feature values can be replaced by these codes, which result in data compression. Each code vector corresponds to and represents a part of the input space, the set of those points in the space that are closer in distance to that code vector than to any other code vector. Therefore, KSOM attempts to enhance prediction by amplifying a pattern and discarding noise. The accuracy of modelling, therefore, increases quickly as will be seen in Chapters 7 and 8.

3.7 Summary

In this chapter, the essential background for understanding the subsequent chapters is briefly introduced. The main context of the chapter is an overview of the basic of AI techniques used in this study, namely, backpropagation feedforward Artificial neural networks, Kohonen features map (Kohonen self organising map) and fuzzy logic. The hybrid modelling systems were also presented and discussed.

The Chapter discussed the learning ability of the artificial intelligence techniques to construct nonlinear relationships that can explain the complex relationship within the data without the difficult task of dealing with deterministic nonlinear mathematics. It also presented how these systems deal with the complexity and uncertainty of the system in a manner similar to the human way of thinking and reasoning. Furthermore, the Chapter presented the combination of fuzzy logic and neural networks that make them a versatile tool in modelling highly non-linear systems.

The next Chapter will present the methods and materials used in this research such as the source of the data, the methodology used for data pre-processing and the software tools used in this study.

CHAPTER 4

METHODS AND MATERIALS

4.1 Case study

The methodology of this research work was applied to data from Seafield wastewater treatment plant in Edinburgh, UK and further details about this plant and the data are given in the next section. Additionally, data from East Calder and New-bridge treatment works also in the Lothian region of Scotland were used for validation. The three plants are part of the Almond Valley and Seafield project, an environmental regenerating initiative by Scottish Water for Edinburgh city and Lothian regions. The three plants are operated by Thames Water under a private finance initiative. Treated wastewater from East Calder and New-bridge is discharged to the Almond River as shown in Figure 4.1. The population equivalents (pe) of all the treatment plants in the Edinburgh area are given in Table 4.1 (Hill and Hare, 1999).

The River Almond is designated as “sensitive area” under the terms of the urban wastewater treatment regulations (UWWTR) (Hill and Hare, 1999). This means that the discharges from its works would have to meet new more stringent standards and water quality objectives set by the Scottish Environmental Protection Agency (SEPA) in order to reduce the eutrophication phenomena. For example, the maximum discharged BOD and COD are 25 mg/l and 125 mg/l respectively; while Suspended Solids (SS) is not reported except where the works only provides primary treatment as consents vary from one works to another (SEPA, 2006).

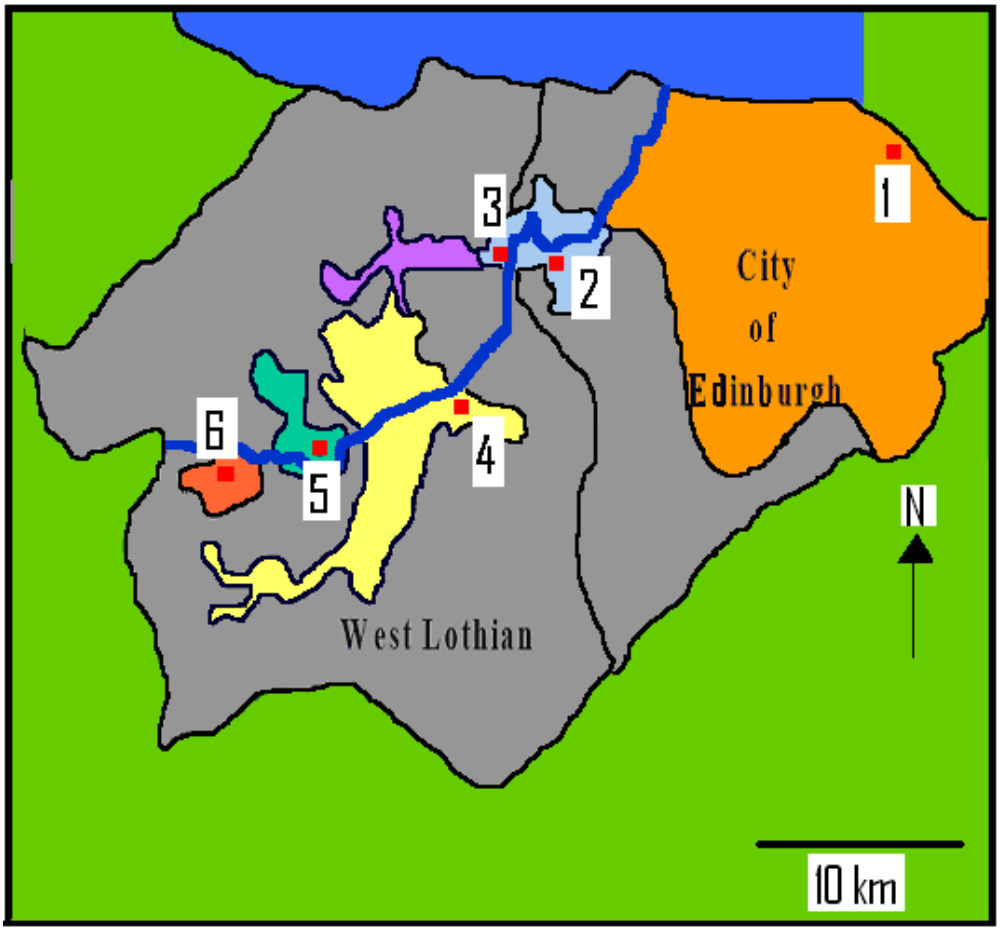


Figure 4.1 *The Almond Valley and Seafield, Edinburgh, Catchment Area. 1 is Seafield WWTP, 2 is Newbridge WWTP, 3 Broxburn WWTP, 4 East Calder WWTP, 5 Blackburn WWTP, 6 Whitburn WWTP.*

Table 4.1 *Population Served in the Almond valley and Seafield project*

Treatment plants	Population Served		Type of treatment (Secondary/tertiary)
	1999	2023 projection	
Seafield (Edinburgh)	480000	520000	Activated sludge
East Calder	65000	75000	Activated sludge, Bio-tower and sand filtration
New-Bridge	19500	22500	Activated sludge and sand filtration
Whit-burn*	11200	13000	Activated sludge
Blackburn*	11500	12700	Trickling filter

* These works do not form part of this study

Catchments served by the Almond valley and Seafield works contain both separate and combined sewerage systems with a number of combined storm overflows discharged to local watercourses. Each of the works has primary and secondary treatment, with some have additional tertiary treatment (see Table 4,1). All have storm tanks which come into operation at about 3 dry weather flows and discharge to the river when full.

The East Calder plant employs a conventional activated sludge for secondary treatment and a biotower for nitrification. Final polishing of the effluent through a system of sand filters takes place prior to discharge. The much newer Newbridge plant uses a combined nitrification-denitrification activated sludge system for its biological oxidation stage. Additionally, prior phosphorus chemical precipitation takes place at inlet. The final effluent is also passed through a system of rapid sand filters prior to being discharged.

The Seafield wastewater treatment plant is located in the eastern part of Edinburgh city. It receives water from Edinburgh catchment and adjacent contributions. The sewered catchment comprises many different sources of wastewater such as domestic effluent, industrial discharge and rainwater seepage. The outfall is situated adjacent to Portabello beach, designated a bathing beach, and the proposed multimillion pound housing, leisure, business and continental ferry development at Leith Docks.

The Seafield treatment plant relies on a conventional activated sludge secondary treatment system. It aims to reduce the BOD and COD of the flow to such a level that there is no risk to the ecology of the receiving water body. The works discharges treated effluent through a long sea outfall to a point some 2.8 km offshore. Storm water is also discharged through the long sea outfall after 6mm screening and retention in 4 no 10,000 m^3 tanks.

The Seafield plant comprises 8 circular sedimentation tanks, 4 rectangular non-nitrifying aeration lanes, and 8 circular final settlement tanks. The main treatment is preceded by six screens (spacing: 6 mm) and four Detritor grit removal units. An overview of Seafield wastewater treatment plant is shown in Figure 4.2. Table 4.2 presents the design characteristics of the Seafield treatment plant.

Table 4.2 *The design characteristics of Seafield treatment works*

Process Unit		
Storm Water Tanks	Existing number	4 No. Rectangular
	Length	100 m
	width	30 m
	Depth	3.4 m
	Volume	10,200 m^3
	Square Area	3,000 m^2
Screening	Existing number	6
	Type	6 mm screens
Grit Removal	Existing number	4
	Diameter	15 m
	Depth	2.3 m
	Volume	1,626 m^3
	Square Area	2826 m^2
Primary sedimentation	Existing number	8 No. Circular
	Diameter	55.0 m
	Depth	3.9 m
	Volume	9261 m^3
	Square Area	2375 m^2
Aeration Lanes	Existing Number	4
	Length	70 m
	Width/lane	21 m
	Depth	6.1 m
Final Settlement Tanks	Existing number	8 No. Circular
	Diameter	45.0 m
	Depth	2.5 m
	Floor Slope	5 ^{o}
	Volume	3974 m^3
	Square Area	1590 m^2

Figure 4.2 *a. Areal photograph of Seafield wastewater treatment plant*

Figure 4.2 b *Map of layout of Seafield wastewater treatment plant, 1 is the screen house, 2 are the detritors, 3 Grit washing mechanism, 4 sedimentation tanks, 5 storm tanks, 6 aeration tanks, 7 final settling tanks, 8 UV treatment unit, 9 outfall tunnel*

The average final effluent BOD concentration was around 10 mg/l with respect to the provided example data record, which is considerably much lower than the design consent of 25 mg/l was taken from an internal report commissioned by the treatment plant operator.

4.2 Data

4.2.1 Data collection system

The Seafield wastewater treatment plant is equipped with data collecting systems. These systems are used for monitoring, automatic control, and as a decision base for operational strategies. The data are passed to the data acquisition unit (SCADA), located in the main control room, which store the data as databases. SCADA software allows set points to be altered and to communicate with the user via friendly interface screens and associated alarms. The system can alter the set points for the controlling parameters, e.g. desludging time, dissolved oxygen (DO), Return and waste activated sludge (RAS), (WAS).

4.2.2 Data Description

Historical daily database describing the operation of the Seafield activated sludge treatment plant in Edinburgh (Scotland, UK) for a period of approximately three years with a total of 1066 data vectors were obtained from Thames Water (plant operator). These data come from different sources. On-line data are gained directly from sensors and these include flow, temperature, and pH. Off-line data or manual samples are derived variables involving several intermediate steps before being presented in the record sheet. These variables include Specific Sludge Volume Index (SSVI), Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Suspended Solids (SS), and Ammonia Nitrogen (NH4). The third category of data is those calculated based on a combination of the above

measurements and include BOD-load, food to micro-organisms ratio (F/M), and Sludge Age. The performance of the three treatment plants is monitored on a routine basis by Thames Water staff and in some occasions by Scottish Water staff. Composite samples are collected at 9:30 and consist of "shots" taken over the previous 24 hours. These samples are collected from the works influent, primary clarifier effluent and secondary effluent streams. The frequency with which samples are analysed varies. The parameters measured are discussed in the next supsections.

4.2.2.1 Biochemical Oxygen Demand (BOD)

The BOD test seeks to measure the biochemical oxygen demand exerted by the sample over a fixed period. It is therefore apparent that the oxygen uptake in the BOD bottle may not necessarily reflect the oxygen requirements of the wastewater treatment process having to treat the waste. The BOD_5 is often quoted as being between 60% and 70% of the ultimate BOD (Orhon and Artan 1994; Spellman, 2003). The oxygen demand is an extremely important measurement of wastewater quality as it measures the potential for oxygen depletion in the water, and therefore is an important indicator of organic pollution. It is necessary to assess the overall efficiency of treatment processes as this has a direct bearing on the quality of the final effluent and the economy of the process. Considering the time for the analysis the BOD test is certainly not suitable for operation/control purposes. Furthermore, BOD is not a single point value but is time dependent, also is not precise measurement and the reproducibility is quite poor.

4.2.2.2 Chemical Oxygen Demand (COD)

COD is widely used to characterise the organic strength of wastewater. The test measures the amount of oxygen required for chemical oxidation of organic matter in the sample to CO_2 and H_2O. In the COD test, biodegradable as well as non-biodegradable material is oxidised; however, cannot distinguish between

biodegradable and inert organic matter. There is no uniform relationship between the COD and BOD of wastewater except that the COD value must be greater than BOD (Olsson and Newell, 1999). An empirical correlation of COD to BOD for a particular wastewater can be determined which is useful as this method takes only hours. The COD/BOD or BOD/COD ratio provides an estimate of the proportion of biodegradable organic matter present in wastewater (Rustum et al., 2008-a).

4.2.2.3 Dissolved Oxygen (DO)

DO is one of the most important and useful measurement in activated sludge processes and is also the basis for the BOD and Oxygen uptake rate (OUR) tests. DO is measured on-line in the plant using DO probes. The probes have also a sensor for measuring temperature. Aeration is controlled by maintaining a set-point for the DO concentration in the aeration tanks. A sufficient supply of oxygen is important for the removal of carbonaceous materials and essential of the nitrification process. Low oxygen concentration may not only inhibit nitrification, but may also deteriorate sludge settleability, worsen effluent quality and results in predominance of filamentous bacteria (Chen et al, 1993; Spellman, 2003). On the other hand, excessive oxygen supply results in high operational costs. Furthermore, high oxygen supply may lead to excessive nitrification and again to poorly settling sludge. DO concentrations in the aeration tanks are maintained at about 1.5 to 4 mg/l; 2 mg/l is used as a set point. Values above 4 mg/l do not improve operations significantly, but increase the aeration costs considerably. Aeration accounts typically for more than 50% of the total plant energy requirements (Gray, 2004; Spellman, 2003).

4.2.2.4 Mixed Liquor Suspended Solids (MLSS and RAS-MLSS)

The concentration of suspended solids in the mixed liquor (MLSS) and returned activated sludge (MLSS-RAS) are investigated on daily basis. A single discrete sample is collected and sent for analysis each day. The level of MLSS is

maintained by recycling the settled sludge (RAS) at a level sufficient to treat the incoming organic load, but there is a lack of reliable on-line information on how to control the RAS. Generally, increasing the MLSS produces an older, denser sludge, while decreasing MLSS produces a younger, less dense sludge. The mixed liquor suspended solids concentration is controlled by manual adjustment of wastage rates to achieve a value of approximately (2500 mg/l, special communication with the staff). This corresponds to a sludge age of between 3.3-7.5 days.

4.2.2.5 Quantification of settling properties (SSVI)

Although there are now many automated system for quantification of settling properties, such as installing a measuring system that tracks the sludge blanket or concentration profiles in the full scale clarifier, or another optical systems that reveal the relation between sludge flocs structure and the settling properties (Vanrolleghem et al., 1996; Olsson and Newell, 1999); the Seafield plant still relies on the traditional way of quantifying sludge settleability by measuring the Stirred Specific Sludge Volume Index test (SSVI). This test is performed on samples of mixed liquor in order to have an idea about the settleability of the activated sludge.

4.2.2.6 Flow

Monitoring of flow in wastewater treatment plant is important for the determination of mass balance. Therefore, flow rates are measured at three points in the Seafield treatment works:

a. Flow to full treatment

Flow to the treatment plant is restricted to a maximum of three times average dry-weather flow. Limitations of flow help to prevent extreme hydraulic shocks that would result in further deterioration of the treatment plant performance.

b. Return activated sludge (RAS)

The setting of the return activated sludge rate is important for solids control in the activated sludge process. it is because it maintains a constant MLSS concentration that constitutes one of the most common methods of solids control, resulting in an equalised and improved effluent quality. The RAS rate, necessary to maintain a constant target MLSS concentration, can be calculated from a mass balance. In the Seafield wastewater treatment plant, the return sludge system includes pumps, a timer or variable speed drive to regulate pump delivery and a flow measurements device to determine actual flow rates. Generally, the RAS is related to the treatment plant inflow rate by a constant factor between 0.4 and 1.5, thus the sludge rate is increased with increasing flow (Harremoës et al., 1993; Spellman, 2003). However, this control principle may lead to deterioration of the plant performance by imposing an additional hydraulic load to the clarifier as well as by a hydraulic shock to the thickener caused by sharp changes in the RAS rate (Spellman, 2003).

c. Waste or surplus activated sludge (WAS)

The WAS rate determines the rate at which sludge is removed from the activated sludge system. If a constant mean cell residence time (sludge age) and a constant MLSS are to be maintained, the wastage rate can be calculated from a simple mass balance involving the target MLSS concentration in the aeration tank and the MLSS concentration in the effluent and in the RAS (Tchobanoglous and Burton, 1991; Spellman, 2003). In the Seafield wastewater treatment plant, the WAS withdrawal is accomplished by adjusting valves on the return system.

4.2.2.7 Sludge Age

The length of time, or mean cell residence time, that the biomass remains in the secondary treatment stage is known as the sludge age. The conventional sludge age in UK is around 5 days and Seafield is not an exception with the sludge age, in the available data ranging between 3.35 and 7.52 days (Lessard and Beck, 1993).

4.2.2.8 Food to Micro-organisms ratio (F/M)

In the UK, traditional practice has used the F/M ratio to control the activated sludge process, in units of Kg BOD5/Kg MLSS (or sludge loading rate). The food to micro-organisms ratio for Seafield treatment plant was between 0.08 and 0.23 kg BOD_5 applied per Kg MLSS day.

Other water quality parameters such as pH and Temperature are also measured and recorded.

4.2.3 Data treatment

A major problem was the quality of the data from the full-scale wastewater treatment plant. For any modelling strategy, the quality of the outputs heavily depends on the quality of the inputs. Therefore, data treatment provides techniques on how measurement data can be validated and how the quality of data can be improved. This is important to obtain reliable analysis results.

Different methods for exploring the available data were applied. Typically, descriptive statistics such as the mean, maximum, minimum were calculated. In addition, histograms were plotted. All measurements were examined with respect to erroneous values, missing data, possible outliers (See Rustum and Adeloye, 2007). Missing values and outliers were located and replaced by “NAN”(Not a Number) in order to cope with the MATLAB requirements, then they were

replaced as will be discussed in Chapter 5 (see also Rustum and Adeloye, 2007-a). In addition, data were manipulated in order to cope with the modelling strategy, e.g. the data were standardized before being fed to the ANN as described in Chapter 3.

Erroneous Measurements were filtered out from the data using a priori knowledge of the process. Statistical characteristics of the measured process variables are described in the more detail in Chapter 5. An important feature of the data is the large number of missing values, which would have made modelling by other methods such as regression analysis impractical.

4.4 Computer Software

The developed models were implemented using MATLAB 7 programming language with Neural Networks and Fuzzy Logic toolboxes (Version 7, Release 14, Mathworks, Inc). Kohonen Self Organizing Maps were built and visualized using SOM Toolbox for MATLAB 5, developed at the Laboratory of Computer and Information Science (CIS) at Helsinki University of Technology. Supporting statistical analysis was conducted using Statistical Toolbox and various functions in MATLAB.

The MATLAB programming language was chosen for model development because NN and ANFIS require intensive matrix computations. The Fuzzy Logic and Neural Network Toolboxes of MATLAB provide comprehensive support for design, implementing, and simulation of the models rapidly. Their consistent methodology and modular organization provide a flexible framework for experimentation, and simplify customization. The work was preformed using the available advices in the documentation of the software's package and in the literature.

4.5 Selecting model structure

In this work, an ANN and ANFIS will be adopted as the framework of the model structure, in addition to the KSOM for data analysis and features extraction. When developing the models there are difficulties that can be encountered. The task of choosing the appropriate model parameters, i.e. number of inputs and outputs that the model has to take into consideration, the number of neurons in the hidden layer in case of ANN and the shape of membership functions in case of ANFIS. Unfortunately, this is an empirical exercise that has to be carried out using trial and error approach until satisfactory results are obtained. That is due to the little information available to guide the user in the selection of particular model. Therefore, there is a need to evaluate those models and assess their predictive capabilities.

4.5.1 KSOM modelling Strategy

The topology of the KSOM grid was chosen to be the plain two –dimensional discreet map, the rational being that it is easy to visualise and to interpret. The topology of the grid was hexagonal, which is more homogeneous with respect to the directions on the KSOM plane than other frequently used alternative, rectangular, though, according to Nikkila (2005), the results can be very similar with both choices. The procedures for map training and visualisation are followed as presented in Chapter 3.

4.5.2 Artificial Neural Network modelling strategy

The neural networks models developed in this study were trained in MATLAB programming language using neural network graphical user interface according to the following steps:

1. Data Pre-processing: the data are normalized using the mean and standard deviation in order to improve the performance of the model.

2. Creating a feed foreword back propagation network.

3. Choosing Training and Learning functions: training and leaning functions are mathematical procedures used to automatically adjust the network weights and biases. The MATLAB includes several training function but the Levenberg-Marguard back propagation (*Trainlm*) was used.

4. Choosing the performance function to calculate and monitor network efficiency during training. MSE is used to measure the network error in this study.

5. Selecting transfer functions: transfer functions transform the output of each network layer according to a desired linear or nonlinear mapping. Tan-sigmoid, the most commonly used transfer unction, is used in the hidden layer and linear (*purline*) is used in the output layer.

6. Post-processing: this can be achieved by examining the predictive power of the model with testing data set that has not been used during training.

7. Presenting the results of training, validation and testing in figures and tables.

4.5.3 ANFIS modelling strategy

The ANFIS models were developed using the graphical user interface of the Fuzzy logic Toolbox according the following steps:

1. Loading data (training, testing, and checking) ;

2. Generating an initial FIS model ;

3. Choosing the FIS model parameter optimization method: backpropagation or a mixture of backpropagation and least squares (hybrid method). In this work the second approach is applied;

4. Choosing the number of training epochs and the training error tolerance;

5. Training the FLS model, this training adjusts the membership function parameters and plots the training (and/or checking data) error plot(s) in the plot region.

4.6 Models performance evaluation criteria

Once a model structure has been chosen and the network trained, the selected model needs to be evaluated. In practice, the accuracy of a model is determined by the 'goodness of fit' between outputs of the model and the system given the same input. Hence, some validation tests need to be considered. Generally, the accuracy of a model must be evaluated for three sets of data samples. These data sets are: training data that express the effectiveness of learning, validation data set that used to save the model from overfitting problem, and the testing data set that measure the generalisation capability of the network. There is a need to point out that the testing data set should ideally not have previously been presented to the network and it must represent the entire operation range. In this work, the following evaluation criteria have been considered.

1. The average absolute error (AAE) measures the mean error of the predictions.

$$AAE = \frac{1}{N}\sum_{i=1}^{N} \left(\left|x_i - x_i^{'}\right|\right) \tag{4.1}$$

2. The relative average absolute error (RAAE) is the AAE scaled by the range (maximum subtracted by the minimum) of the data. For a well-performing model, the RAAE should be as small as possible.

$$RAAE = \frac{AAE}{Max - Min} \tag{4.2}$$

3. The normalized root mean squared error (NRMSE) measures the mean root squared error scaled by the standard deviation of the values.

$$NRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - x_i^{'})^2}}{S} \tag{4.3}$$

4. The mean square error (MSE) which is defined as:

$$MSE = \frac{\sum(x_i - x_i^{'})^2}{N} \tag{4.4}$$

5. The correlation coefficient (R) measures the similarity of the shapes of the original and predicted time series and ranges between -1 and 1; the absolute value of the correlation coefficient for perfect predictions is unity.

$$R = \frac{N\sum x_i x_i^{'} - \sum x_i \sum x_i^{'}}{\sqrt{[N\sum x_i^2 - (\sum x_i)^2][N\sum x_i^{'2} - (\sum x_i^{'})^2}} \tag{4.5}$$

6. The classification error (CE) measures the fraction of over- and under-predictions for a variable $CE_i = x_i - x_i^{'}$. For purely random residuals, one would expect a 50:50 split between PR and NR. Otherwise, the model could be considered to be either upward or downward biased.

a. The positive rate (PR) is the fraction of over-predictions in the set of predicted values.

$$PR = \frac{count \quad CE_i \quad (if \quad CE_i \quad < \quad 0)}{N} \times 100\% \tag{4.6}$$

b. The negative rate (NR) is the fraction of under-predictions in the set of predicted values .

$$PR = \frac{count \quad CE_i \quad (if \quad CE_i \quad > \quad 0)}{N} \times 100\% \tag{4.7}$$

where

N is the number of samples;

x_i is the actual value;

$x_i^{'}$ is the value predicted by the model.

Max and Min are the maximum and minimum data points of the range.

$CE_i = x_i - x_i^{'}$

S is the standard deviation given by

$$S = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}} \tag{4.8}$$

$\bar{x}$ is the sample mean given by

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} \tag{4.9}$$

4.7 Visualizing the performance of the model

Although the evaluation criteria are very useful, probably the most valuable assessment of the model can be made using simple plots. These plots will compare the predictions made by the model with the actual data. It can give an indication of under and over-fitting data and will illustrate the model performance during training, validation and testing data sets. Several visualization techniques have been used in this work such as time series plots of the predicted and observed, Error distribution plot, scatter plots, etc.

4.8 Summary

This Chapter presents the methodology applied in this study. It starts with a thorough description of the case study used in work, which are three treatment works in the Almond River in Edinburgh, UK. Then a description of the data used and its characteristics is presented. An overview of computer softwares used in this study is also presented briefly. The Chapter ends with the performance evaluation criteria used to asses the developed modes.

Thus the previous chapters have covered all the basis of the methodology, the data and the various assumptions required to understand the analysis presented in the next applications. Each of the next chapters represents a separate application of the previously described tools to the case study data and treatment works.

CHAPTER 5

APPLICATION I: ACTIVATED SLUDGE DATA PRE-PROCESSING USING THE KOHONEN SELF ORGANISING MAP

5.1 Introduction

Modelling the activated sludge wastewater treatment plant plays an important role in improving its performance. However, there are many limitations of the available data for model identification, calibration, and verification, such as the presence of missing values and outliers. Although, many treatment plants are equipped with properly designed data collection systems, there is often little or no attention paid to the quality of the data (Rosen, 1998). Thus, situations abound for data corruption, such as excessive disturbances, equipment malfunction, and human errors. These lead to some of the problems in the raw plant data such as noise, missing values, and outliers. The resultant effects of these are discontinuities or gaps in the data records and outliers, both of which create severe handicap in modelling and identification of the process. One obvious solution to the problem is to remove records containing the missing values and outliers; however, given the shortness of the available data and the time and expense for their collection, such a luxury cannot be afforded. So a considerable pre-processing of the data is required, both to fill the missing gaps and to replace the outliers with more plausible values.

The data obtained from all the case study wastewater treatment works are not exempt from missing values and outliers. Thus, the first in the series of applications of the novel tools developed in this study involved use of the KSOM

to replace outliers and missing values from the high dimensional data set for the Seafield treatment works.

5.2 Outliers and Missing Values

5.2.1 Outliers

An outlier, a sample value that differs notably from the mean of the measurement series, can be caused by many factors such as electromagnetic interference, hostile measurement environment, defective installation, insufficient maintenance, or erroneous handling of the measurement system and intentional cover-up for lapses of the technician. A problem in detecting outliers is to decide whether they represent a true value or whether they are false due to disturbances in the measurement system.

Detection of outliers can be accomplished by using redundant sensors (Barnett and Lewis, 1994). In the redundant sensors, at least two sensors (or measurements) are used and an outlier is indicated when the sensors or the analysed samples do not deliver the same value (within a reasonable margin). However, this is an expensive procedure because it requires a large number of sensors or samples. In the activated sludge process, in which the wastewater is normally treated in parallel lanes, it is possible to use measurements from another lane to make such a validity check. The conditions in the various lanes are rarely exactly the same but if the configurations do not differ significantly, one could use the information from one lane to validate the other. However, in this case, there are no parallel measurements to validate the outliers. In addition, even where this is the case, it will be necessary to combine such evidence with a more formal one for detecting outliers such as the statistical *Z-score* and *modified Z-score* (Fallon and Spada, www.ewr.cee.vt.edu).

In the *Z-score* test, assuming that the data have a normal distribution, the mean and standard deviation of the entire data set are used to obtain a Z-score for each data point as in Equation 5.1.

$$z_i = \frac{(x_i - \bar{x})}{S} \quad ; \text{i= 1,2,……, N} \tag{5.1}$$

where S is the sample standard deviation and $\bar{x}$ is the sample mean given by Equations 4.8 and 4.9 respectively and N is the sample size. A test heuristic is that an observation with a *Z-score* greater than 3 should be labelled as an outlier.

The problem with the *Z-score* approach is that its estimate depends on both the sample mean and standard deviation, both of which are affected by the outliers, particularly if N is small, i.e. below 20 (McBean and Rovers, 1998). To overcome this problem, a modified *Z-score* test based on outlier-resistant estimators, such as the median of absolute deviation about the mean (MAD), is used in place of the standard deviation to compute the *Z-score* in Equation 5.1, where MAD is expressed by Equation (5.2) (Fallon and Spada, www.ewr.cee.vt.edu, last visit 12/1/2008). The condition for labelling an observation as an outlier using the modified z-score is 3.5 < Z < -3.5.

$$MAD = median(\left| x_i - \bar{x} \right|) \tag{5.2}$$

However, the above statistical techniques are not robust enough for labelling outliers since they depend on a number of assumptions, notably that a sample is normally distributed. Also, as illustrated for the *Z-score* method, the test statistic depends on parameters such as mean, median and standard deviation, which can be significantly affected by the outliers being detected. Therefore, a straightforward and practical method for the off-line detection of outliers is to

manually label the outliers by examining the time series plot. In time series, the human eye has a remarkable ability to pick out outliers with good result and by a careful investigation of the time series, manual detection of outliers can be as good as any more formalised method. Manual outlier detection from time series plots is certainly preferable when preparing data for model identification or training because it gives the model builder a sense for the data and also what can be expected from the model (Rosen, 1998). Consequently, the manual labelling approach was applied in the study to identify the outliers, although comparisons with the results of the *Z-score* and the *modified Z-score* approaches will be presented.

5.2.2 Missing Values

A missing value is caused by a sensor that does not deliver a measurement value, or by a fault in the measurement tools, or even by human mistakes. Depending on the measurement equipment, missing values can appear in the record as, for instance, blanks, zeros or negative values for entities limited to positive values. Therefore, missing values are often simple to detect in a data record.

Missing values are a serious problem as they distort the dynamic properties of the signal. Where outliers have been correctly identified, they too must be removed and replaced with more plausible values, thus exacerbating the sparseness of the data record. In order to perform a dynamic analysis, all missing values must be estimated, as failure to estimate them makes the complete sample difficult to be used. They may lead to severe problems in model identification process, particularly when tools such as Artificial Neural Networks and ANFIS are used. Since these tools require long periods of good and reliable data, it is important that the number of missing values be kept at a minimum.

5.3 Replacing outliers and missing values

If there are relatively few missing points, there are some models which can be used to estimate values to complete the series, such as replacing missing values with the mean or median of the data. Also interpolation and extrapolation may be applied. In interpolation, a curve is projected between known data points to infer the value at points between. In extrapolation the value of unknown data points are predicted by projecting a function beyond the range of known data points. A more elaborate approach is to express any variable which is missing in value, in terms of regression over the other variables using the available data, and then to use the regression function to fill the missing values ((MacDonald and Zucchini, 1997; Harvey, 1989; Bishop, 1995). Regression will work best if the number of water quality parameters having missing values in their records is small; otherwise developing different predictive regression equations for a large number of water quality parameters will be time wasting. Back propagation artificial neural networks modelling (ANNs) described in Chapter 3 offers a solution for multi-variable prediction but the performance of ANNs tends to decrease rapidly as the number of output variables increases, particularly when the output variables are not highly correlated (Adeloye and De Munari, 2006). Moreover, ANN models are affected by the missing values in the input space.

Far from actually removing identified outliers from a data set, their influence on estimates of summary statistics of the data can be tempered or even completely eliminated through the use of trimmed means, other scale estimators apart from standard deviation (e.g. MAD) and Winsorization (McBean and Rovers, 1998). In calculations of a trimmed mean, a fixed percentage of data is dropped from each end of an ordered data, thus eliminating the outliers. The mean is then calculated using the remaining data. Windsorization involves accommodating an outlier by replacing it with the next highest or next smallest value as appropriate.

However, using these types of models to predict missing values or outliers in a long time series is difficult and often unreliable, particularly if the number of values to be in-filled is relatively high in comparison with total record length. The accuracy of the estimate depends on how good and representative the model is and how long the period of missing values extends (Rosen and Lennox, 2001).

The activated sludge treatment plant is a dynamic process, so any variable is dependent, not just on the historical time series of the same variable but also on several other variables or parameters of the process. In other words, the problem is an exercise in multivariate analysis rather than the univariate approach of most of the traditional methods of estimating missing values and outliers; a multivariate model will therefore be more representative than a univariate one for predicting missing values. The KSOM offers a simple and robust multivariate model for data analysis, thus providing good possibilities to estimate missing values, taking into account its relationship or correlation with other pertinent variables in the data record. In comparison to other data-driven modelling paradigms such as multi-layer perceptron artificial neural networks (MLP ANNs) and classical multivariate regression analysis, the KSOM is not hindered by missing values. Moreover, time sequences of data is not a problem when compared to classical time series analysis (Vesanto et al., 2000).

5.4 Overview of Predicting Missing values using KSOM

The calibration of traditional predictive models with data is a supervised learning problem, because there is an output X_{out} which needs to be constructed from the input X_{in}. Multi-layer perceptron artificial neural networks (MLP-ANN) use this approach as described in Chapter 3. The KSOM algorithm, also described in detail in chapter 3, is designed for unsupervised learning, in that there is no “teacher” as such in the process that compares the desired output with the model output. In

other words, the difference between the KSOM and the usual supervised approach such as the MLP-ANN is as follows (see Figure 5.1): In the multi-layers perceptron neural network (MLP), the vector X_{in} (t) is presented to the network input, while the X_{out}(t) is used at the network output to compute explicitly an error signal that guides learning. However, the KSOM learns to associate or correlate the inputs and outputs of the mapping without explicit computation of an error signal (Barreto and Araujo, 2004), as it just finds the similarities between the input vectors and the KSOM neurons or weight vectors to determine best matching units, BMU.

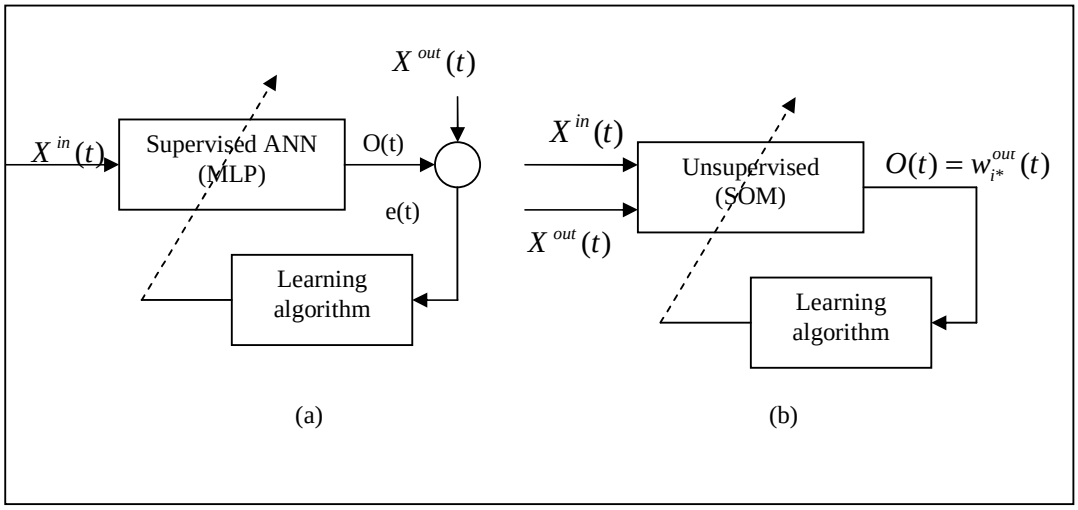


Figure 5.1 *Differences between (a) supervised and (b) KSOM identification*

Consequently, the KSOM can be easily used as a lookup-model. Taking any set of known values (the input variables or data sample), their BMUs from the KSOM can be determined and the corresponding prototype vector can be used to give values to the other variables which are missing in the input vector (Alhoniemi et al., 1997). In a sense, this is like predicting missing values using linear regression albeit in a multivariate platform. The general regression of X_{out}(t) on X_{in}(t) is the expectation of the output given the input and hence the calibrated regression

model can be used to predict $X_{out}(t)$ if the $X_{in}(t)$ is known. In the same vein, since the codebook vectors of the KSOM represent the local averages of the input vector, the map can be used for the prediction of missing components of an incomplete input vector. This is achieved by searching for the best matching unit (BMU) using the known vector components of the input vector and the output then gives an approximation of the unknown components of the vectors as shown in Figure 5.2 (Obu-Cann et al., 2001).

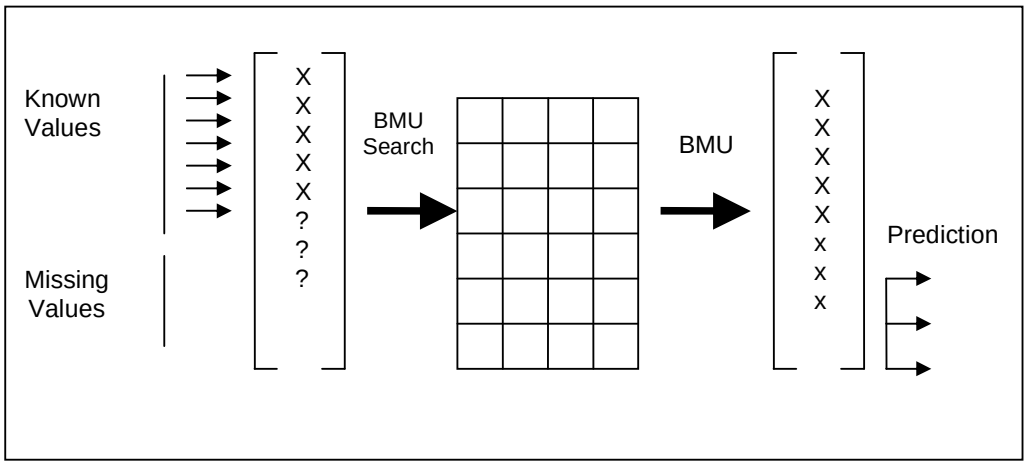


Figure 5.2 *Prediction of missing components of the input vector using the Kohonen Self-organizing Map. BMU = best matching unit.*

5.5 Experimental data

The application is applied to the daily records from the operation of the Seafield wastewater treatment plant in Edinburgh, UK, presented in Chapter 4, during a period of about three years with a total of N=1066 data samples (or records). Summary statistics of the measured process variables are shown in Table 5.1.

As shown in Table 5.1, there are large numbers of missing values which cannot be thrown away. In addition, the missing values occur randomly within the data array. Thus, although the maximum number of missing values is 310 (for the

stirred sludge volume index, SSVI), the number of potentially discardable incomplete daily records or vectors in the data is much more than this number given the non-synchronization of the missing values. In the Seafield data, the total number of data records discardable as a result of non-synchronization was 496.

Table 5.1 also contains the number of identified outliers using the three methods described previously. In general, the modified Z-score method tends to identify more outliers than either the visual inspection or the Z-score method. Indeed, the Z-score method produced the least number of outliers (148) as against 387 by the modified Z-score and 228 based on visual inspection. Because of the restrictive assumptions underpinning the Z-score and the modified Z-score approaches, however, the identified outliers using the visual inspection method were taken as the outliers for the subsequent analysis. These outliers were also removed and treated as missing values to be estimated so as to preserve the true dynamic history of the process as exemplified in the data.

The data in Table 5.1 relate to the secondary treatment stage of the treatment plant. The decision to focus on the biological stage of treatment is because the secondary stage is often the terminal treatment offered at treatment plants discharging to inland rivers or coastal environments. The secondary treatment process helps to remove a substantial proportion of the SS and BOD_5 in the wastewater; it therefore plays a significant role in meeting the quality objectives set for such receiving systems. An analysis such as the KSOM to estimate the missing values should therefore provide complete data for analysing and modelling the biological activated sludge process in wastewater treatment plant.

Table 5.1 *Summary statistics of the measured variables at Seafield Treatment plant*

Variables	Unit	Measurements			Number of missing values	Number of outliers		
		Average	Minimum	Maximum		Visual Inspection	Z-score	Modified Z-score
Influent Flow	m^3/d	259427	171367	466486	19	23	18	54
Influent BOD5	mg/l	65	15	180	105	1	5	22
Influent SS	mg/l	68	3	268	87	7	22	41
WAS Rate	m^3/d	3822	802	6016	146	15	9	20
MLSS	mg/l	2240	1126	4180	246	16	5	25
RAS MLSS	mg/l	4984	1748	1014	303	15	7	28
SSVI	ml/g	92	31	165	310	7	4	14
Sludge Age	Days	5	1	32	225	13	11	30
Actual F/M	Day^{-1}	0.15	0.015	0.43	292	23	4	12
Effluent Flow	m^3/d	250174	65000	461926	1	17	16	58
Effluent SS	mg/l	28	3	190	14	24	24	29
Effluent COD	mg/l	50	15	173	15	48	18	42
Effluent BOD5	mg/l	9	2	351	8	19	5	12

5.6 SOM analysis

The computation for training and searching for the Best Map Units *(BMUs)* was done starting with the default values for the learning rate (=0.5) and neighbourhood radius *(=max(l_1,l_2)/4)* parameters in the SOM Toolbox. Where l_1 and l_2 are the dimensions of the map as presented in Chapter 3, Equation 3.11. In computing the size (and dimension) of the map, the Toolbox uses the formulations in Equations 3.10 and 3..11, but adjust the final map units *M* such that it is equal to the product of l_1 and l_2 exactly. In making this final adjustment, the estimated number of map units may be slightly different from that obtained with Equation 3.10. The analysis led to map size *M = 168* map units which is slightly different from the $M \approx 164$ obtainable using Equation 3.28 with N=1066 and sides l_1 and l_2 of 14 and 12 respectively. Table 5.2 contains other characteristics of the trained KSOM.

Table 5.2 *Characteristics of the trained KSOM*

Normalisation Method	"var": $x^{\setminus} = (x - \bar{x})/\sigma_x$
Codebook	168*13
Neighbourhood function	Gaussian
M size	14*12
Lattice	"Hexa"
Shape	Sheet
Final quantisation error	1.801
Final topographic error	0.066

5.7 Results and discussion

The component planes for each of the 13 variables are shown in Figure 5.3. Each component plane can be thought as a "sliced" version of the KSOM, because it consists of the values of single vector variable in all map units. In other words, the component planes show the value of the variables in each map unit (Vesanto et

al., 2000). These planes are built using colour levels to show the value of a given feature of each KSOM unit in the two dimensional lattice, such that the lighter the colour, the higher the relative component value of the corresponding weight vector.

These component planes help to illustrate visually the relationship between the various parameters or characteristics of the wastewater treatment plant. For example, by looking at the upper left hand of the component planes, we can see that high sludge age is associated with a low waste activated sludge (WAS) rate. This is to be expected given the relationship between the sludge age and the wastage sludge rate (Equation 5.3):

$$\theta_c = \frac{VL}{Q_w L_W} \tag{5.3}$$

where θ_c is sludge age; V = volume of the reactor; L is the MLSS (mixed liquor suspended solids) in the aerator; L_w is the MLSS in the waste activated sludge and Q_w is the waste activated sludge rate. It is therefore to be expected that a combination of low L_w , low Q_w and high L will produce a high θ_c. Other notable relationships visible from the component planes is the low effluent SS, BOD_5 and COD concentrations associated with low hydraulic loading rate to the aerator, which is a natural result of the higher retention time caused by a low hydraulic loading. The complete correlation matrix for all 13 variables of the prototype vectors is shown in Table 5.3, and although this is a simple tool for examining the linear relationship between various variable, its results seem to agree with the indications of the cross-correlation provided by the much more complex KSOM analysis that resulted in the component planes.

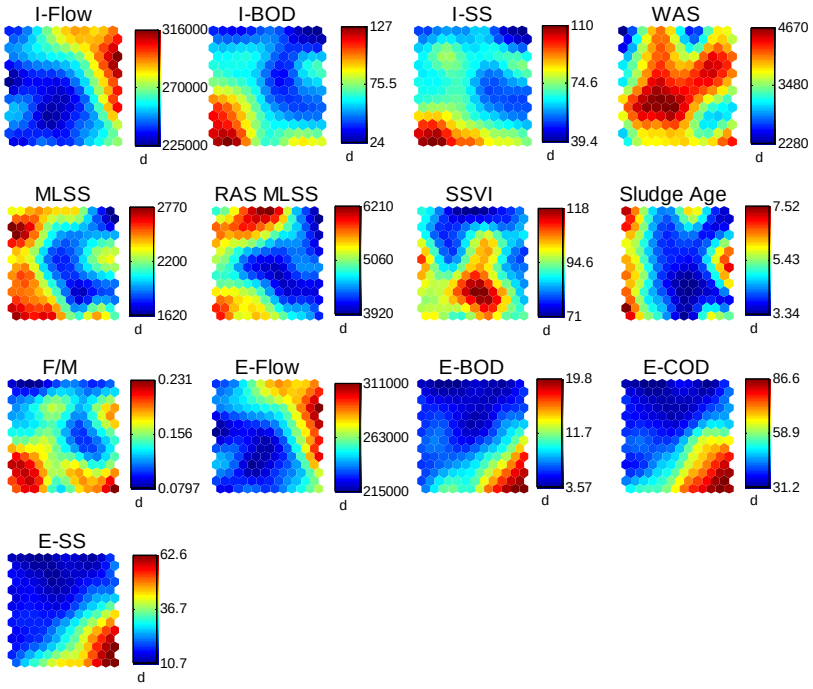


Figure 5.3 *The component planes of the KSOM (I= influent, E=Effluent)*

The performance of the KSOM in predicting the various characteristics is demonstrated in Figure 5.4. In general, the performance is good as further confirmed by the associated statistical indices presented in Table 5.4. For most of the effluent characteristics, the correlation coefficient is generally above 0.90. The model has also particularly done well in predicting the sludge age and F/M ratio, two of the most commonly used parameters for controlling the activated sludge process. This offers some promise for the real-time control of the activated sludge treatment process using these parameters.

Table 5.3 *Correlation matrix for variables in the features*

Variables	Flow	Influent BOD	Influent SS	WAS Rate	MLSS	RAS	SSVI	Sludge age	F:M	effluent flow	Effluent BOD	Effluent COD	Effluent SS
Flow	1												
Influent BOD	-0.43	1											
Influent SS	-0.44	0.88	1										
WAS Rate	-0.16	0.15	0.13	1									
MLSS	-0.25	0.63	0.56	-0.21	1								
RAS-MLSS	-0.18	0.35	0.45	-0.30	0.77	1							
SSVI	-0.62	0.3	0.29	0.20	-0.19	-0.31	1						
Sludge age	0.09	0.42	0.21	-0.46	0.74	0.42	-0.29	1					
F/M	-0.16	0.84	0.74	0.25	0.22	0.02	0.35	0.13	1				
Effluent flow	1.00	-0.42	-0.43	-0.18	-0.24	-0.17	-0.62	0.10	-0.15	1			
Effluent BOD	0.03	0.24	0.19	-0.13	-0.10	-0.34	0.31	-0.05	0.45	0.06	1		
Effluent COD	0.03	0.12	0.10	-0.09	-0.21	-0.44	0.38	-0.16	0.34	0.05	0.96	1	
Effluent SS	0.09	0.1	0.09	-0.13	-0.21	-0.43	0.30	-0.15	0.33	0.11	0.97	0.98	1

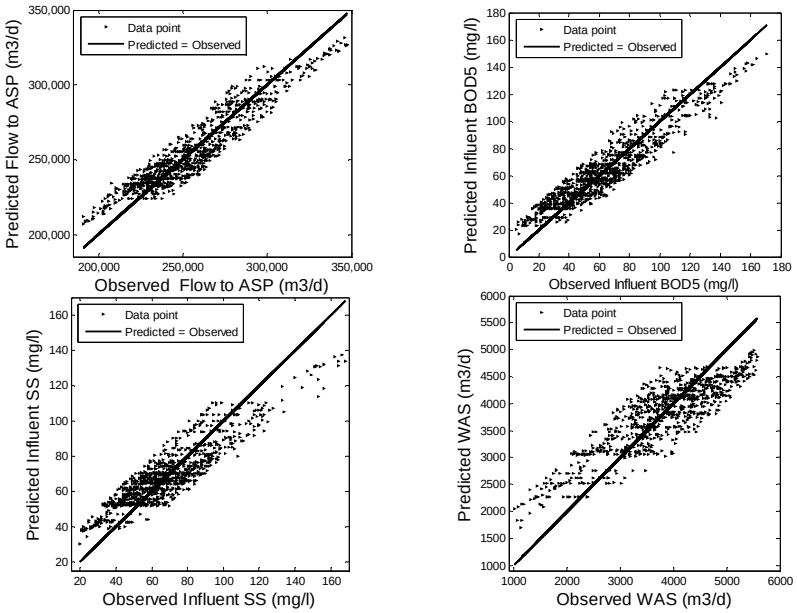


Figure 5.4a *Performance of the KSOM in predicting the process variables*

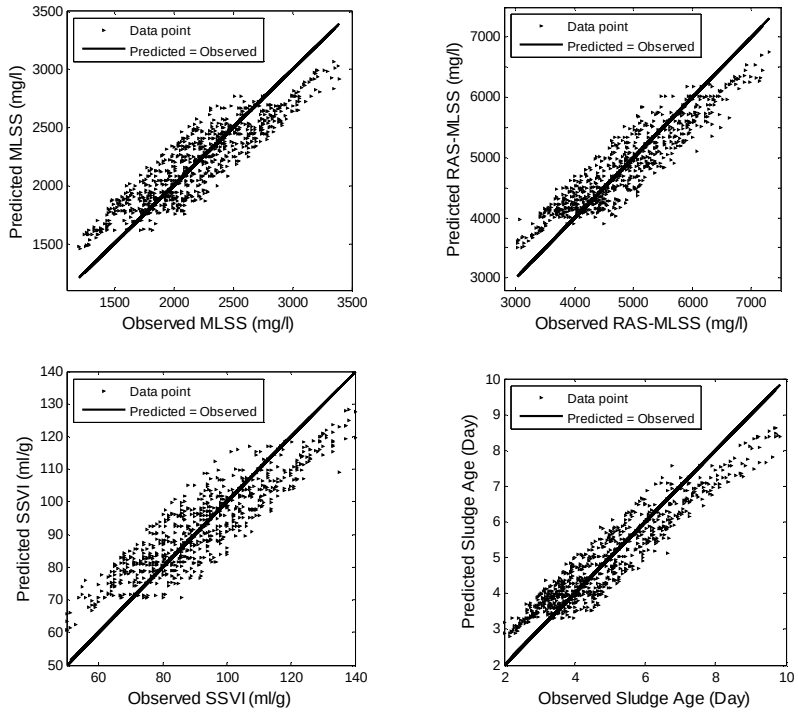


Figure 5.4b *Performance of the KSOM in predicting the process variables*

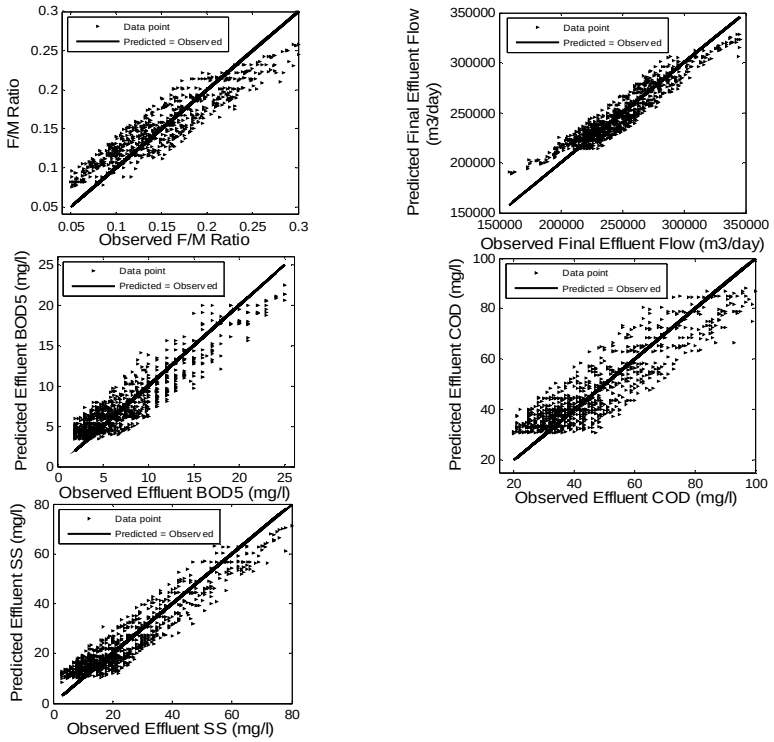


Figure 5.4c *Performance of the KSOM in predicting the process variables*

Table 5.4 *Correlation between measured and KSOM-predicted variables*

Variables	Correlation Coefficient (R)
Flow to ASP	.945
Influent BOD_5 to ASP	.943
Influent SS to ASP	.898
Waste activated sludge rate (WAS)	.879
MLSS	.912
RAS	.933
SSVI	.905
Sludge age	.950
Actual F/M	.934
Effluent Flow	.946
Effluent BOD	.932
Effluent COD	.914
Effluent SS	.950

Visually, Figures 5.5 and 5.6 respectively show the comparison between the estimated and the measured values for the effluent concentrations of BOD_5 and SS. Unlike Figure 5.4, Figures 5.5 and 5.6 help to illustrate how well the KSOM outputs have matched the observed data temporally. In general, the KSOM outputs have correctly reproduced the peaks and troughs in the observed time series data. The predicted missing values are also shown in Figure 5.5 and Figure 5.6, from which it can be seen that their trend is in conformity with the overall trend of the observed data series.

A further analysis was carried out to test whether the sample skewness coefficients of the residuals are statistically zero. This is required to ensure that the residuals have a normal distribution. The sample skew for a variable x_i can be estimated using Equation 5.4.

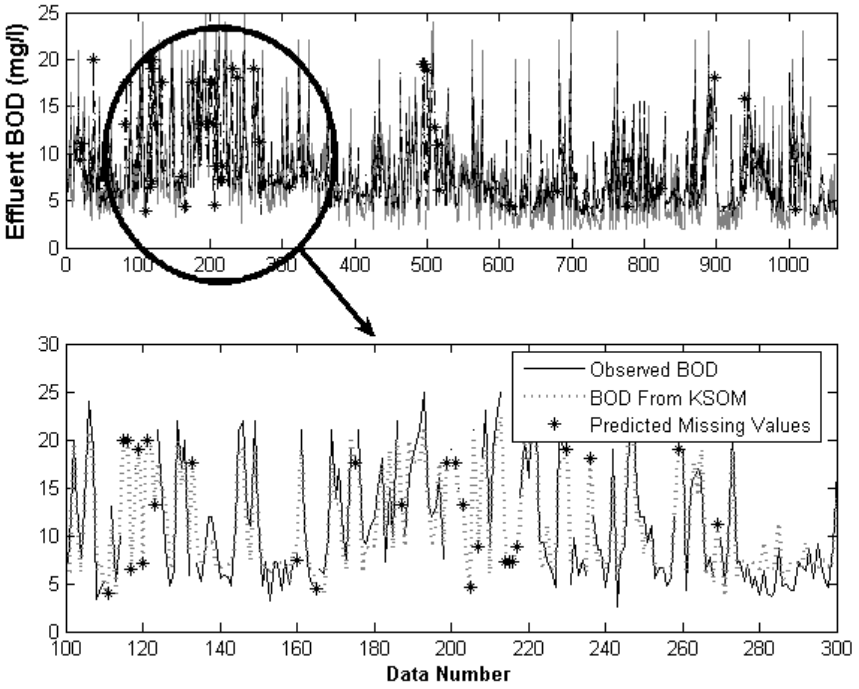


Figure 5.5 *Comparing the observed and KSOM predicted time series plots for Effluent BOD_5*

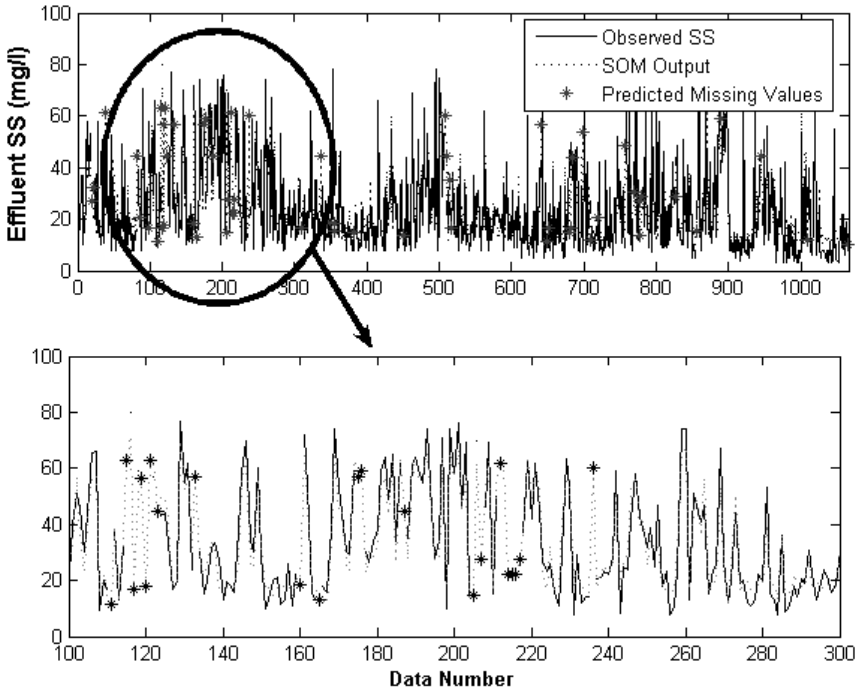


Figure 5.6 *Comparing the observed and KSOM predicted time series plots for effluent SS*

$$skew = \frac{N}{(N-1)(N-2)} \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{S} \right)^3 \tag{5.4}$$

where N is the sample size, $\bar{x}$ is the sample mean and S is the sample standard deviation. Based on the null hypothesis that the skew is zero, the skew coefficient will have a normal distribution with a mean of zero and variance of 6/N. Therefore, The 95% confidence interval for a zero skew is $[-1.96\sqrt{6/N}, +1.96\sqrt{6/N}]$. If the estimated sample skew coefficient lies within this interval, then the null hypothesis cannot be rejected at the 5% level. The results of the hypothesis testing for all 13 variables are shown in Table 5.5 from which it is clear that the residuals associated with most of the characteristics are distributed as normal. The only exceptions are the influent SS, F/M and the

effluent COD whose test statistics fall marginally outside the 95 % confidence interval for a zero skew.

Table 5.5 *Result of approximate normality test for the residuals*

Variables	N	$-1.96 \times \sqrt{\frac{6}{N}}$	$+1.96 \times \sqrt{\frac{6}{N}}$	Skew coefficient	Normal (Y/N) (Skewed)
Influent Flow	1002	-0.152	0.152	0.020	Y
Influent BOD_5	978	-0.154	0.154	0.154	Y
Influent SS	982	-0.153	0.153	0.169	N
WAS	887	-0.161	0.161	-0.093	Y
MLSS	788	-0.171	0.171	0.026	Y
RAS	703	-0.181	0.181	0.055	Y
SSVI	327	-0.265	0.265	0.013	Y
Sludge age	815	-0.168	0.168	0.152	Y
F/M	726	-0.178	0.178	-0.193	N
Effluent flow	1032	-0.149	0.149	-0.082	Y
Effluent BOD_5	1020	-0.150	0.150	0.143	Y
Effluent COD	1011	-0.151	0.151	0.163	N
Effluent SS	1005	-0.151	0.151	0.144	Y

5.8 Discussion of Results

The performance of the KSOM was compared with the use of simple linear regression and back propagation ANN for predicting the effluent BOD_5. On the basis of the correlation matrix shown in Table 5.3, the independent (i.e. input) variables for both the regression and ANN were chosen as the effluent COD (R = 0.96), effluent SS (R = 0.97) and the F/M ratio (R = 0.45). Since complete records are required for these two techniques, only the 770 data records with no missing values in these four parameters were used. Of these, 500 data records were used for model calibration and 270 were used for model testing. The final regression model was:

$$BOD = -1.01 + 11.59F/M + 0.075COD + 0.146SS \quad (R = 0.75) \qquad (5.5)$$

The backpropagation ANN had a single hidden layer and the optimum number of neurons in this hidden layer was found to be 25. Table 5.6 compares the performances of the regression and ANN with the KSOM. In general, the linear regression model had the least performance of the three approaches. The backpropagation ANN was a much more improved approach than the regression, particularly during training, but its performance is still inferior to that of the KSOM. It should be noted that the KSOM statistics quoted in Table 5.6 relate to the reduced 770 sample size unlike those in Table 5.4, which relate to the entire 1,066 data record. A further advantage of the KSOM is that the same map can be used for predicting any missing value in any variable, whereas if the missing variable were to change from the BOD_5, new regression and ANN models would have to be developed. Additionally, the KSOM is not affected by missing values, implying that it is unnecessary to carry out any pre-processing for identifying complete records before the method can be applied. Both the regression and ANN approaches require complete records and hence extensive pre-processing of the data is required before they can be applied.

Table 5.6 *Comparing KSOM, regression and ANN for predicting the effluent* BOD_5

Modelling Method	Data set (Training 500 data point/Testing 270 Data points)	Correlation	MSE (Mean Square Error)	AAE (Average Absolute Error)
Regression	Training	0.75	20	2.5
	Testing	0.76	11.3	2.6
BP-ANN (25 neurons in the hidden layer, best number from trial and error)	Training	0.94	5.0	1.6
	Testing	0.88	5.4	1.5
KSOM	Training	0.96	5.5	1.2
	Testing	0.95	6.0	0.9

5.9 Conclusion

In this chapter, raw data of operational process variables obtained from the Seafield activated sludge wastewater treatment plant Edinburgh, UK during a period of about three years have been modelled to replace outliers and missing values using the Kohonen Self Organising Map (KSOM). Each sample comprises 13 quality and process variables.

After the iterative training of the KSOM, each of the 1066 samples was associated with an output unit known as the best map unit (BMU). The outliers or missing values were then replaced with the corresponding component from the BMU. The results demonstrated that the KSOM is an excellent tool for replacing outliers and missing values in high dimensional data sets. The predicted missing values are plausible and show a trend not dissimilar to that of the observed measurements. These results cannot be obtained from traditional time series models due to the multivariate, time varying and highly non-linear nature of the process. The method is simple, computationally efficient and highly accurate. The method was also shown to outperform linear regression and supervised-learning ANNs. The method therefore provides useful tool for a process engineer who is faced with improving the performance of the WWTP, given the usually incomplete and noise process time series data.

The next Chapter present a novel methodology based on the same techniques presented in this Chapter to predict the biological oxygen demand concentration in wastewater based on its correlation with other water quality parameters that can be measured within three hours or in real time using on-line hardware sensors. This allows for a timely intervention and cost reduction during problem diagnosis.

CHAPTER 6

APPLICATION II: APPLYING KOHONEN SELF-ORGANIZING MAP AS A SOFTWARE SENSOR TO PREDICT THE BIOCHEMICAL OXYGEN DEMAND

6.1 Introduction

The dissolved oxygen used during the bacterial oxidation of organic pollutants, usually determined under standard conditions (APHA, 1998), is known as the biological oxygen demand (BOD). The BOD is a widely applied parameter for assessing water quality, the bio-treatability of wastewater, performance of wastewater treatment operations, and organic loadings to treatment plants for the purpose of sizing aeration facilities. The BOD is also important for discharge consents for wastewater treatment plant facilities in many countries (EEC, 1991).

However, the usefulness of the BOD for the effective monitoring and control of water pollution and wastewater treatment plant performance is severely constrained by the long time it takes to obtain an estimate, which precludes its use in 'real time' control. For example, considering a sustained low dissolved oxygen content in a river due to the discharge of large concentrations of organic pollutants, it will take at least five days before the problem can be effectively diagnosed; any remedial actions to be taken are likely to be too late. This is because the traditional bioassay method for determining the BOD requires incubation for at least 5 days; hence the term 5-days BOD or BOD_5. Consequently, it has become very desirable to have a rapid and accurate inferential model for BOD prediction, thus removing the time delay associated with the laboratory-based bioassay method.

Due to the long time it takes to obtain the standard BOD_5 , some have argued that the BOD data are not useful for the purpose of process control, and have therefore advocated the use of the chemical oxygen demand (COD) instead (Mohanty et al., 2002; Scholz, 2006). However, in contrast to the COD, the BOD only characterizes the biodegradable component of organic matter in wastewater; it is therefore the most important water quality parameter for assessing biotreatability. Furthermore, previous attempts to calibrate BOD against COD have had only limited success, because of the wide variability often observed in the relationship between the two water quality variables (Mohanty et al. 2002; Scholz, 2006). Thus, although the COD is a faster water quality parameter to determine than the BOD, it is certainly not a direct substitute.

An alternative to measure BOD is the development of biosensors, which are devices for the detection of an analyte that combines a biological component with a physicochemical detector component. Biosensors indirectly measure the BOD via a short term (e.g. < 30 minutes) BOD_{st} and a calibration curve which relates this BOD_{st} to the BOD (see Karube et al., 1977 for the pioneering work in this area). Consequently, biosensors are now commercially available, but as will be revealed in the next section, they do have several limitations such as their high maintenance costs, limited run lengths and the need for frequent reactivation, and their inability to respond effectively to rapidly changing water quality characteristics to which wastewater treatment streams can sometimes be subjected to, particularly during storm events (Praet et al., 1995). Another important limitation with the use of biosensors is the uncertainty associated with the calibration function for translating the BOD_{st} to the BOD_5.

Because of versatility of the KSOM in handling high dimensional data and establishing correlation or the cause-effect relationship in a multivariable framework, the KSOM was used in application II to develop intelligent models

for making rapid inferences about BOD_5 using other easy to measure water quality parameters, which, unlike BOD_5, can be obtained directly and reliably using on-line hardware sensors. This will make the use of BOD_5 for on-line process monitoring and control a more plausible proposition. In comparison to other data-driven modelling paradigms such as multi-layer perceptrons artificial neural networks (MLP ANN) and classical multi-variate regression analysis, the KSOM is not inhibited by missing data. Moreover, time sequencing of data is not a problem when compared to classical time series analysis (Lee and Scholz, 2006a,b).

6.2 Review of alternatives to bioassay determination of the BOD_5

Attempts to improve the rapidity with which the BOD_5 estimates are obtained have led to the development of BOD_5 biosensors. A microbial BOD_5 biosensor usually consists of microbial cells immobilized on an oxygen electrode. The first generation of such sensors was developed by Karube et al. (1977) for BOD_5 determination within minutes. However, biosensors do not measure the BOD_5 directly, but they record the enhanced respiration rate of the immobilized microbes caused by the influx of biodegradable organics in the presence of sufficient oxygen (Kim and Kwon, 1999). The graphical representation of the signals of this activity, which is picked up by the dissolved oxygen probe, is known as the respirogram, which can be correlated with the BOD_5 after calibration as seen from Figures 6.1 and 6.2. It is therefore to be expected that the outcome of any BOD_5 determination exercise with the biological electrodes will depend on the species composition of the immobilized micro-organism community, in addition to other boundary conditions and variables such as temperature, pH, the quantity of biodegradable organics, and the presence of toxins and other growth inhibiting substances (Scholz, 2006).

The original biosensor by Karube et al. (1977) used *Trichosporon cutaneum*, but later researchers applied other microbial species such as *Bacillus subtilis* (Tan et al., 1992; Riedel et al., 1988), Klebsiella oxytoca (Ohki et al., 1994) and activated sludge (Sakai et al., 1995). Praet et al. (1995) provided a detailed and critical review of available biological electrodes. They concluded that current BOD_5 biosensors are not a panacea for the limitations caused by the five days time delay associated with conventional bioassay BOD_5 determination.

Considering that hardware biosensors do not provide a satisfactory solution, work has also been carried out on the development of model-based software sensors for the rapid, on-line estimation of BOD_5 and other water quality variables such as COD, heavy metals and nutrients (Lee and Scholz, 2006a,b; Mohanty et al. 2002). In contrast to hardware biosensors, software sensors attempt to obtain the BOD_5 directly and require little or no maintenance as it is based on finding the nonlinear correlation between the BOD5 and the other readily available water quality parameters. Furthermore, software sensors are readily updated if more data become available, unlike hardware sensors, which usually require huge expenses and time to regenerate when the biological cells within the probe are becoming ineffective.

The majority of the available software sensors are based on data driven modelling approaches such as artificial neural networks presented in Chapter 3. The areas of application in water management predominantly include water and wastewater treatment plants, but also sewer systems and storm water runoff management control systems (Shen et al., 2006; Lee and Stenstrom, 2005; Lee and Scholz, 2006a; Mohanty et al., 2002). For example, Brydon and Frodsham (2001) developed a MLP ANN to predict the settled sewage BOD_5 for a wastewater treatment plant in England. Two models with different numbers of input variables

were successfully developed to predict the overall trends in BOD_5 time series data. However, the measured peaks were often not well predicted.

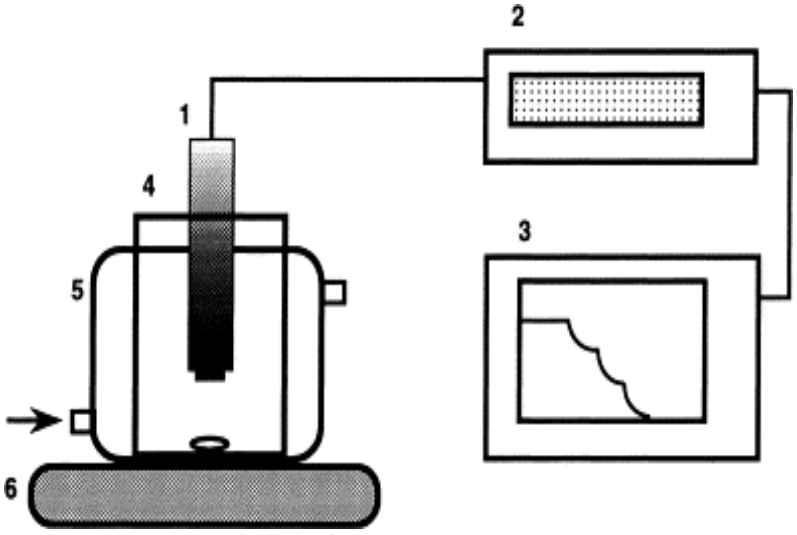


Figure 6.*1 Construction of the BOD sensor, (1) Oxygen electrode, (2) digital multimeter, (3) recorder, (4) beaker, (5) thermostat and (6) magnetic stirrer (Chee et al., 1999).*

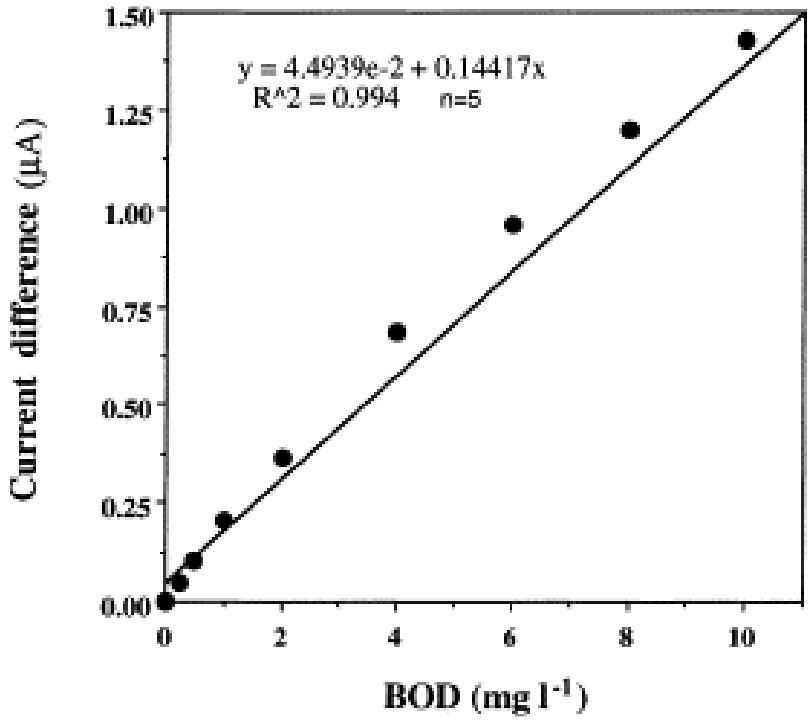


Figure 6.2 *Calibration graph for the BOD sensor using artificial wastewater solutions at pH 7.0, 30°C and 40 mg of immobilized microorganism (Chee et al., 1999)*

Other applications of artificial neural networks to predict water quality parameters including BOD_5 were discussed by Ellis et al. (1990), Hiraoka et al. (1990), Lee and Scholz (2006a,b), and Scholz (2006). The main objective of most of these studies was to provide a BOD_5 estimate in a fraction of the time required for conventional bioassay determination, thus making the BOD_5 a more realistic water quality parameter for real-time process control.

In the current application of the tools developed in this study, unsupervised KSOM instead of the MLP ANN was used as BOD_5 software sensor. The advantages of the KSOM over the MLP ANN are that the former can handle missing values and gaps in the data set without the need for a priori data infilling exercise. Moreover, KSOM can be used to visualize features in the data, thus providing easy to comprehend pictorial evidence of correlations between important water quality parameters.

6.3. Methodology

6.3.1 Case Study

The application of the KSOM for prediction purposes was described in Chapter 5. As a summary, first, a KSOM is trained using the training data set. Then to predict a set of variables as part of an input vector, these variables are first removed from the vector and the depleted vector is subsequently presented to the KSOM to identify its BMU. The values for the missing variables are then obtained by their corresponding values in the BMU.

The KSOM was applied to data from the three wastewater treatment plants described in Chapter 4. The data from the largest Seafield treatment works in Edinburgh were used for model development (calibration and verification) while data from the two smaller works at East Calder and New Bridge, both outside Edinburgh, were used for independently validating the model. All the three works employ the activated sludge treatment process for the secondary biological oxidation; however, given that the emphasis of the study was on the influent (raw sewage) stream, the type of secondary biological process employed will have no bearing on the validity of the modelling results. Consequently, the model would equally apply were the secondary process to be the trickling filter or any other process technology.

Exhaustive details about the Seafield works have been presented in Chapter 4; since the other two works also employ the activated sludge to the treatment process, they are essentially similar to the Seafield except for size. Summary of the data at the three works is shown in Table 6.1. All the data were provided by Thames Water plc, the plant operators. The Seafield data comprises daily measurements of raw sewage quality characteristics including the flow, BOD_5, COD, the suspended solids (SS), etc. for a period of approximately three years, giving a total of 1066. An important feature of the three data sets is the large number of missing values, which would have made modelling by other methods such as regression analysis and MLP ANN impractical. Of the 1066 Seafield vectors, 800 were used for model development and the remaining 266 were used for verification of the model. All the 1091 data vectors (East Calder works) and 1375 data vectors (New Bridge works) were used for independently validating the model. As shown in Table 6.1, the main characteristics at both East Calder and New Bridge are within the range of the Seafield values.

6.3.2 Numerical Analysis and Modelling

The self-organizing map (SOM) toolbox for Matlab 5 was used for this case study. The toolbox was developed by the SOM team at the Helsinki University of Technology, Finland (http://www.cis.hut.fi). An initial pre-processing of the data provided by Thames Water for input into Matlab 5 was carried out using Microsoft Excel, such as changing the missing data or erroneous measurements into NAN (Not a Number) to cope with the Matlab requirement.

Given the objective of the application, i.e. to develop a model for the rapid estimation of BOD_5, a decision had to be made on which of the available water quality variables to include alongside BOD_5 in the KSOM. This decision was based on two considerations: the rapidity with which the variables could be measured or evaluated using hardware sensors and the degree of association between the variable and the BOD_5. The former consideration is important for the rapid prediction of the BOD_5. Therefore, the inflow, COD, SS, ammonia-nitrogen, pH and temperature (Table 6.1) were chosen, because they can be measured within three hours or by on-line hardware sensors to allow BOD_5 prediction in virtually ‘real’ time.

For modelling purposes, the complete data set was subdivided into two sets; a first set of 800 observations to train the model (training set) and a second set of 266 observations to test the model (validation set). Three different KSOM models were trained with the training data set, with each KSOM model having different input variables alongside BOD_5. The corresponding list of variables for each KSOM model is given in Table 6.2.

Table 6.1 *Water quality of the inflow characteristics*

Plant	Symbol	Description	Unit	Min.	Max.	Mean
Seafield (1066 data points)	Inflow	Flow rate to the treatment plant	m^3/d	100000	674000	300261
	COD	Chemical oxygen demand	mg/l	74	880	350.58
	SS	Suspended solids	mg/l	15	580	164.27
	NH_4	Ammonia-nitrogen	mg/l	0.50	35.84	13.83
	pH	pH	-	7.09	9.20	6.10
	T	Temperature	^oC	9	19	14
	BOD_5	Five days @ 20°C biological oxygen demand	mg/l	19	244	117
East Calder (1091 data points)	Inflow	Flow rate to the treatment plant	m^3/d	18956	56052	30921
	COD	Chemical oxygen demand	mg/l	186	730	410
	SS	Suspended solids	mg/l	34	354	169
	pH	pH		6.8	7.8	7.33
	NH4	Ammonia-nitrogen	mg/l	3.85	30.14	18.23
	BOD_5	Five days @ 20°C biological oxygen demand	mg/l	30.5	152.5	96.4
New Bridge (1375 data points)	Inflow	Flow rate to the treatment plant	m^3/d	7235	28389	13550
	COD	Chemical oxygen demand	mg/l	113	778	395
	SS	Suspended solids	mg/l	38	361	166
	NH_4	Ammonia-nitrogen	mg/l	3.18	38.33	17.51
	pH	pH		6.1	7.7	7.2
	BOD_5	Five days @ 20°C biological oxygen demand	mg/l	34.93	189.1	103.8

Table 6.2 *The structure of the trained Kohonen self-organizing map models with a map size of 18×8 and a hexa lattice*

Map no.	Input variables	Quantization error	Topographic error
1	Influent flow, COD, SS, NH4, pH, Temperature, BOD5	1.176	0.059
2	Influent flow, COD, SS, NH4, pH,BOD5	0.975	0.055
3	Influent Flow, COD, SS, NH4,BOD5	0.723	0.022

The estimated BOD_5 concentrations from the BMU of the training set were compared with the actual concentrations to evaluate the performance during training. Following training, the models were then validated with the validation data set, which was not used during the training phase. In the verification phase, the BOD_5 was omitted from the input vectors, essentially implying that BOD_5 values were 'missing'. The BMU for each verification input vector were then determined to derive the missing BOD_5 values. Finally, BOD_5 concentrations obtained with the BMU were then compared with the actual observations to evaluate the performance during verification. The performance of the models during training and verification was evaluated using the evaluation criteria described in Chapter 4 (See section 4.6).

6.4 Results

The component planes for each variable of the three KSOM are shown in Figure 6.3 to illustrate the associations between variables. Each component plane can be viewed as a 'sliced' version of the KSOM, because it consists of the values of single vector variables in all map units. In other words, the component planes show the values of the variables in each map unit (Vesanto et al., 2000). These planes are filled using colours to show the individual values of a given feature of

each KSOM unit in the two dimensional lattice, such that the lighter the colour grey, the higher the relative component value of the corresponding weight vector.

Thus, the component planes help to visually illustrate the relationship between BOD_5 and other variables considered in this analysis. For example, by looking at the upper left hand of the component planes, it can be seen that low BOD_5 concentrations are associated with a high influent flow rate, which is expected given the dilution effects of high inflow rates. Furthermore, low BOD_5 concentrations in the influent are associated with low COD, SS and ammonia-nitrogen concentrations. While BOD_5 is positively correlated with SS, ammonia-nitrogen and COD, it is negatively correlated with the inflow rate. Furthermore, visual inspections of the component planes of pH and temperature do not reveal any obvious association with BOD_5. In particular, the relationship between BOD_5 and pH is weak (Figure 6.3).

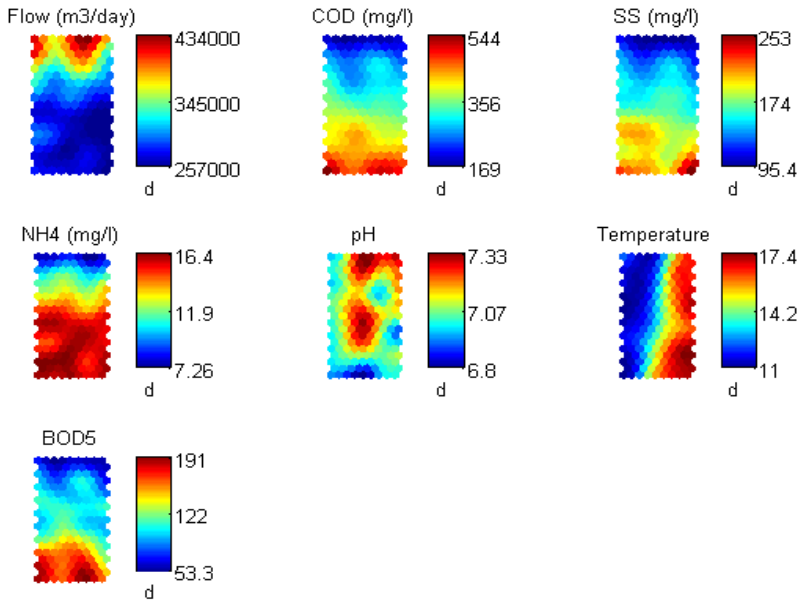


Figure 6.3(a) *Component planes for the Kohonen self-organizing map (KSOM) model 1*

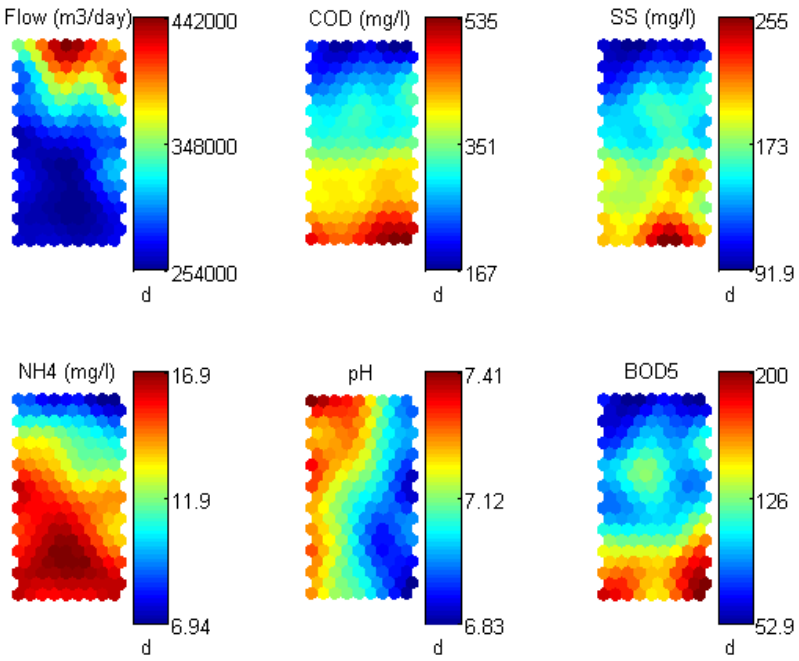


Figure 6.3(b) *Component planes for the Kohonen self-organizing map (KSOM) model 2*

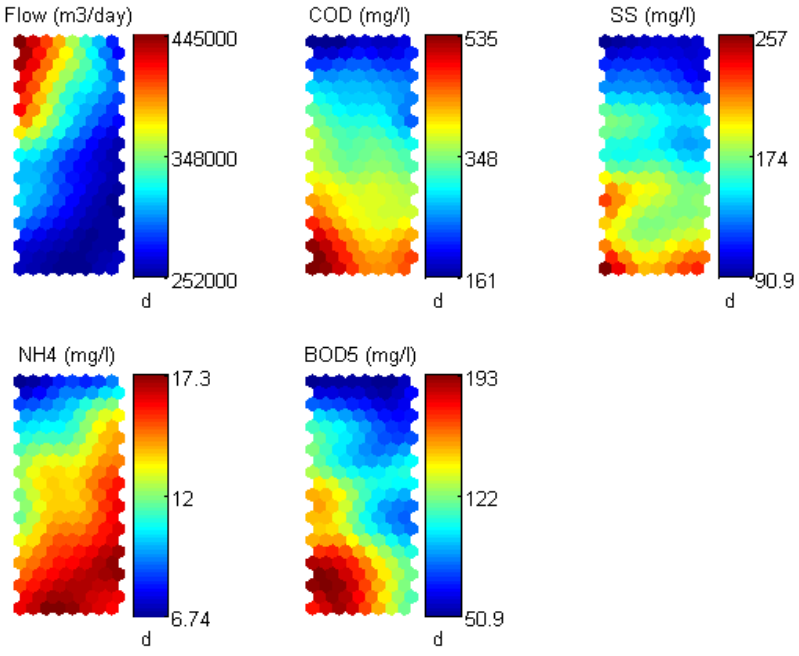


Figure 6.3(c) *Component planes for the Kohonen self-organizing map (KSOM) model 3.*

The structures of the three trained KSOM are summarised in Table 6.2. All the KSOM models have the same map size. However, the third map has the smallest quantization and topographic errors, which indicate that this is the best model for prediction purposes. Table 6.3 summarizes the errors during training and verification, which reinforce the superiority of the third KSOM model. For example, the correlation coefficient between the measured and KSOM-predicted BOD_5 during training was 0.93 for map 3 in comparison to 0.86 and 0.91 for maps 1 and 2, respectively. Map 3 also performs much better than maps 1 and 2 on the basis of the AAE: 4.34 as opposed to 18.86 and 14.78 for maps 1 and 2, respectively, during training.

Assuming that the errors are normally distributed, the approximate 95% confidence interval is shown by Equation 6.1.

$$95\%Conf = AAE \pm 1.96\frac{S}{\sqrt{N}} \tag{6.1}$$

where 95% Conf is the 95% confidence interval; AAE is the average absolute error; S is the standard deviation of the absolute errors (Table 6.3); and N is the sample size, which is 800 for the training set and 266 for the validation set.

The 95% confidence interval for the AAE of KSOM 3 is 4.34 ± 0.4 during training and 4.92 ± 0.55 during verification (Table 6.3). In comparison, these errors are considerably lower than the errors obtained by Brydon and Frosham (2001) for their two MLP ANN models also predicting BOD_5.

The superiority of map 3 is also confirmed by the classification error (CE) as shown in Table 6.3. The fraction of the false positive rate is nearly equal to the fraction of the false negative rate. This implies little bias in predictions obtained by map 3 in comparison to predictions obtained by maps 1 and 2. The CE for

maps 1 and 2 during verification were worse than during training, which is a clear evidence of the significant upward bias in the BOD_5 prediction for these two models.

Both temperature and pH correlated weakly with the influent BOD_5 (as noted earlier) but were included because of the theoretical importance of these two variables for the reaction kinetics in water and wastewater treatment (Mohanty et al., 2002; Scholz, 2006). The fact that the third KSOM, which contains neither the pH nor temperature, is performing better than the first two models, which included these variables (Table 6.2), confirms that both temperature and pH are not directly linked to BOD_5.

Because of the relative superiority of the map 3, further analysis was only undertaken with this map. Figures 6.4a and 6.4b show the time series of the observed and predicted BOD_5 during testing and verification, respectively. These show that the performance of the KSOM is very good and in agreement with corresponding evidence presented in Table 6.3. The plot of residuals is shown in Figure 6.4c indicating that the residuals are random as expected from the approximate 50:50 split of the positive and negative prediction fractions of map 3.

The observed and modelled values have been presented in Figures 6.5a (training) and 6.5b (verification) to better demonstrate how well the KSOM 3 is predicting the BOD_5. Most data points are contained within the 95% prediction limits, particularly during verification. These findings support the use of the KSOM model for predicting mean BOD_5 concentrations. Moreover, the 95% prediction limits during verification completely enclosed all the plotted points, implying that individual BOD_5 concentrations can be predicted with sufficient accuracy using KSOM map 3.

6.5 Discussion of Results

As shown above, the KSOM has performed reasonably well during calibration and verification with the Seafield data. However, this cannot be taken as evidence that the model is sufficiently general as to perform well when applied to an independent data set. In other words, to ascertain the true ability of the developed model to generalize, it must be applied to independent data sets unrelated to those used for the model development.

Consequently, the KSOM model 3 was applied to the East Calder and New Bridge data sets for the purpose of validation. The validation exercise followed the same procedure used for verifying the model earlier, i.e. in which the BOD_5 measurement in each measured vector was removed, assumed missing. The validation results are summarised in Table 6.4. Also included in Table 6.4 is the summary of the verification results at Seafield. In general, the performance of the model is good. In particular, the correlation coefficient between the observed and predicted BOD_5 during validation was very close to that obtained for the verification at Seafield. However, much more satisfying is that the model was able to better match the mean, minimum, and maximum concentrations at the validation sites than at the Seafield calibration sites.

All of the above is proof that the KSOM model developed in this work is an adequate predictor for the BOD_5, thus meeting the objective of this application which is to develop a rapid, software sensor for the BOD_5. However, like any modelling exercise, the developed model has its limitations which should be borne in mind when applying the model. The first limitation is that the model has been developed using historical daily data only. The number of samples taken for each variable differs from one variable to another, which could have affected the accuracy and precision of the measurements. This is a result of inconsistencies in

the 'real' industrial sampling regime, which was not designed for research purposes.

Table 6.3 *Summary statistics of the three Kohonen self-organizing map models for predicting the biochemical oxygen demand during training and verification. The verification values are in parentheses.*

Map no.	Statistics	Mean (mg/l)	Standard deviation (mg/l)	Mini-mum (mg/l)	Maxi-mum (mg/l)
1	Measured	126 (92)	45.83 (32.32)	19 (22)	244 (192)
	Estimated	124 (99)	36.99 (26.24)	52 (52)	192 (186)
	Absolute error	18.86 (14.00)	14.560 (16.000)	0.03 (0.02)	89.92 (61.00)
	Correlation	0.86 (0.87)			
	AAE	18.86 (14.00)			
	RMAE	0.084 (0.082)			
	NRMSE	0.520 (0.547)			
	Fraction positive	45% (64%)			
	Fraction negative	55% (36%)			
2	Measured	126 (92)	45.83 (32.32)	19 (22)	244 (192)
	Estimated	124 (98)	38.92 (27.14)	52 (52)	201 (193)
	Absolute error	14.78 (12.16)	12.320 (9.590)	0.01 (0.02)	93.91 (43.32)
	Correlation	0.91 (0.90)			
	AAE	14.78 (12.16)			
	RMAE	0.066 (0.072)			
	NRMSE	0.420 (0.479)			
	Fraction positive	47% (66%)			
	Fraction negative	53% (34%)			
3	Measured	126 (92)	45.83 (32.32)	19 (22)	244 (192)
	Estimated	125 (96)	40.67 (28.33)	50 (50)	195 (187)
	Absolute error	4.34 (4.92)	11.530 (8.890)	0.00 (0.00)	68.16 (47.06)
	Correlation	0.93 (0.91)			
	AAE	4.34 (4.92)			
	RMAE	0.055 (0.050)			
	NRMSE	0.360 (0.414)			
	Fraction positive	52 (53%)			
	Fraction negative	48 (47%)			

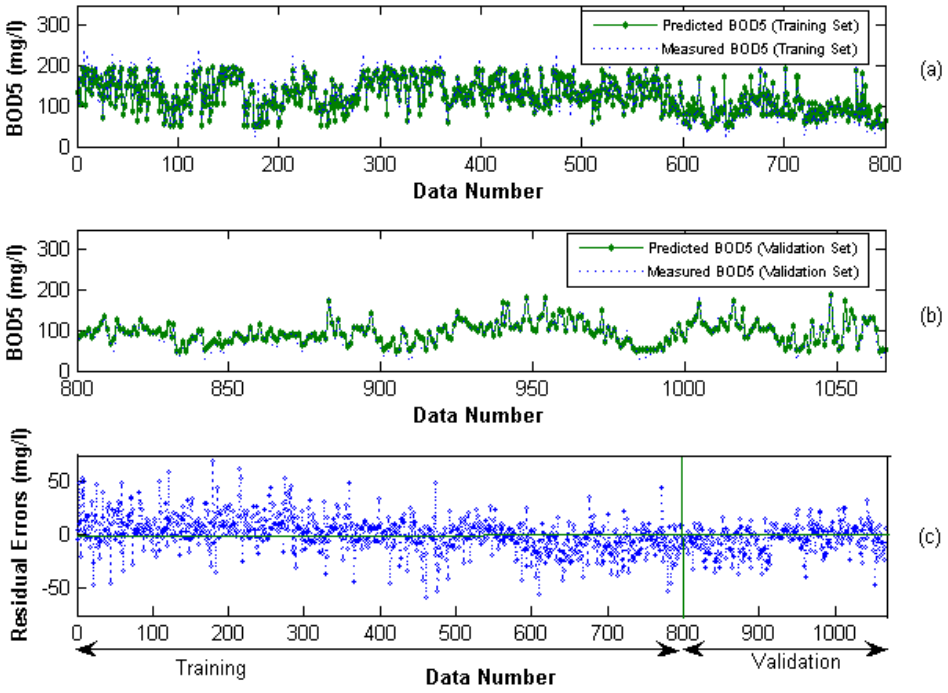


Figure 6.4 *Comparison of the observed and predicted five BOD_5 concentrations of the Kohonen self-organizing model 3 during (a) training and (b) verification. The corresponding residuals are shown in part (c).*

Furthermore, the number of data points used to develop the model limits its overall performance. As a general rule of thumb, if more data points are available for the model development, the performance of the model is likely to increase, because it can extract more patterns out of a larger data set. Finally, the model has been developed using raw wastewater quality characteristics. However, its application can be widened if more data from other sources such as river water or treated effluent wastewater are available for analysis. More data from different sources would allow the model to capture the patterns of data from a wider range of scenarios. This may also lead to an increase of the model performance.

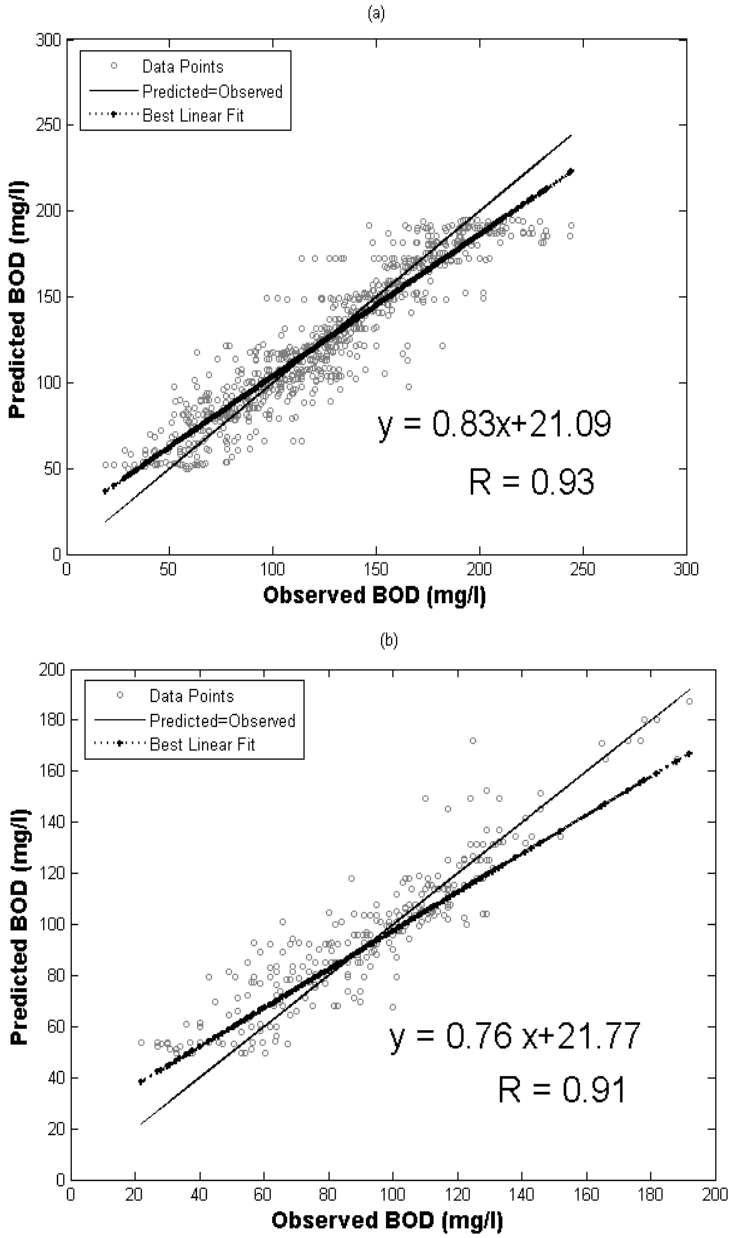


Figure 6.5 *The performance of the Kohonen self-organizing map model 3 in predicting BOD_5 during (a) training and (b) verification.*

Table 6.4 *Summary statistics of the validation for the Kohonen self-organizing map model number 3 for five days @ 20℃ biochemical oxygen demand predictions with data from Seafield, East Calder, and New Bridge.*

	Seafield	East Calder	New Bridge
MSE	79	98	112
AAE	4.92	7.99	8.82
Correlation	0.91	0.88	0.87
Fraction positive (%)	53	44	53
Fraction negative (%)	47	56	47
Minimum measured	22	31	35
Minimum predicted	52	50	50
Maximum measured	192	152	189
Maximum predicted	186	154	182
Mean measured	92	96	104
Mean predicted	96	96	103

The proposed software sensor is fundamentally different from any biosensor, and only an empirical comparison rather than an objective statistical comparison is therefore possible. Biosensors do not measure BOD_5 directly or estimate actual BOD_5 concentrations based on historical BOD_5 data such as the proposed model, but they rely on a linear correlation between the responses of a sensor and the BOD values. Considering that a large variety of organisms could be used as biosensors, a comparison with the proposed model is meaningless. Moreover, large numbers of biosensor data are rare and were not available for the case study site or the sites used for validation purposes. The only journal paper so far that refers to biosensors and neural networks for BOD_5 determination in the same context has been written by Roche et al. (2006). However, neural networks were only used to model direct biosensor performances.

The performance of the KSOM was compared with the use of simple linear regression for predicting the BOD_5 On the basis of the correlation between BOD and COD. 800 data records were used for model calibration and 266 were used for model testing. The final regression model was [BOD = 18.24+ 0.296 COD] with correlation coefficient R = 0.75 during calibration data and 0.71 during testing. As

expected, the KSOM model performed better than the simple linear regression model.

6.6 Summary

The current work presents a completely novel methodology based on the use of the Kohonen self-organizing map (KSOM) models to predict five-days @ 20°C biochemical oxygen demand (BOD_5) concentrations in wastewater, using raw sewage data obtained at three wastewater treatment plants in Scotland. Extensive testing and validation of the model shows that the model is sufficiently general to predict the BOD_5 readily using variables, which can be measured within three hours or in real-time using on-line hardware sensors, thus making it possible to estimate BOD_5 very rapidly. This allows for a timely intervention and cost reduction during problem diagnosis.

The proposed BOD_5 software sensor methodology is an alternative to BOD biosensors because the BOD_5. The software sensor does not require calibration and cannot be negatively affected by toxins and other inhibitors. Moreover, the software sensor is very dynamic and can be readily updated when additional data become available, thus enhancing its accuracy. Furthermore, the performance of the software sensor is much better than the performance of simple linear regression between BOD and COD.

The KSOM tool used for the development of the software sensor can readily deal with missing values in one or more of the input variables without significantly affecting the accuracy of the model. The proposed methodology is applicable for other water and environmental engineering problems, and this work could therefore be regarded as a teaching aid.

CHAPTER 7

APPLICATION III: MODELLING ACTIVATED SLUDGE PROCESS USING HYBRID KSOM-ANN

7.1 Introduction

As stated in Chapter 2, with tighter regulations on the receiving water quality, it is important to limit point source pollution by improving the performance of wastewater treatment plants. Controlling treatment plants through modelling is technically the most feasible and maybe least costly way of achieving a sustainable improvement in performance. This is because modelling the wastewater treatment units can help the operator to test some corrective actions on the computer and, in this way, identify the corrective actions that give better performance. However, modelling the activated sludge process (ASP) has many problems; all these problems give the ASP its nonlinear characteristics and time-varying parameters. Thus, most approaches to modelling the ASP using mechanistic paradigms have relied on numerous simplifying assumptions in order to make the problem tractable.

In Application III of this work, an alternative approach involving neural computing has been applied to model the ASP. Artificial neural networks (ANNs) can be used to model any complex, nonlinear and dynamic systems without the need to specify the functional form of the governing relationship a priori (Landeras et al., 2008). However, basic multi-layered perceptron (MLP-ANNs) are affected by the quality of the data such as noise and missing values, which can make effective training difficult. To solve this problem, a model based on the hybrid Kohonen self-organising map (KSOM) and multi-layered perceptron

artificial neural networks (MLP-ANN) was used. The best map units of the measurement vectors over the KSOM were used as inputs to the MLP-ANN to reduce the effects of noise and uncertainty in the measurement data, and to replace the missing elements. This hybrid KSOM-ANN modelling paradigm is illustrated in Figure 7.1.

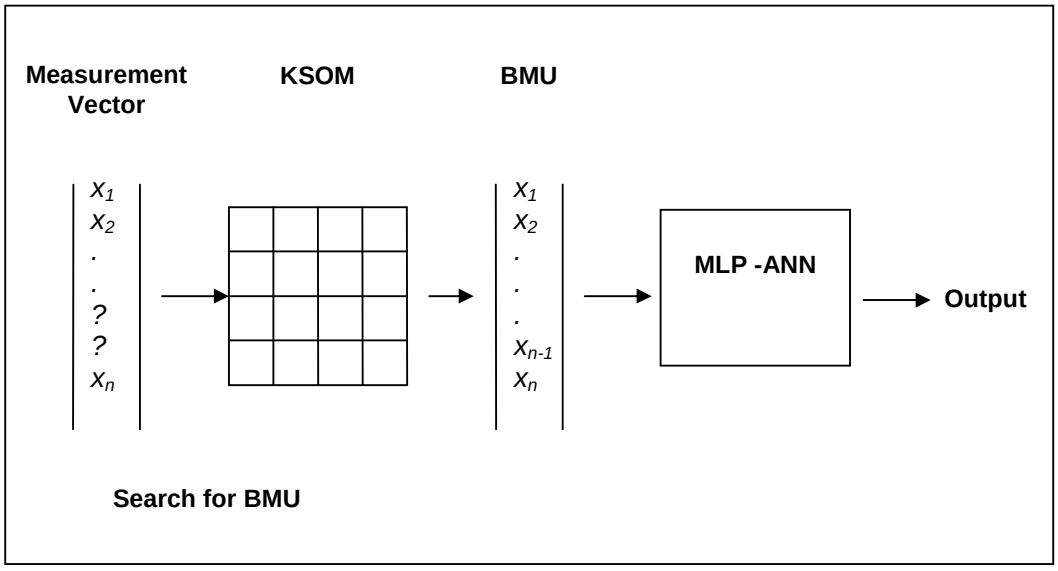


Figure 7.1 *Diagrammatic representation of the integrated KSOM-ANN modelling strategy*

In the application, two situations were thus investigated for the prediction of the effluent BOD_5 and SS concentrations: using the MLP-ANN on raw data; and using MLP-ANN of features extracted using the KSOM (i.e. the hybrid KSOM-ANN).

7.2 Methodology

The theory and mathematical basis of artificial neural networks (ANNs) have been described in Chapter 3. ANNs consist of a set of artificial neurons which are called nodes, and they have connections between them, called weights. Optimal values for these weights are obtained by training the network. The most

commonly used form of ANNs is the multi-layer perceptron (MLP). In general, a three-layered MLP, trained using the Levenberg-Marquardt algorithm, can approximate any function with sufficient accuracy (Hagan et al., 1996; Daliakopoulos et al., 2005). The activation functions chosen were the tan sigmoid transfer function in the hidden layer and the linear function in the output layer.

For both modelling paradigms, the choice of input variables was based on examining of correlation matrix. A correlation analysis, based on the features of the data, was performed between the variables to establish, at a preliminary stage, which of the inputs has the most influence for the performance of ASP(See Table 7.1). The highest 5 correlated variables with the effluent BOD_5 and effluent SS, apart from effluent COD which is a quality output like effluent BOD and effluent SS, were found to be BOD-Load, which has a correlation coefficient of 0.34 and 0.21 with effluent BOD_5 and SS respectively; DO has a correlation coefficient of -0.27 and -0.21 with effluent BOD_5 and SS respectively; RAS-MLSS has a correlation coefficient of -0.34 and -0.47 with effluent BOD_5 and SS respectively; F/M has a correlation coefficient of 0.44 and 0.33 with effluent BOD_5 and SS respectively, and T has a correlation coefficient of 0.20 and 0.33 with effluent BOD_5 and SS respectively. These were thus chosen as possible input variables.

To overcome the over-fitting problem, the early-stop rule was used which necessitated dividing the Seafield data, described in chapter 4, into three subsets for training (500 data points), validation (200 data points) and testing (366 data points). The validation data set was used to stop the training when the errors in this set begin to increase during the training, following a sustained period in which the error fell as seen from Figure 7.2. The testing set was used to assess the ability of the ANN to generalise. The input and target data were normalised in order to have zero mean and unit standard deviation. The outputs of the trained

networks were post-processed, using the inverse of the pre-processing transformation, to be compared with the desired outputs.

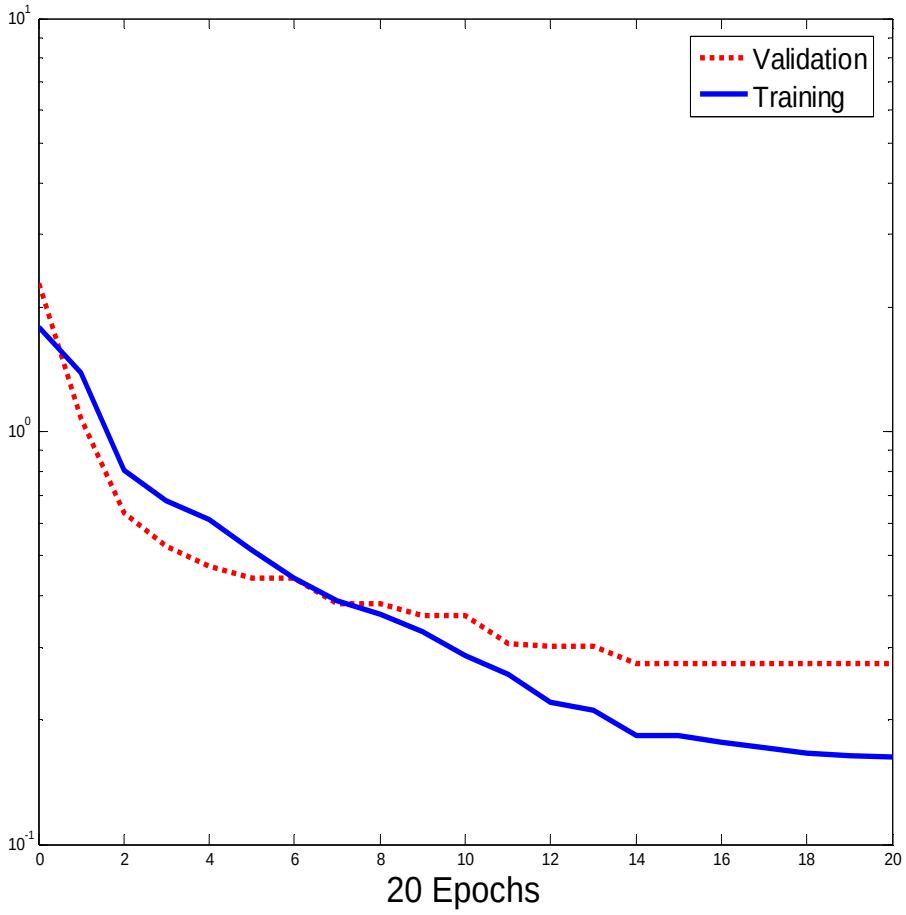


Figure 7.2 *Training and validation errors for Model 1(F) with 5 hidden neurons*

Table 7.1 Correlation matrix of the features

	Flow to ASP m3/d	Infl. BOD to ASP mg/l	BOD load kg/day	Infl. SS to ASP mg/l	DO mg/l	WAS Rate m3/d	MLSS mg/l	RAS MLSS mg/l	SSVI ml/g	Sludge Age day	F/M	final Effluent Temp deg. C	Final Effl. Flow m3/d	BOD5 mg/l	COD mg/d	SS mg/l
Flow to ASP1 (m^3/d)	1.00															
Influent BOD5 to ASP (mg/l)	-0.35	1.00														
BOD load (kg/day)	-0.17	0.98	1.00													
Influent SS to ASP (mg/l)	-0.47	0.91	0.86	1.00												
DO (mg/l)	0.25	-0.57	-0.54	-0.51	1.00											
WAS Rate (m^3/d)	-0.14	0.10	0.09	0.01	0.06	1.00										
MLSS (mg/l)	-0.29	0.60	0.56	0.59	-0.19	-0.35	1.00									
RAS MLSS (mg/l)	-0.33	0.38	0.33	0.49	-0.05	-0.43	0.86	1.00								
SSVI (ml/g)	-0.62	0.33	0.23	0.35	-0.47	0.25	-0.18	-0.29	1.00							
Sludge Age (day)	0.16	-0.01	0.01	0.02	-0.08	-0.93	0.40	0.33	-0.20	1.00						
F/M	-0.11	0.91	0.95	0.78	-0.56	0.21	0.32	0.09	0.32	-0.11	1.00					
Final Effluent Temp. (deg. C)	0.08	-0.34	-0.33	-0.37	-0.02	0.46	-0.80	-0.89	0.45	-0.40	-0.12	1.00				
Final Effluent Flow (m^3/d)	0.92	-0.35	-0.19	-0.46	0.25	-0.19	-0.23	-0.26	-0.60	0.21	-0.14	0.04	1.00			
Effluent BOD5 (mg/l)	0.23	0.29	0.34	0.16	-0.27	-0.04	-0.07	-0.34	0.26	0.17	0.44	0.20	0.22	1.00		
Effluent COD (mg/l)	0.22	0.19	0.24	0.08	-0.23	0.03	-0.18	-0.45	0.33	0.10	0.35	0.37	0.20	0.96	1.00	
Effluent SS (mg/l)	0.31	0.14	0.21	0.03	-0.21	-0.01	-0.21	-0.47	0.25	0.14	0.33	0.33	0.30	0.97	0.98	1.00

Although the correlation analysis carried out gave an initial idea of the relative importance of the different variables that affect effluent BOD_5 and SS concentrations, the best way to select the ideal input variables for a neural network is to train several models with different combinations of inputs and choose the one with best generalisation performance. This approach is time consuming but was used in this study. Therefore, several network architectures, in terms of the number of input neurons, were trained and tested. All of these architectures have one hidden layer. The output variables were the effluent BOD_5 and SS concentrations. The models were evaluated using three criteria namely, correlation coefficient (R), mean square error (MSE) and average absolute error (AAE) as described in Chapter 4.

To reach the suitable network architecture for the MLP-ANN, simulations were run for several of inputs and several of assumed numbers of hidden neurons. All the networks share the same specifications: Three-layer (input, hidden and output layers) feedforward neural networks, Backpropagation learning algorithm, with Levenberg-Marquard optimization technique, Tan-sigmoid transfer function is used in the hidden layer and the linear transfer function is used in the output layer. The number of hidden neurons and the number of inputs nodes were found through a trial and error process. The number of hidden neurons was set to range from 5 to 40. The number of learning iterations was set to be 100 epochs or learning cycles. For best training performance, all of the data presented to the neural network for training were normalized ensuring no signal dominates the training process. This would certainly ensure that all the input signals apply the same influence throughout the training process. The normalization was done by deducing the mean and dividing by the standard deviation giving a transformed variable with a mean of 0 and variance of 1.

Total of seventy two models were thus trained and tested. Models 1 to 36, denoted by M1 to M36, were trained and tested using the raw data with different number of input variables and different number of neurons in the hidden layer. Another set of models

(denoted by M1 (F) to M36 (F)), were trained and tested using the features of the data or their best map units over the KSOM. It is assumed at the beginning of the study that the features of the data will improve the performance of ANNs models since they reduce the effect of noise, outliers in the measurement system. Additionally, these features also allow the networks to be used when there are some missing values, satisfying one of the purposes of this study.

7.3 RESULTS AND DISCUSSION

7.3.1 Artificial neural networks models using raw data

Table 7.2 shows the details of the 36 models trained with the raw data. In the first 12 models, M1-M12, there were 5 inputs and 2 outputs with different number of neurons in the hidden layer. As stated previously, the early stopping technique was employed in which the training process was stopped when the validation error started to increase. This ensures that overfitting dose not occurs. Figure 7.2 shows the training and validation errors for model M1(F) as an example. In figure 7.2, the validation error started to rise after 14 epochs whilst the training error continues to fall. Stopping the training at epoch 14 is therefore warranted to avoid over-fitting. This is much lower than the maximum 100 epochs specified for the learning process.

The next set of 12 models, denoted by M13 to M24, had 4 inputs (BOD-Load, RAS-MLSS, F/M, and Temperature), 2 outputs (BOD_5 and SS) with different numbers of neurons in the hidden layer. The last set of 12 models, denoted by M25 to M36, have 3 inputs (BOD-Load, RAS-MLSS, F/M, and Temperature), 2 outputs (BOD5 and SS) with different numbers of neurons in the hidden layer.

Table 7.3 and Table 7.4 also Figure 7.3 and Figure 7.4 show the comparison of the error criteria of the models developed in the three cases during training, validation and testing for effluent BOD_5 and Effluent SS respectively. It can be seen from these results that using the raw data were hopeless and not acceptable for modelling

purpose. The matter that led us to investigate the possibility of using the features rather that the raw data as well be seen in the next section.

Table 7.2 *The structure of the developed ANN models using raw data*

Inputs	Model Number	Number of neurons in the hidden layer
BOD-Load DO RAS-MLSS F/M T	M1	5
	M2	7
	M3	10
	M4	13
	M5	16
	M6	18
	M7	20
	M8	23
	M9	26
	M10	30
	M11	35
	M12	40
BOD-Load RAS-MLSS F/M T	M13	5
	M14	7
	M15	10
	M16	13
	M17	16
	M18	18
	M19	20
	M20	23
	M21	26
	M22	30
	M23	35
	M24	40
BOD-Load RAS-MLSS F/M	M25	5
	M26	7
	M27	10
	M28	13
	M29	16
	M30	18
	M31	20
	M32	23
	M33	26
	M34	30
	M35	35
	M36	40

Table 7.3 *Statistics analysis of modelling ASP using ANN to predict effluent BOD_5*

Model Number	MSE $(mg/l)^2$			AAE (mg/l)			Correlation (R) %		
	Training	Validation	Testing	Training	Validation	Testing	Training	Validation	Testing
M1	18.11	18.23	26.35	3.25	3.38	4.14	61.07	43.97	11.95
M2	22.73	17.01	17.64	3.60	2.88	3.09	49.54	51.70	32.02
M3	36.15	24.33	27.13	6.54	4.00	4.33	52.10	15.97	6.88
M4	19.16	21.23	30.04	3.40	3.54	4.53	56.88	33.22	8.58
M5	18.97	21.28	26.95	3.44	3.58	4.27	58.16	30.49	10.22
M6	21.28	20.85	32.75	3.69	3.46	4.67	49.13	33.89	3.96
M7	23.22	22.88	28.54	3.90	3.93	4.56	41.17	33.31	9.83
M8	21.01	23.00	24.64	3.64	3.82	4.13	50.81	21.61	13.27
M9	19.79	20.46	26.25	3.42	3.59	4.28	57.31	32.35	9.18
M10	17.24	23.42	28.97	3.23	3.80	4.45	64.25	25.16	1.96
M11	18.69	26.37	30.29	3.39	4.16	4.66	60.00	11.60	3.56
M12	23.44	24.90	32.04	3.86	3.97	4.84	43.80	15.12	2.38
M13	24.70	21.01	23.34	3.91	3.64	4.08	34.33	28.76	21.75
M14	25.88	22.48	27.63	4.09	3.80	4.55	28.05	23.28	9.56
M15	25.94	22.81	27.08	4.02	3.85	4.55	27.26	24.63	12.25
M16	61.23	36.62	28.49	5.80	4.35	3.64	4.60	33.95	13.21
M17	24.03	24.17	26.66	3.87	3.48	4.03	38.49	31.29	14.20
M18	24.25	24.95	29.81	3.92	4.03	4.72	36.77	11.27	1.87
M19	20.97	24.16	28.88	3.59	4.01	4.47	50.73	16.21	4.59
M20	23.07	25.23	29.42	3.85	4.04	4.69	42.75	10.52	6.00
M21	23.28	26.47	31.13	3.86	4.17	4.79	41.21	2.47	6.82
M22	25.95	23.87	30.32	4.26	4.05	4.84	30.34	31.52	15.74
M23	26.63	31.29	38.15	4.38	4.70	5.48	30.69	7.21	14.84
M24	42.07	26.58	23.67	4.56	3.74	3.99	10.45	6.32	6.18
M25	25.36	22.44	28.02	4.01	3.75	4.50	31.05	17.80	0.75
M26	31.20	22.28	26.45	4.25	3.66	4.43	5.11	10.30	2.37
M27	25.36	20.36	23.47	4.08	3.55	4.15	30.44	32.00	15.48
M28	22.35	22.46	27.26	3.77	3.79	4.43	44.86	24.20	0.28
M29	23.39	21.81	25.11	3.76	3.66	4.15	41.44	19.28	1.02
M30	23.79	22.26	25.30	3.85	3.72	4.27	39.03	17.08	2.97
M31	25.00	21.26	26.70	3.99	3.67	4.42	32.82	27.36	1.56
M32	21.40	21.26	28.42	3.69	3.67	4.57	48.50	27.36	0.79
M33	22.06	23.51	31.19	3.78	3.80	4.74	45.98	18.20	3.97
M34	22.07	22.21	28.73	3.79	3.71	4.53	45.86	23.07	1.44
M35	24.61	21.91	27.21	3.90	3.56	4.25	35.87	24.25	4.23
M36	26.49	32.93	34.01	4.32	4.52	5.01	28.95	0.76	4.23

Table 7.4 *Statistics analysis of modelling ASP using ANN to predict SS*

Model Number	MSE $(mg/l)^2$			AAE (mg/l)			Correlation (R) %		
	Training	Validation	Testing	Training	Validation	Testing	Training	Validation	Testing
M1	193.13	150.29	289.40	10.39	9.76	13.51	62.38	50.00	24.85
M2	230.48	134.09	222.94	11.31	8.67	11.51	54.68	58.53	42.25
M3	231.83	181.01	265.34	11.90	11.22	13.54	50.80	34.30	29.26
M4	208.85	168.55	304.40	11.12	10.22	14.26	57.70	43.00	27.57
M5	204.70	153.68	281.85	10.92	9.66	13.30	58.99	47.71	27.05
M6	223.18	158.37	293.29	11.22	9.77	13.52	53.54	45.63	25.80
M7	349.28	208.04	299.02	13.60	10.51	12.82	39.04	32.56	10.87
M8	229.70	181.37	298.21	11.53	10.83	14.21	52.50	34.85	20.29
M9	213.63	168.79	302.51	11.02	10.68	14.42	57.73	41.20	19.08
M10	182.98	180.55	299.09	10.21	10.61	13.69	65.27	40.20	21.26
M11	201.09	188.03	313.90	10.81	11.17	14.55	60.43	35.79	20.22
M12	257.44	173.46	305.99	12.49	10.33	14.13	43.66	38.24	24.25
M13	276.66	179.95	252.63	13.01	10.91	13.14	32.92	30.24	34.56
M14	289.25	182.70	293.17	13.49	11.13	14.43	25.74	31.89	21.48
M15	286.89	208.72	303.62	13.22	12.13	14.94	27.49	18.21	18.54
M16	599.50	296.63	393.30	17.40	12.28	13.74	3.76	11.29	33.49
M17	257.45	292.78	465.13	12.57	13.49	18.08	41.47	9.94	18.43
M18	284.23	194.47	315.41	13.21	11.48	15.09	30.07	25.00	18.57
M19	232.24	213.83	350.65	11.76	12.19	15.37	50.66	17.50	4.47
M20	263.48	223.05	383.94	12.83	12.39	16.69	38.85	17.87	5.74
M21	265.02	218.69	353.73	12.87	12.33	15.85	38.46	14.48	0.59
M22	277.36	198.97	313.42	12.99	11.79	15.04	32.42	25.82	18.30
M23	294.44	197.43	290.55	13.34	10.99	13.78	28.80	25.49	22.67
M24	521.02	262.74	310.98	16.84	11.47	12.23	2.17	39.55	22.25
M25	283.23	191.88	315.33	13.06	11.46	14.98	30.29	23.63	10.76
M26	344.58	182.18	296.77	13.41	10.73	14.44	12.11	25.05	7.66
M27	278.81	190.99	303.83	13.23	11.54	14.89	31.81	30.84	13.61
M28	248.72	199.59	324.52	12.22	11.67	14.91	44.51	30.19	8.50
M29	254.75	200.03	357.18	12.60	11.71	15.59	42.18	30.43	9.21
M30	267.55	191.47	313.51	12.85	11.33	14.79	36.85	28.18	5.64
M31	274.24	194.62	334.59	12.82	11.47	15.48	34.11	24.78	2.93
M32	248.42	202.96	328.51	12.39	11.95	15.18	44.52	23.88	6.84
M33	247.58	215.28	367.77	12.45	11.97	16.03	44.85	23.71	0.84
M34	248.99	193.59	338.51	12.40	11.49	15.16	44.25	30.67	3.53
M35	283.48	191.98	339.17	12.70	11.10	14.85	33.73	27.87	3.31
M36	292.63	298.66	401.52	14.34	14.17	17.23	29.65	4.51	4.78

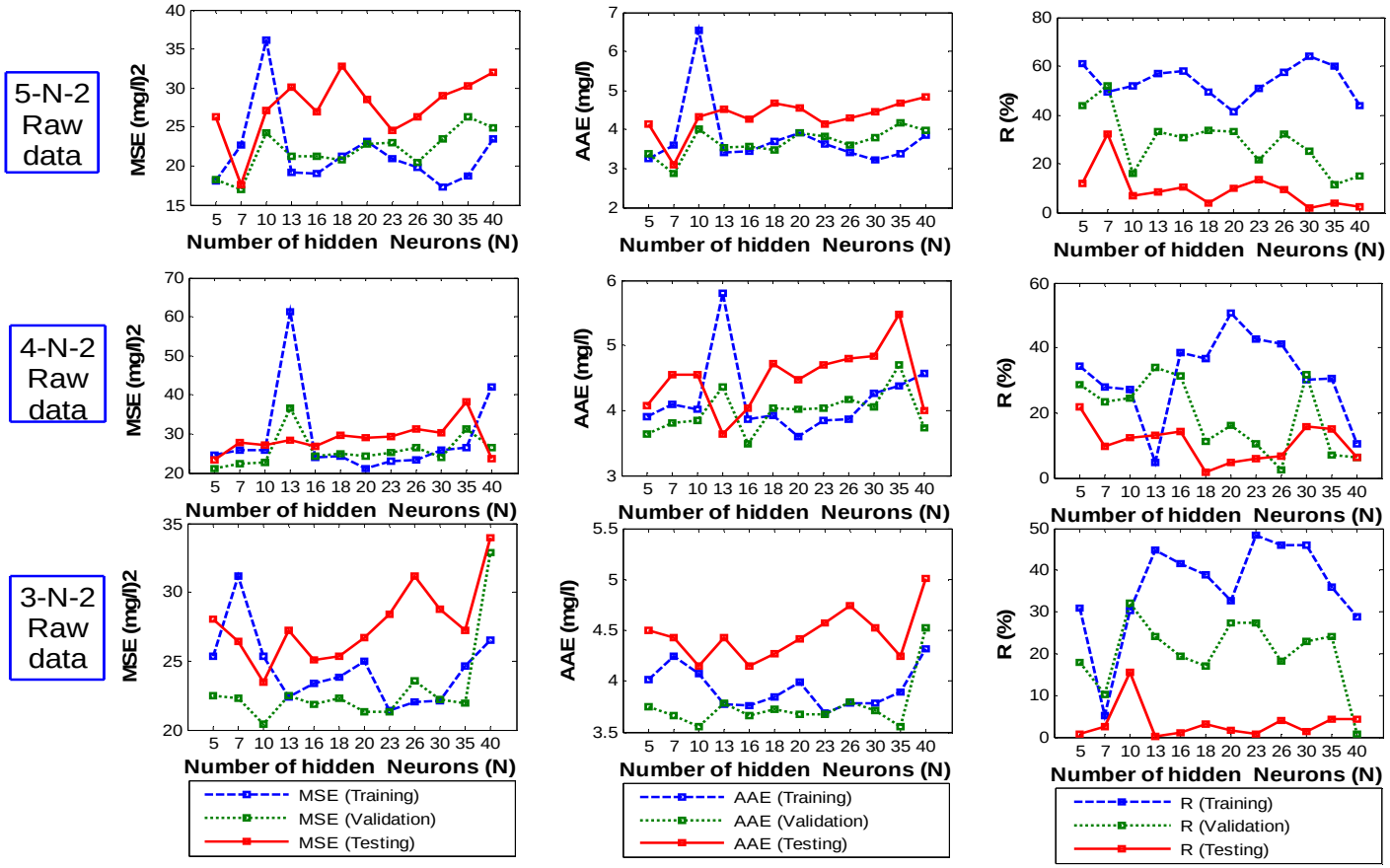


Figure 7.3 *Selection of hidden neurons using raw data (effluent BOD)*

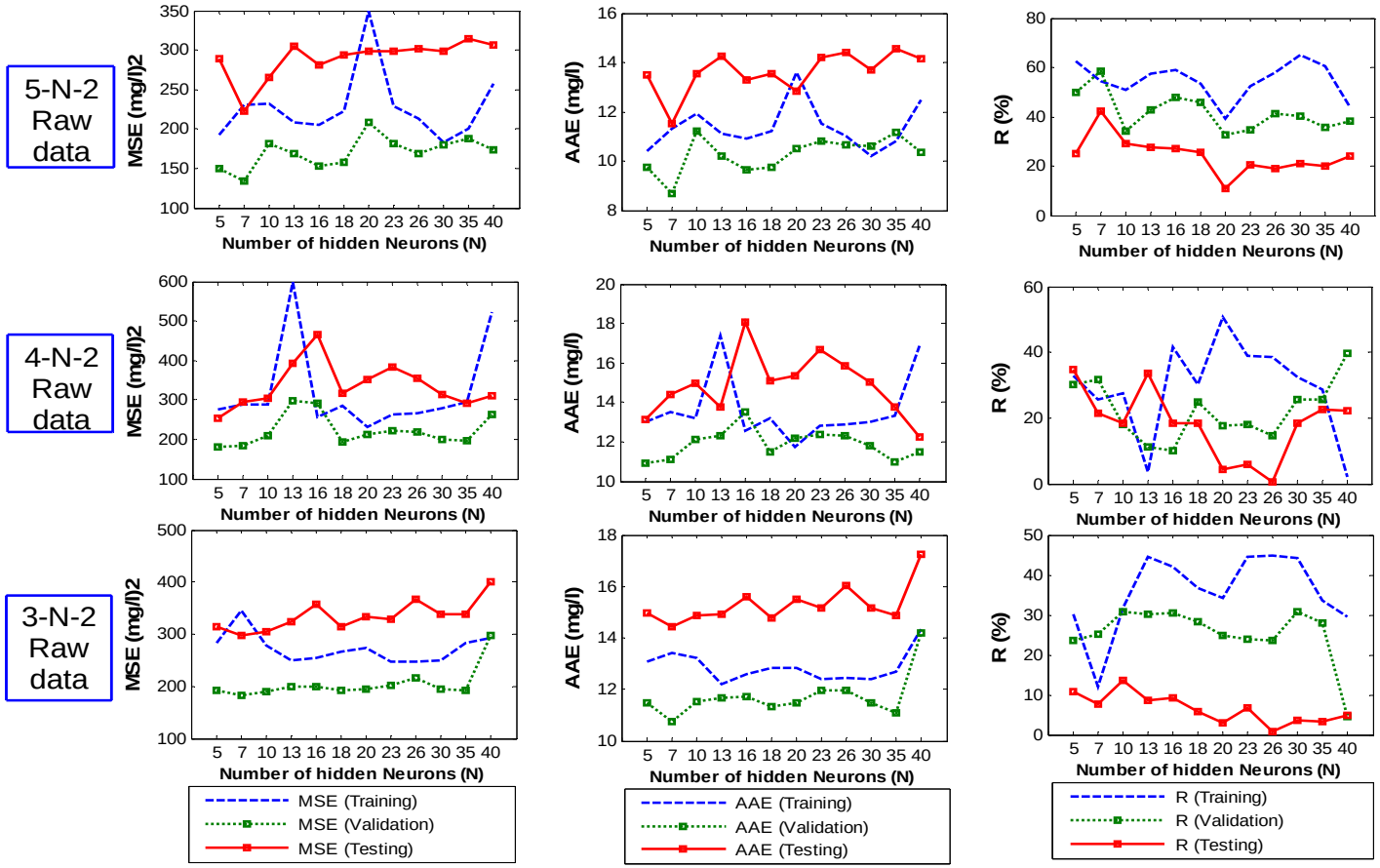


Figure 7.4 *Selection of hidden neurons using raw data (effluent SS)*

7.3.2 Artificial neural networks using the features of the data

Table 7.5 shows the details of the 36 models trained with the features of the data. In the first 12 models, M1(F)-M12(F), there were 5 inputs and 2 outputs with different number of neurons in the hidden layer were trained and tested. As stated previously, the early stopping technique was employed in which the training process was stopped when the validation error started to increase. This ensures that overfitting dose not occurs.

The next set of 12 models, denoted by M13(F) to M24(F), had 4 inputs (BOD-Load, RAS-MLSS, F/M, and Temperature), 2 outputs (BOD_5 and SS) with different numbers of neurons in the hidden layer. The last set of 12 models, denoted by M25(F) to M36(F), have 3 inputs (BOD-Load, RAS-MLSS, F/M, and Temperature), 2 outputs (BOD_5 and SS) with different numbers of neurons in the hidden layer.

Table 7.6 and Table 7.7, Figure 7.5 and Figure 7.6 show the comparison of the error criteria of the models developed in the three cases during training, validation and testing for effluent BOD_5 and Effluent SS respectively.

7.3.3 Comparison between the models

The evaluation criteria (i.e. MSE, AAE, and R) were calculated for each architecture and the number of hidden neurons corresponding to the best performance during testing data set for each case was selected. Since the training was stopped according to the validation error, the number of epochs varied for each architecture. Table 7.8 shows the optimal architecture for each model and the number of epochs corresponding to the selected model.

Table 7.5 *The structure of the developed ANN models using KSOM extracted features*

Inputs	Model Number	Number of neurons in the hidden layer
BOD-Load DO RAS-MLSS F/M T	M1(F)	5
	M2(F)	7
	M3(F)	10
	M4(F)	13
	M5(F)	16
	M6(F)	18
	M7(F)	20
	M8(F)	23
	M9(F)	26
	M10(F)	30
	M11(F)	35
	M12(F)	40
BOD-Load RAS-MLSS F/M T	M13(F)	5
	M14(F)	7
	M15(F)	10
	M16(F)	13
	M17(F)	16
	M18(F)	18
	M19(F)	20
	M20(F)	23
	M21(F)	26
	M22(F)	30
	M23(F)	35
	M24(F)	40
BOD-Load RAS-MLSS F/M	M25(F)	5
	M26(F)	7
	M27(F)	10
	M28(F)	13
	M29(F)	16
	M30(F)	18
	M31(F)	20
	M32(F)	23
	M33(F)	26
	M34(F)	30
	M35(F)	35
	M36(F)	40

Table 7.6 *Statistics analysis of modelling ASP using KSOM-ANN to predict effluent BOD_5*

Model Number	MSE $(mg/l)^2$			AAE (mg/l)			Correlation (R) %		
	Training	Validation	Testing	Training	Validation	Testing	Training	Validation	Testing
M1F	12.62	14.50	12.14	2.67	2.74	2.57	74.92	60.99	60.90
M2F	11.55	13.17	12.86	2.56	2.62	2.72	77.39	65.34	58.32
M3F	11.65	13.30	11.55	2.60	2.61	2.54	77.64	66.18	63.34
M4F	10.35	11.62	11.95	2.35	2.45	2.51	80.29	69.82	62.11
M5F	10.77	12.01	11.96	2.43	2.41	2.50	79.96	68.30	62.00
M6F	10.34	11.54	11.37	2.32	2.41	2.51	81.80	68.31	65.39
M7F	10.67	12.30	12.45	2.38	2.41	2.39	79.59	67.35	60.26
M8F	11.53	13.45	12.64	2.52	2.61	2.52	77.75	63.30	59.22
M9F	11.76	14.00	13.22	2.55	2.60	2.66	77.50	62.47	56.37
M10F	10.25	11.44	12.92	2.32	2.40	2.60	80.51	69.55	59.42
M11F	11.48	12.64	12.84	2.53	2.54	2.65	77.82	67.36	58.54
M12F	11.80	13.04	11.04	2.56	2.55	2.40	77.44	66.70	65.46
M13F	13.72	15.11	14.89	2.75	2.90	3.00	72.21	56.58	49.17
M14F	14.74	15.21	13.22	2.88	2.74	2.62	69.82	57.70	56.16
M15F	14.44	16.28	13.17	2.79	2.82	2.64	70.99	54.25	56.59
M16F	13.03	13.95	14.15	2.69	2.68	2.77	73.81	60.69	54.40
M17F	11.81	13.46	12.71	2.55	2.60	2.61	77.06	63.42	58.50
M18F	15.37	15.65	13.09	2.96	2.77	2.64	67.96	54.58	56.80
M19F	10.19	12.12	13.85	2.32	2.45	2.71	80.40	66.97	56.79
M20F	10.76	12.68	12.55	2.37	2.54	2.55	79.47	65.26	60.66
M21F	11.42	13.71	13.30	2.43	2.60	2.62	78.04	62.99	56.00
M22F	10.61	12.43	12.94	2.37	2.49	2.64	79.69	65.58	59.46
M23F	11.22	12.83	12.92	2.43	2.53	2.60	78.76	65.66	58.69
M24F	10.90	12.70	12.74	2.39	2.51	2.54	79.02	65.07	59.35
M25F	22.62	15.50	14.86	3.74	2.86	3.01	44.89	55.44	49.67
M26F	21.50	15.75	16.44	3.64	2.95	3.14	48.88	53.27	42.96
M27F	21.05	15.53	16.31	3.58	2.77	3.09	50.68	55.19	44.04
M28F	12.91	14.57	14.47	2.68	2.68	2.73	74.22	61.51	52.66
M29F	20.23	15.56	15.73	3.50	2.82	2.97	54.58	54.98	45.91
M30F	15.34	14.63	14.20	2.91	2.75	2.85	68.08	58.89	52.51
M31F	15.86	14.46	14.50	2.97	2.69	2.80	66.96	60.57	51.48
M32F	13.51	14.35	14.25	2.66	2.70	2.80	72.67	60.00	53.35
M33F	13.48	12.95	13.15	2.68	2.64	2.71	72.56	65.42	57.96
M34F	13.30	13.92	13.50	2.72	2.69	2.73	74.00	61.86	58.77
M35F	13.26	13.64	12.79	2.65	2.62	2.58	73.34	63.10	59.01
M36F	15.83	14.53	13.72	2.99	2.77	2.66	67.42	59.18	54.76

Table 7.7 *Statistics analysis of modelling ASP using KSOM-ANN to predict effluent SS*

Model Number	MSE $(mg/l)^2$			AAE (mg/l)			Correlation (R) %		
	Training	Validation	Testing	Training	Validation	Testing	Training	Validation	Testing
M1F	140.37	130.17	169.34	8.77	8.39	9.44	74.35	58.82	62.47
M2F	136.61	130.10	205.00	8.65	8.31	11.73	75.15	58.45	55.62
M3F	136.28	123.06	211.00	8.58	7.87	11.64	75.49	63.00	60.28
M4F	122.03	106.96	205.00	7.99	7.54	11.54	78.30	67.49	60.17
M5F	128.07	108.13	204.00	8.14	7.56	11.52	77.41	66.73	58.90
M6F	121.04	106.20	162.73	7.62	7.51	8.49	78.32	67.53	65.04
M7F	127.01	117.28	210.00	8.06	7.79	11.43	77.21	63.70	59.06
M8F	135.78	128.58	217.00	8.44	8.26	11.56	75.46	59.07	56.98
M9F	137.77	124.72	175.95	8.53	7.89	9.28	75.25	61.54	60.02
M10F	124.91	115.93	167.91	7.95	7.74	8.86	77.61	64.70	62.02
M11F	135.36	120.92	179.09	8.39	7.95	9.53	75.52	62.06	58.13
M12F	138.51	119.07	160.83	8.56	7.88	8.90	74.95	63.64	64.85
M13F	157.86	132.02	191.50	9.29	8.62	10.24	70.29	56.62	54.85
M14F	158.86	134.87	179.33	9.27	8.36	9.48	70.34	57.16	59.27
M15F	158.76	141.77	177.63	9.12	8.59	9.22	70.73	55.60	61.61
M16F	151.43	128.69	182.43	8.91	8.41	9.51	71.76	58.17	57.29
M17F	137.72	124.66	166.98	8.48	8.05	9.15	75.07	61.16	62.91
M18F	164.93	143.73	176.18	9.49	8.78	9.34	68.89	52.80	60.63
M19F	124.73	118.18	174.04	8.05	7.85	9.36	77.51	63.72	60.31
M20F	129.55	114.77	168.12	8.09	7.76	8.84	76.69	64.30	62.91
M21F	136.43	125.43	180.13	8.35	7.99	9.27	74.92	61.08	57.89
M22F	127.99	118.34	168.60	8.07	7.88	9.09	76.94	62.91	61.74
M23F	134.52	114.20	177.77	8.25	7.80	9.39	75.69	64.57	59.53
M24F	130.24	122.60	176.79	8.16	7.99	8.91	76.50	61.55	59.40
M25F	240.08	148.83	188.53	12.03	8.97	10.17	47.84	50.17	55.64
M26F	233.39	149.47	196.76	11.92	9.22	10.43	49.81	48.65	52.58
M27F	223.14	157.78	188.54	11.49	9.47	9.91	53.39	50.00	55.83
M28F	152.28	132.67	191.45	9.11	8.11	9.53	71.59	59.89	55.13
M29F	220.48	152.86	200.09	11.34	9.02	9.97	55.20	49.27	51.28
M30F	178.98	132.81	184.73	9.87	8.40	9.78	65.22	57.05	56.23
M31F	181.61	142.06	189.33	9.90	8.49	9.57	64.75	54.22	55.17
M32F	156.15	128.78	186.25	9.07	8.07	9.54	70.70	59.36	56.13
M33F	159.46	125.12	162.75	9.12	8.07	9.22	69.79	60.05	61.15
M34F	155.02	123.97	178.04	9.06	8.95	9.51	71.36	61.40	60.60
M35F	154.49	114.66	160.24	9.03	8.06	9.05	70.39	58.06	61.40
M36F	178.73	135.17	187.52	9.92	8.37	9.38	65.83	55.67	55.59

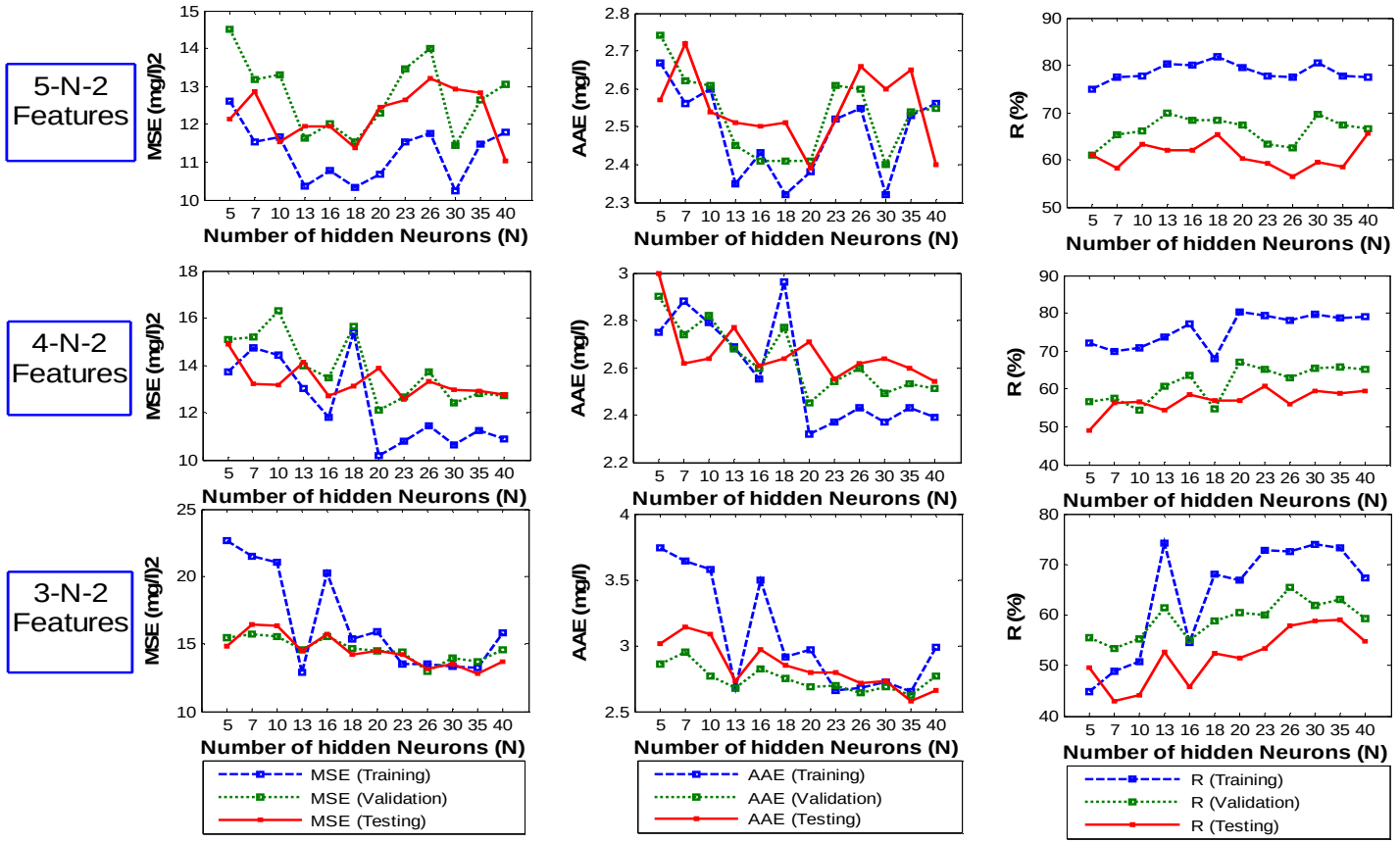


Figure 7.5 *Selection of hidden neurons using KSOM features (effluent BOD)*

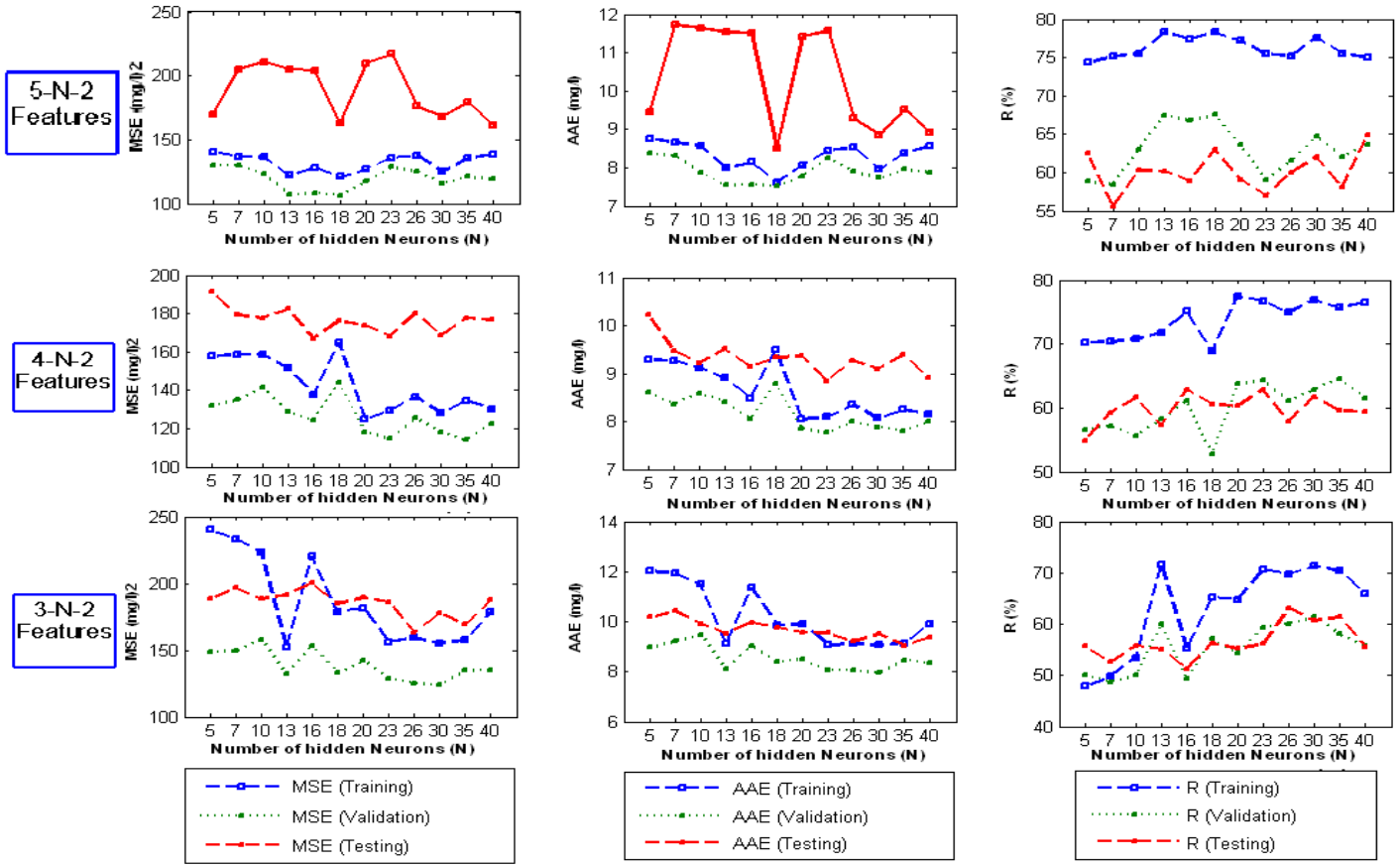


Figure 7.6 *Selection of hidden neurons using KSOM features (effluent SS)*

Table 7.8 *Characteristics of neural network models*

Model	Architecture	Number of epochs
M I	5-7-2	11
M II	4-5-2	12
M III	3-10-2	9
M I (F)	5-18-2	22
M II (F)	4-23-2	33
M III (F)	3-35-2	23

In order to find the best model, the evaluation criteria obtained during training, validation and testing for the best model in each case were compared. Table 7.9 and Table 7.10 show the values of MSE, AAE, and R for BOD_5 and SS respectively. The highlighted values are the best performance achieved, that is the lowest MSE and highest correlation R (%). The same results are represented in a graphical form in Figure 7.7. From Table 7.9, 7.10 and Figure 7.7, it is clear that using the features of the data produced better performance than using the raw data itself to train the models.

This confirms the fact that KSOM features improve the performance of ANN. Also, From Table 7.9 and Table 7.10, and Figures 7.7, it is clear that model number M1 (F) (M6F in Table 7.6) has the best performance. The 18-node architecture can be taken as a compromise best structure since no significant improvement in all the three performance criteria occurs when the number of neurons is increased beyond 18.

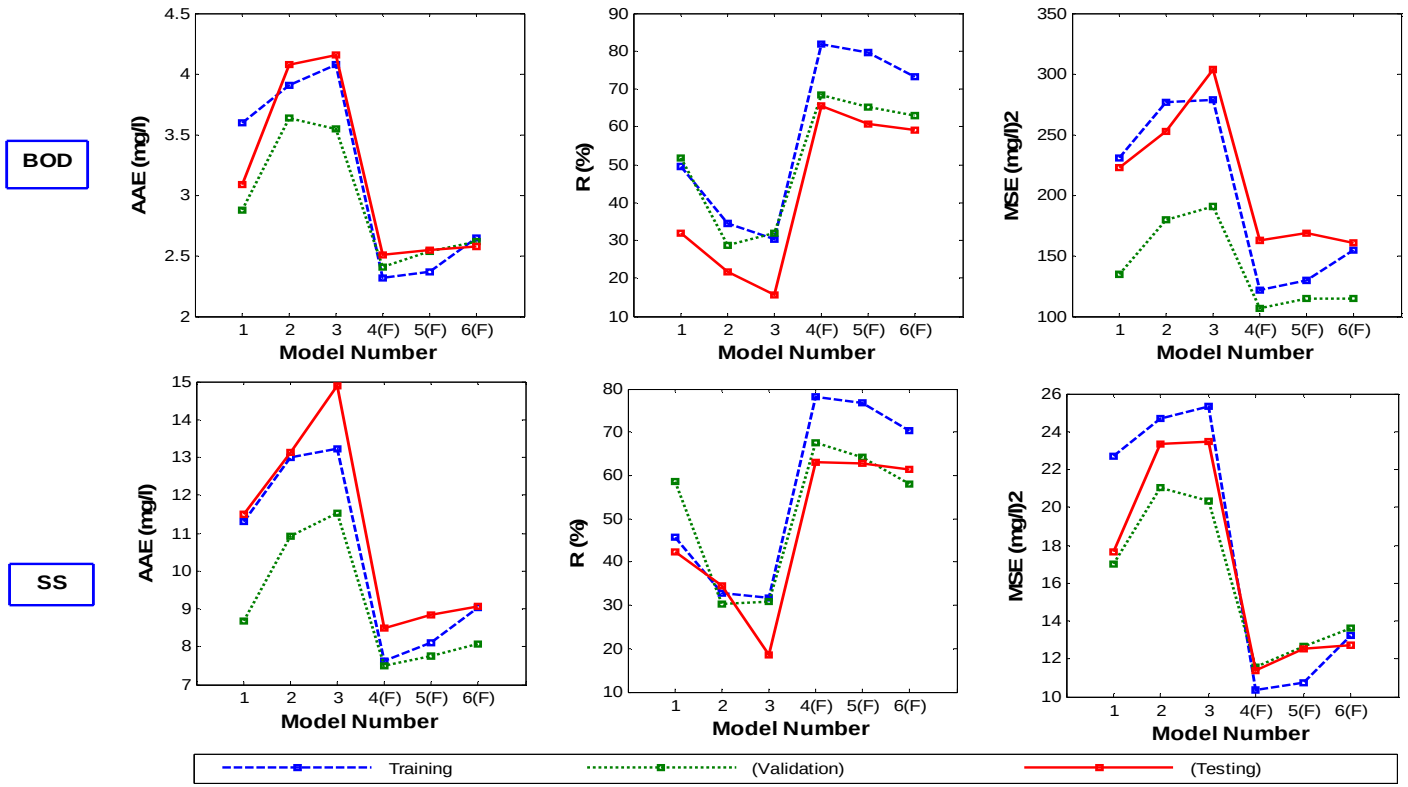


Figure 7.7 *Comparison between models*

Table 7.9 *Comparison of statistical coefficient for effluent BOD for the best model with each category of models*

Model Number	Architecture	MSE $(mg/l)^2$			AAE (mg/l)			Correlation (R) %		
		Training	Validation	Testing	Training	Validation	Testing	Training	Validation	Testing
M I	5-7-2	22.73	17.01	17.64	3.60	2.88	3.09	49.54	51.71	32.02
M II	4-5-2	24.70	21.01	23.34	3.91	3.64	4.08	34.33	28.76	21.75
M III	3-10-2	25.36	20.36	23.47	4.08	3.55	4.15	30.44	32.00	15.48
M I (F)	5-18-2	10.34	11.54	11.37	2.32	2.41	2.51	81.80	68.31	65.49
M II (F)	4-23-2	10.76	12.68	12.55	2.37	2.54	2.55	79.47	65.26	60.66
M III (F)	3-35-2	13.26	13.64	12.74	2.65	2.62	2.58	73.34	63.10	59.01

Table 7.10 *Comparison of statistical coefficient for effluent SS for the best model with each category of models*

Model Number	Architecture	MSE $(mg/l)^2$			AAE (mg/l)			Correlation (R) %		
		Training	Validation	Testing	Training	Validation	Testing	Training	Validation	Testing
M I	5-7-2	230.48	134.0	222.94	11.31	8.67	11.51	45.68	58.53	42.25
M II	4-5-2	276.66	179.95	252.63	13.01	10.91	13.14	32.92	30.24	34.56
M III	3-10-2	278.81	190.99	303.83	13.23	11.54	14.89	31.81	30.84	18.50
M I (F)	5-18-2	121.04	106.20	162.73	7.62	7.51	8.49	78.32	67.53	65.04
M II (F)	4-23-2	129.55	114.77	168.12	8.09	7.76	8.84	76.69	64.30	62.91
M III (F)	3-35-2	154.49	114.66	160.24	9.03	8.06	9.05	70.39	58.06	61.40

It is also evident from Tables 7.9 and 7.10 and Figure 7.7 that for each specific number of neurons in the hidden layer, the performance of the model is better using the features of the raw data than using the raw data itself. The relative superiority of the features-derived models is also evident when both the MSE and AAE are considered. This is because the features have eliminated the noise in the raw data set, which affected the performance of the basic ANN.

Because the ANN model with 18 neurons in the hidden layer using the features method, Model M I (F) has the best performance, further analysis was only done with this model. Table 7.6 and 7.7 show clearly that the performance results were good. However the best way of assessing the model prediction is by looking at the predicting plots and analysing wither they have good agreement. Figure 7.8 and 7.9 show in the time series plots the comparison between the model predictions with the targets for BOD_5 and SS during training, validation and Testing. Visual investigations of these figures demonstrates that the model is able to predict the output during operation conditions as the two lines (predicted and observed) are overlapped in the majority of the points. Figure 7.10 and 7.11 show the residuals plots of the model during training, validation and testing. Figures 7.12, 7.13, and 7.14 show scatter plot of the measured and predicted BOD_5 during training, validation and testing respectively. Figures 7.15, 7.16, and 7.17 show the scatter plot of the measured and observed of SS during training, validation and testing respectively. Most of the data points are around the predicted equals observed line.

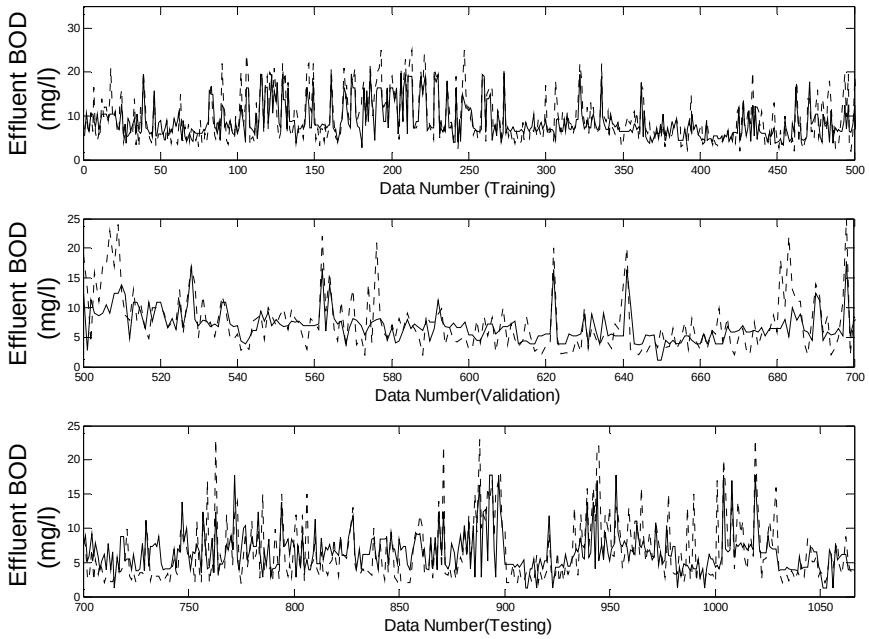


Figure 7.8 *Comparison of the observed (doted) and predicted (solid) effluent BOD values during training, validation and testing*

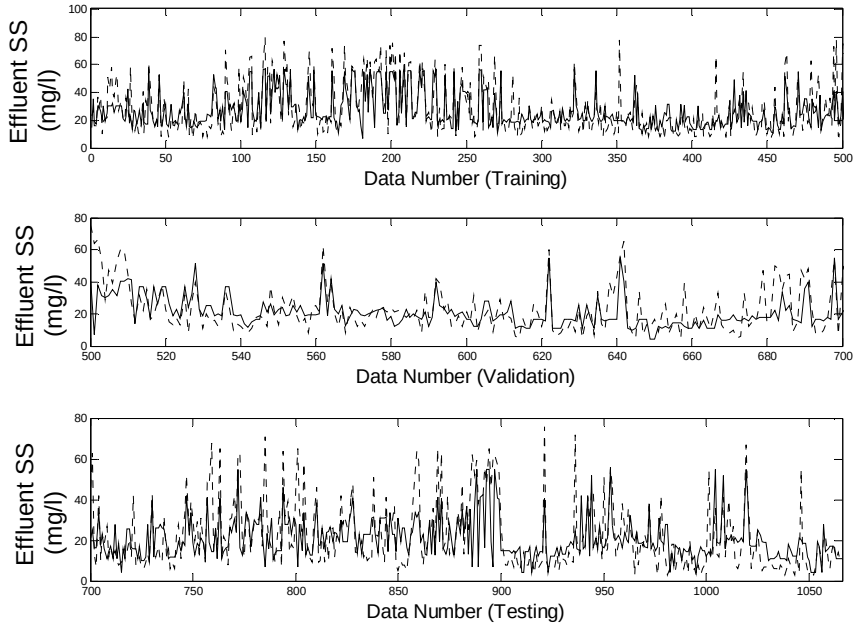


Figure 7.9 *Comparison of the observed (doted) and predicted (solid) effluent SS values during training, validation and testing*

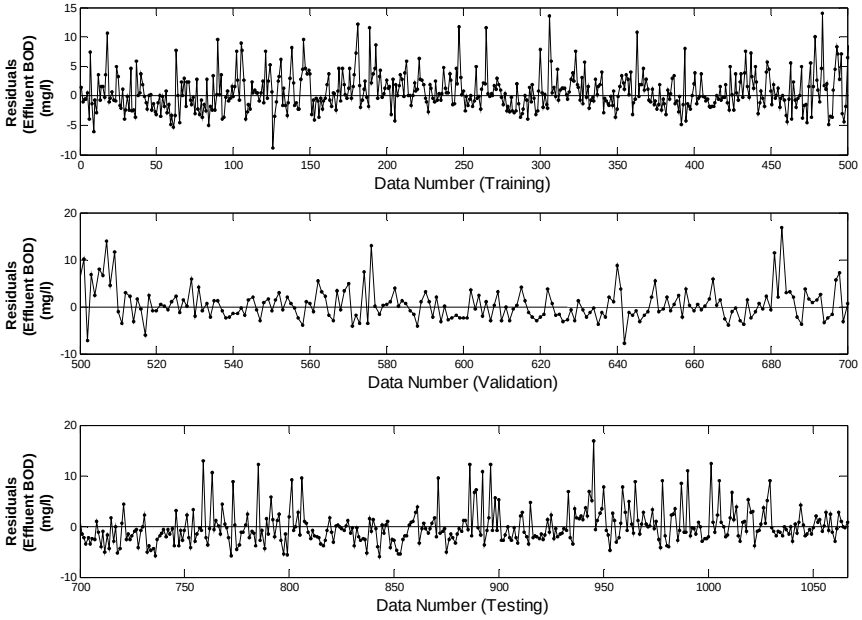


Figure 7.10 *Residuals of the model during training validation and testing*

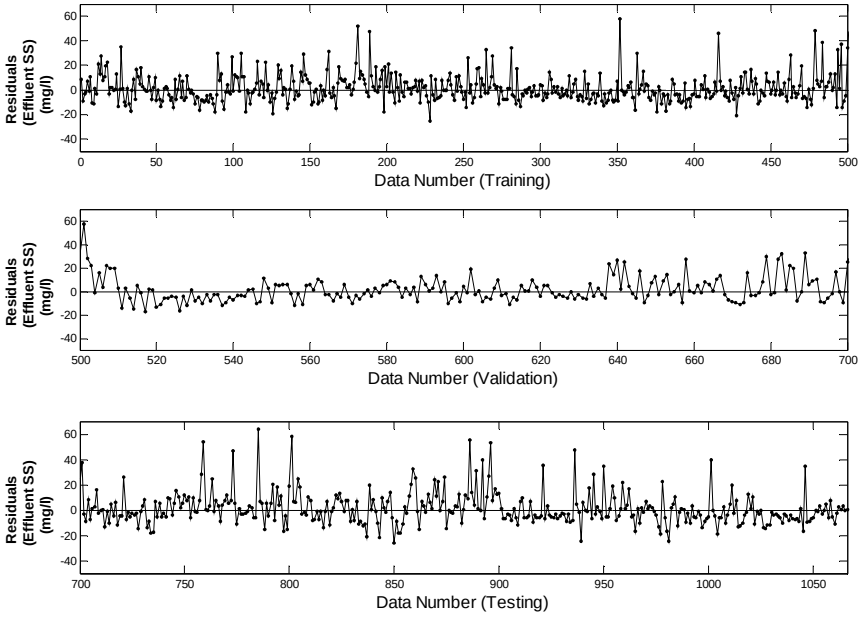


Figure 7.11 *Residuals of the model during training validation and testing*

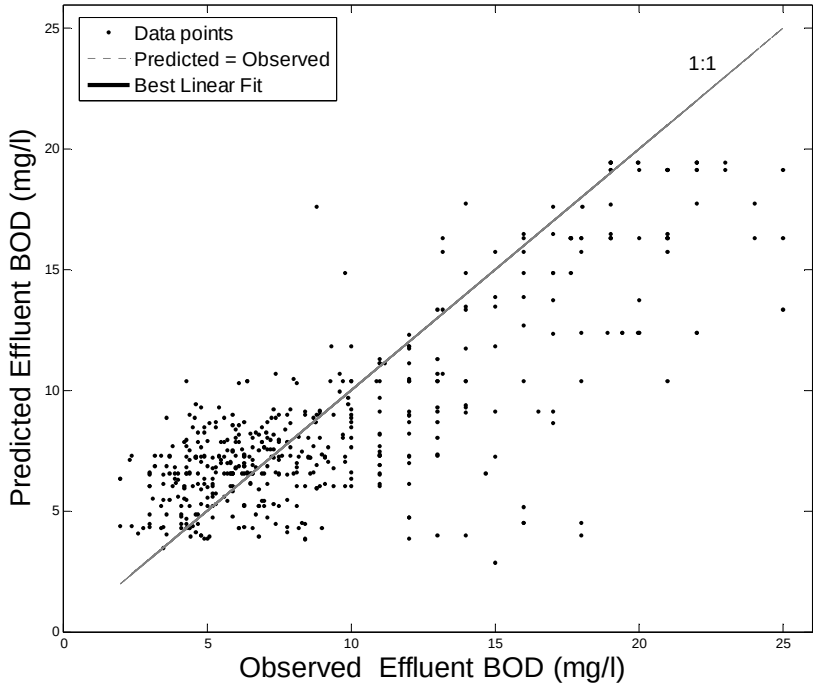


Figure 7.12 *the performance of the model in predicting the effluent BOD during training*

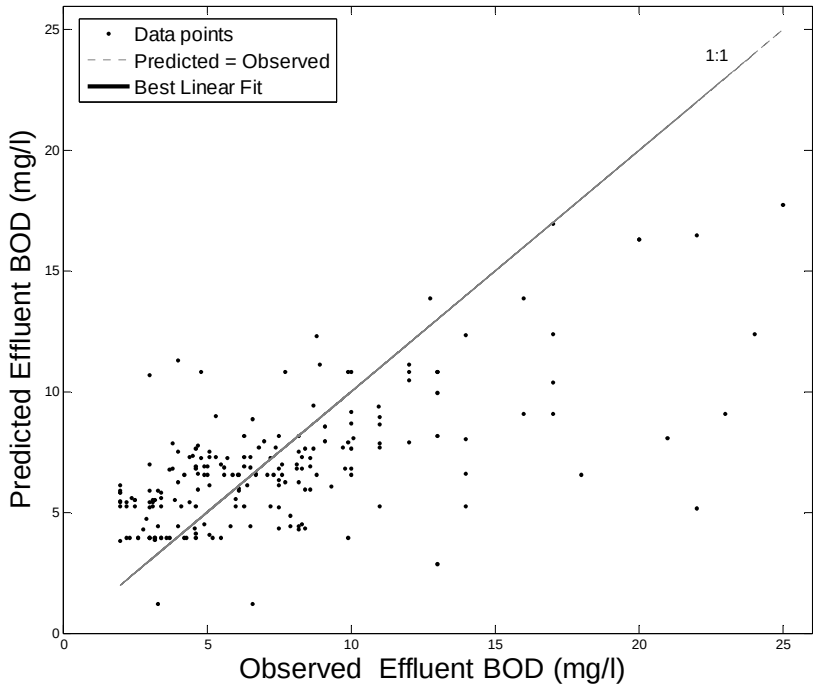


Figure 7.13 *The performance of the model in predicting the effluent BOD during validation*

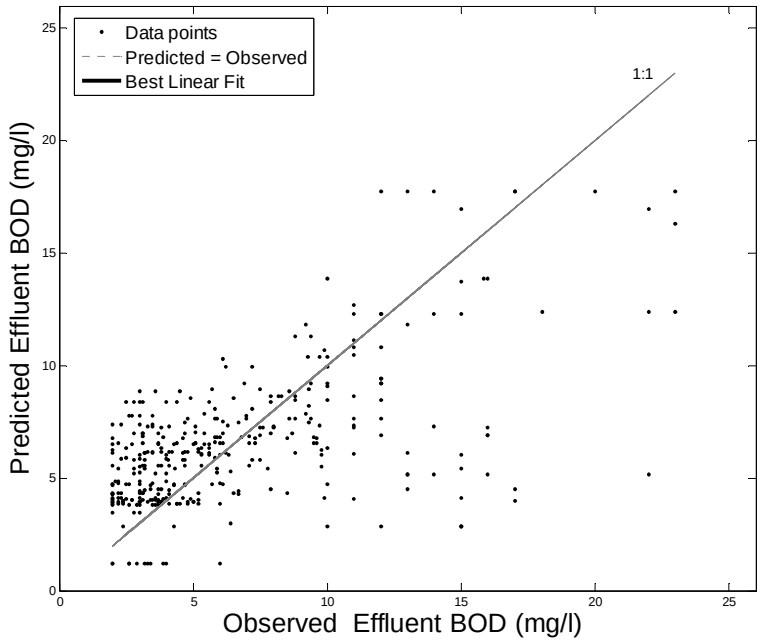


Figure 7.14 *The performance of the model in predicting the effluent BOD during testing*

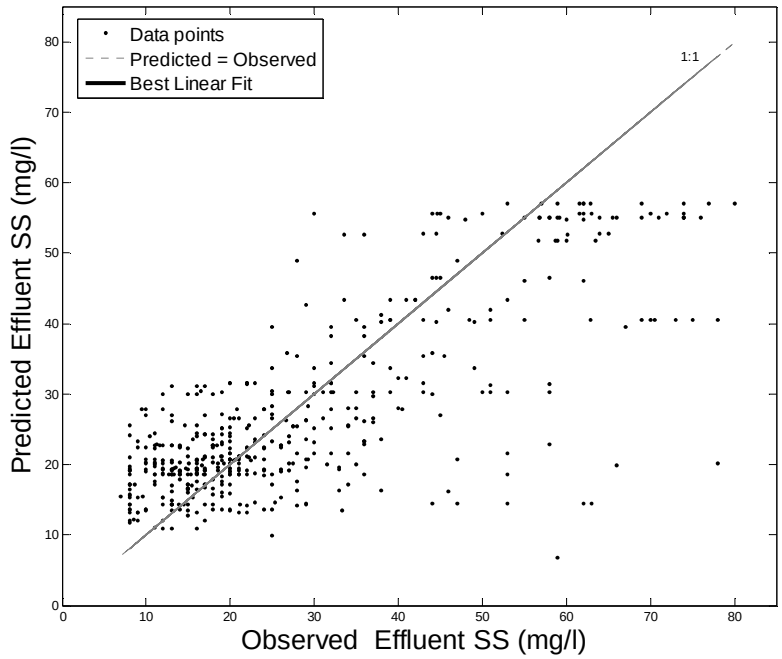


Figure 7.15 *The performance of the model in predicting the effluent SS during training*

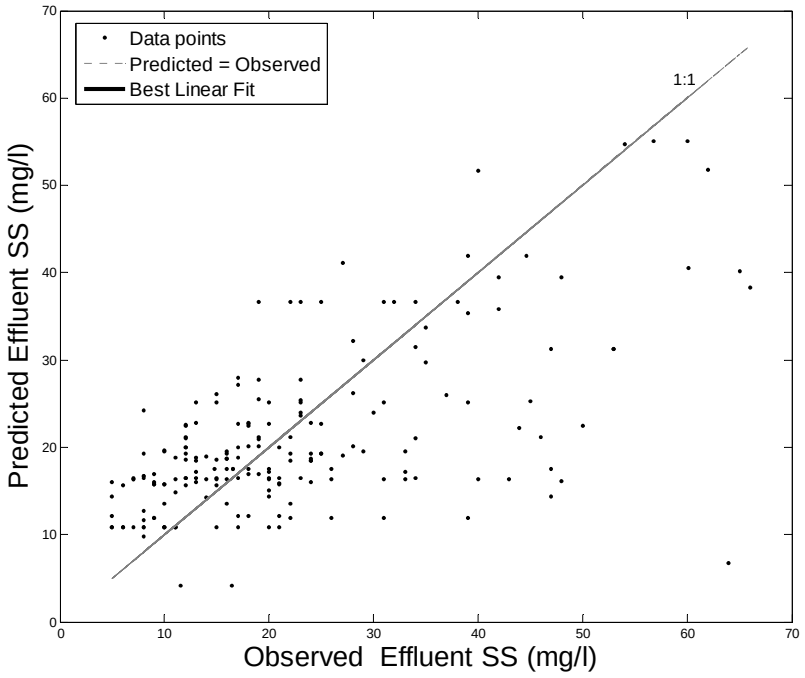


Figure 7.16 *the performance of the model in predicting the effluent SS during validation*

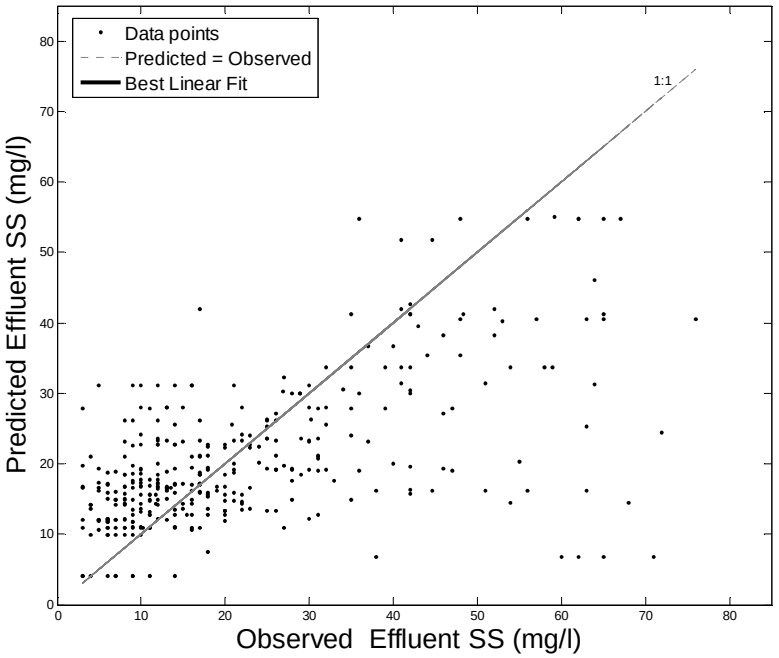


Figure 7.17 *the performance of the model in predicting the effluent SS during testing*

7.4 Conclusion

The current application used a new methodology based on a hybrid supervised-unsupervised artificial neural network to improve the performance of the basic backpropagation neural network method in modelling the activated sludge wastewater treatment plant. The method was applied to data taken from the Seafield wastewater treatment plant in Edinburgh, UK, during a period of about three years. Input variables were selected based on their correlation with the effluent BOD_5 and SS, which were the target prediction variable. Several ANN models with different numbers of neurons in the hidden layers were developed. For each model, two types of data were used, the first one is the raw data set and the second one is the extracted features of the raw data using the Kohonen self-organising map. The results showed that the models using the features were better than those using the raw data.

The findings prove the ability of KSOM to improve the performance of modelling using basic back-propagation neural networks, particularly when the available data are noisy, a common problem with the process data of wastewater treatment plants. Furthermore, the KSOM can readily deal with missing values in one or more of the input variables without significantly negative impacts on the accuracy of the model (see Rustum and Adeloye, 2007a). Results obtained prove that KSOM-ANN present a versatile tool in modelling ASP and provided an alternative methodology for predicting the performance of WWTPs.

The developed models have several advantages as they were able to predict the effluent BOD_5 and SS without explicit considering the mathematical relationship between the inputs and the outputs. However, it is necessary to underline the limit of using the methodology if the database is insufficient as it might lead to erroneous interpolations or restricted to narrow range of operating conditions. In other words, the developed models are not able to be applied to another plant with

input data outside the range of the inputs used for training. Therefore, care must be taken when extrapolating these results to other nonlinear systems. Furthermore, there are problems concerned with identifying the neural network elements and parameters. Examples of such elements and parameters are the number of hidden layers, the number of neurons in each layer, training function, and transfer function. Therefore, finding the best general model is time consuming as it is highly empirical, and the only way to optimize the models is by searching for the best network parameters iteratively.

CHAPTER 8

APPLICATION IV: MODELLING ASP USING HYBRID KSOM-ANFIS

8.1 Introduction

In the previous chapter, the possibility of modelling ASP using artificial neural network was investigated. The results indicated that the KSOM extracted features improve the performance of backpropagation multilayer perceptron ANN. The learning algorithms were found to be suitable for modelling ASP. However, the concept or knowledge cannot be clearly expressed in a human understandable way in the form of if-then rules. In contrast, Fuzzy logic systems (FLS) are more favourable, in that their behaviours can be explained using fuzzy rules. In addition, FLSs can easily be interpreted in human understandable terms rather than with numerical quantities. In other words, fuzzy logic models depend not only on black box such as ANN, but are also based on a combination of knowledge of the system and operational experience.

As discussed exhaustively in Chapter 3, the basic idea behind neuro-fuzzy combination is to design a system that uses a fuzzy system to represent knowledge in an interpretable manner and have the learning ability of neural network to adjust its membership functions and parameters in order to enhance the system performance. Consequently, the main drawbacks of both individual systems could be avoided, i.e., the black box nature of neural networks, and the problem of selecting suitable membership values for FLSs (Jang, 1993). The combination can constitute an interpretable model that is capable of learning, as NNs, and reasoning, as FLSs (Firat and Güngör, 2007). Using this technique makes it

possible to adjust the membership functions automatically from data by using neural network learning algorithms. The trained membership functions can provide a better understanding about the properties of the database.

Because human-determined membership functions using trial and error are subject to the differences from one person to another and from time to time, they are rarely optimal in terms of reproducing desired outputs. To solve this problem, in 1993, Jang and his colleagues started looking at the FLS as adaptive network (Adaptive Network Based on Fuzzy Inference System, ANFIS (Jang, 1993; Jang and sun, 1995; Jang et al., 1997). This technique joins the linguistic interpretation of FLS with the computational power of neural networks that can be trained through gradient algorithms such as Back-propagation. A first forward pass is performed to determine the network output and a second backward pass is performed to adjust the parameters for better approximation. They found that ANFIS could be easily implemented for a given input/output modelling technique as fully described in Chapter 3.

Consequently, due to the power of KSOM in enhancing the performance of ANN as demonstrated in the previous chapter, the last application of this work investigates the possibility of integrating the KSOM with ANFIS in order to improve the performance of the ANFIS trained with raw data.

8.2 Methodology

The methodology presented in Section 3.4 has been applied to develop a fuzzy model for the ASP. The same Seafield treatment plant data set used in the previous chapter has been used in this application. The developed models were trained and testing using MATLAB 7 programming language with Fuzzy Logic toolboxes (Version 7, Release 14, Mathworks, Inc). Kohonen Self Organizing

Maps were built and visualized using SOM Toolbox for MATLAB 5, developed at the Laboratory of Computer and Information Science (CIS) at Helsinki University of Technology. Supporting statistical analysis was conducted using Statistical Toolbox and various functions in MATLAB.

The key requirement to getting a good model is to choose the appropriate antecedent part variables. However, in activated sludge wastewater treatment plants , it is difficult to take into account the numerous factors that influence the specific bacterial growth rate and its metabolic activities. Therefore, the parameters were chosen from typical variables observed in the daily monitoring record of activated sludge process that used to asses in controlling the process.

It is expected that the accuracy of the model will be improved by increasing the number of the variables in the antecedent part and by increasing the number of membership functions in each variable. However, that leads to a large number of rules and consequence parameters (see Chapter 3 for more details). Therefore, the first question was how many inputs can be put in a single ANFIS model when a model is trained on a personal computer? The second question was how many fuzzy membership functions can be used in each antecedent part in order to divide the input space into fuzzy divisions? That is because the computation cost increases by increasing the number of inputs and the number of membership functions of each input, as revealed in Equation 3.42. In addition, the volume of the training data plays a role in deciding the number of inputs and the number of membership functions associated with each input. For example, the total number of modified parameters must not exceed the total number of training data.

Furthermore, the aim is to adjust the parameters in order to approximate the given samples with the least error and at the same time obtain good generalization ability. To do this, partitioning the available data into training, validation and

testing is performed in order to measure the generalization ability. This is because it is known that, a minimum training error does not necessarily correspond to a minimum testing error, or best generalization (Adeloye and De Munari, 2006).

Since the number of rules in ANFIS and consequently the number of its parameters increases exponentially with the number of inputs, five inputs was the maximum number of input parameters that could be used given the available data base. For example, using five inputs with three membership functions associated with each input, the total number of modified parameters, according to Equation 3.42 will be 1488 which exceeds the number of training data of 500. However, with two membership functions for each input, the total number of modified parameters will be 212, which is okay, because that will leave a significant number of degree of freedom.

To choose the parameters to include, the analysis of the correlation matrix was carried out to determine the most likely parameters for predicting the effluent BOD_5 and effluent SS, the tow target variables. The identified variables were the same as in the previous chapter.

To decide the optimal number of input parameters and the optimal number of membership functions associated with each input, several models have be developed using a variety of inputs and a variety of membership functions associated with each input taking into account that the priority is for inputs with high correlation with the effluent BOD_5 and effluent SS concentrations. Table 8.1 and Table 8.2 present the structure of the models developed and tested in the study using Gaussian membership function.

8.3 Results and Discussion

To ensure good generalization capability of the developed models, the available raw data (1066 data records) were divided into three subsets, training (500 data records), validation (200 data records) to insure that model do not over fit the training data, and testing (366 data records). As was the case with Application III, two scenarios were considered: (a) an ANFIS model on the raw data, and (b) a hybrid KSOM-ANFIS model based features of the extracted using KSOM.

Through evaluating the prediction capability of the developed ANFIS models (M1-M8) using several evaluation criteria such as mean square error (MSE), average absolute error (AAE) and correlation coefficient (R) (Table 8.3 and Table 8.4), it can be seen that the performance of the models was not good. Moreover, the models generated negative values for BOD_5 and SS on occasions as can be seen from Table 8.5 and 8.6. Previously, Miller (2006) encountered negative predictions when using ANFIS to predict the rainfall (precipitation) from raw (noisy) weather data (temperature and humidity). Miller (2006) suggested the use of another technique for pre-processing the data to improve the performance of ANFIS that would deal with the noise by replacing the missing values and omitting the outliers. The original work by Jang (1993) on the ANFIS was validated with noiseless data generated by functional equations; it is therefore not surprising that he did not encounter negative predictions.

Table 8.1 *The structure of the ANFIS models developed and tested in the study for predicting effluent BOD concentrations using the Gaussian membership function**

Model Number	Number of input parameters (N_{input})	Input Parameters	Number of membership functions in each input (N_{mf})	Number of linear Parameters $P_1=l*(N_{input}+1)$	Number of non-linear parameters $P_2=N_{input}*N_{mf}*2$	Total number of parameters $P=P_1+P_2$	Number of Fuzzy Rules $l=(N_{mf})^{Ninput}$
M1	5 (raw)	BOD Load, DO, RAS-MLSS, F/M, Tem.	2	192	20	212	32
M2	4 (raw)	BOD Load, RAS-MLSS, F/M, Tem.	3	405	24	429	81
M3	4(raw)	BOD Load, RAS-MLSS, F/M, Tem.	2	80	16	96	16
M4	3 (raw)	BOD Load, RAS-MLSS, F/M	6	864	36	900	216
M5	3 (raw)	BOD Load, RAS-MLSS, F/M	5	500	30	530	125
M6	3 (raw)	BOD Load, RAS-MLSS, F/M	4	256	24	280	64
M7	3 (raw)	BOD Load, RAS-MLSS, F/M	3	108	18	126	27
M8	3 (raw)	BOD Load, RAS-MLSS, F/M	2	32	12	44	8
M9	5 (features)	BOD Load, DO, RAS-MLSS, F/M, Tem.	2	192	20	212	32
M10	4 (features)	BOD Load, RAS-MLSS, F/M, Tem.	3	405	24	429	81
M11	4 (features)	BOD Load, RAS-MLSS, F/M, Tem.	2	80	16	96	16
M12	3 (features)	BOD Load, RAS-MLSS, F/M	6	864	36	900	216
M13	3 (features)	BOD Load, RAS-MLSS, F/M	5	500	30	530	125
M14	3 (features)	BOD Load, RAS-MLSS, F/M	4	256	24	280	64
M15	3 (features)	BOD Load, RAS-MLSS, F/M	3	108	18	126	27
M16	3 (features)	BOD Load, RAS-MLSS, F/M	2	32	12	44	8

* the choice of Gaussian membership function was because it has just two modified parameter, the centre and the width, hence it requires less number of training data.

Table 8.2 *The structure of the ANFIS models developed and tested in the study for predicting effluent SS using the Gaussian membership functions*

Model Number	Number of input parameters (N_{input})	Input Parameters	Number of membership functions in each input (N_{mf})	Number of linear Parameters $P_1=l*(N_{input}+1)$	Number of non-linear parameters $P_2=N_{input}*N_{mf}*2$	Total number of parameters $P=P_1+P_2$	Number of Fuzzy Rules $l=(N_{mf})^{Ninput}$
M1	5 (raw)	BOD Load, DO, RAS-MLSS, F/M, Tem.	2	192	20	212	32
M2	4 (raw)	BOD Load, RAS-MLSS, F/M, Tem.	3	405	24	429	81
M3	4(raw)	BOD Load, RAS-MLSS, F/M, Tem.	2	80	16	96	16
M4	3 (raw)	BOD Load, RAS-MLSS, F/M	6	864	36	900	216
M5	3 (raw)	BOD Load, RAS-MLSS, F/M	5	500	30	530	125
M6	3 (raw)	BOD Load, RAS-MLSS, F/M	4	256	24	280	64
M7	3 (raw)	BOD Load, RAS-MLSS, F/M	3	108	18	126	27
M8	3 (raw)	BOD Load, RAS-MLSS, F/M	2	32	12	44	8
M9	5 (features)	BOD Load, DO, RAS-MLSS, F/M, Tem.	2	192	20	212	32
M10	4 (features)	BOD Load, RAS-MLSS, F/M, Tem.	3	405	24	429	81
M11	4 (features)	BOD Load, RAS-MLSS, F/M, Tem.	2	80	16	96	16
M12	3 (features)	BOD Load, RAS-MLSS, F/M	6	864	36	900	216
M13	3 (features)	BOD Load, RAS-MLSS, F/M	5	500	30	530	125
M14	3 (features)	BOD Load, RAS-MLSS, F/M	4	256	24	280	64
M15	3 (features)	BOD Load, RAS-MLSS, F/M	3	108	18	126	27
M16	3 (features)	BOD Load, RAS-MLSS, F/M	2	32	12	44	8

Table8.3 *The performance of the ANFIS models to predict effluent BOD_5*

Model Number	Number of input parameters (N)	MSE $(mg/l)^2$			AAE (mg/l)			Correlation coefficient (%)		
		Training	Validation	Testing	Training	Validation	Testing	Training	Validation	Testing
M1	5 (raw)	16.55	14.97	14.88	3.16	2.79	2.94	64.04	56.15	49.65
M2	4 (raw)	17.44	13.41	14.00	3.05	2.57	2.85	62.06	61.99	53.05
M3	4(Raw)	23.29	18.09	16.76	3.78	3.09	3.17	41.32	41.47	38.34
M4	3 (raw)	636.49	28.30	17.96	10.42	3.54	3.00	16.39	13.74	14.64
M5	3 (raw)	250.34	15.68	14.17	5.16	2.75	2.90	10.92	15.35	15.32
M6	3 (raw)	1058.72	266.44	148.99	18.56	9.45	7.71	12.56	15.04	11.49
M7	3 (raw)	23.91	16.54	16.25	3.85	2.93	3.14	38.90	49.07	42.35
M8	3 (raw)	26.41	17.36	17.03	4.06	3.06	3.25	26.21	45.20	37.05
M9	5 (features)	5.73	8.01	6.10	1.85	2.05	1.91	89.20	79.78	83.01
M10	4 (features)	5.44	7.07	5.53	1.79	1.96	1.77	89.78	82.32	84.67
M11	4 (features)	9.74	11.95	8.44	2.39	2.47	2.18	80.72	67.60	75.24
M12	3 (features)	5.88	7.30	5.65	1.86	2.01	1.81	88.88	81.67	84.47
M13	3 (features)	8.65	8.47	6.72	2.28	2.09	1.96	85.08	78.56	81.29
M14	3 (features)	7.51	9.08	6.88	2.12	2.18	1.98	85.71	76.60	80.64
M15	3 (features)	10.07	10.92	8.72	2.42	2.44	2.24	79.98	71.30	74.62
M16	3 (features)	14.81	13.20	11.20	2.99	2.66	2.50	68.59	64.38	65.18

Table 8.4 *The performance of the ANFIS models to predict effluent SS*

Model Number	Number of input parameters (N)	MSE $(mg/l)^2$			AAE (mg/l)			Correlation coefficient (%)		
		Training	Validation	Testing	Training	Validation	Testing	Training	Validation	Testing
M1	5 (raw)	185.45	133.16	193.25	10.37	8.78	10.60	63.38	55.73	53.57
M2	4 (raw)	182.33	121.89	174.17	10.02	8.01	9.91	64.12	60.85	59.71
M3	4(Raw)	255.18	164.01	215.00	12.50	9.78	11.15	42.01	39.90	45.53
M4	3 (raw)	1663.7	1344.3	682.66	24.90	23.95	18.29	21.57	19.71	27.73
M5	3 (raw)	575.75	172.97	201.98	15.66	9.72	10.63	33.47	24.86	18.21
M6	3 (raw)	1021.76	529.56	351.76	20.69	14.54	13.52	21.95	23.76	19.33
M7	3 (raw)	264.41	148.26	221.59	12.45	9.31	11.58	39.16	48.31	42.86
M8	3 (raw)	286.52	159.13	229.62	13.17	9.81	11.79	28.29	42.18	39.52
M9	5 (features)	77.40	65.10	67.37	6.66	6.17	6.60	86.63	81.50	87.02
M10	4 (features)	70.44	56.33	60.52	6.22	5.65	6.54	87.94	84.31	87.34
M11	4 (features)	117.02	91.78	104.78	8.31	7.24	7.63	78.98	72.54	79.26
M12	3 (features)	81.99	59.65	60.73	6.77	5.88	6.64	86.23	83.36	87.22
M13	3 (features)	153.46	99.61	157.50	9.25	7.32	7.67	78.05	72.66	69.66
M14	3 (features)	87.81	67.07	71.06	6.97	6.31	6.04	84.72	81.05	86.12
M15	3 (features)	119.37	82.50	96.45	8.34	6.85	7.29	78.42	76.08	80.31
M16	3 (features)	172.50	117.55	146.62	10.11	7.91	8.68	66.70	63.26	68.21

Table 85 *Statistics summary of the ANFIS models to predict effluent BOD_5*

Model Number	Number of input parameters (N)	Minimum			Maximum			mean		
		Training	Validation	Testing	Training	Validation	Testing	Training	Validation	Testing
Observed		2.00	2.00	2.00	25.00	25.00	23.00	9.32	7.38	6.44
M1	5 (raw)	0.35	2.24	0.15	25.89	14.09	14.53	9.07	7.32	6.97
M2	4 (raw)	0.14	0.76	-0.99	33.23	18.82	14.63	9.08	7.41	6.80
M3	4(Raw)	4.01	1.27	2.27	16.99	13.55	13.62	8.97	7.47	6.95
M4	3 (raw)	-98.85	-13.79	-4.31	240.35	38.44	37.53	10.37	7.58	7.09
M5	3 (raw)	-89.24	1.71	0.99	288.96	19.82	19.02	9.58	7.59	6.91
M6	3 (raw)	-93.35	-26.35	-34.69	215.89	116.20	108.68	14.81	10.78	7.90
M7	3 (raw)	2.76	-1.51	2.07	18.73	14.30	14.12	8.96	7.47	7.08
M8	3 (raw)	2.71	2.55	2.84	14.47	13.49	12.05	8.90	7.45	7.06
M9	5 (features)	3.22	3.28	2.34	21.14	19.79	19.89	9.20	7.27	6.75
M10	4 (features)	2.97	3.32	2.60	21.33	20.20	20.27	9.24	7.34	6.70
M11	4 (features)	3.05	2.64	0.13	21.62	18.50	18.59	9.30	7.15	6.67
M12	3 (features)	2.99	3.57	2.60	21.23	20.44	20.49	9.34	7.42	6.78
M13	3 (features)	2.40	2.45	2.19	25.34	21.43	20.95	9.68	7.21	6.64
M14	3 (features)	2.31	2.39	2.59	21.29	19.30	19.36	9.02	7.22	6.80
M15	3 (features)	1.84	1.89	1.92	21.37	18.17	18.19	9.25	7.13	6.80
M16	3 (features)	3.24	3.46	3.34	21.31	15.88	15.88	9.30	6.98	6.77

Table8.6 Statistic *summary of the ANFIS models to predict effluent SS*

Model Number	Number of input parameters (N)	Minimum			Maximum			mean		
		Training	Validation	Testing	Training	Validation	Testing	Training	Validation	Testing
Observed		7.00	5.00	3.00	80.00	66.00	76.00	28.04	22.41	21.84
M1	5 (raw)	-2.13	4.46	1.43	79.99	46.49	50.55	27.55	22.19	22.64
M2	4 (raw)	-0.97	1.07	-5.42	93.19	54.60	63.89	27.98	21.88	22.42
M3	4(Raw)	7.51	2.94	6.02	54.24	43.28	44.41	27.51	22.24	22.71
M4	3 (raw)	-96.43	-92.29	-75.11	270.13	206.26	133.98	29.35	18.91	21.72
M5	3 (raw)	-52.76	-0.84	-13.93	256.78	77.84	75.96	29.68	23.28	22.40
M6	3 (raw)	-90.00	-96.48	-29.70	224.21	172.41	166.60	26.06	21.32	23.23
M7	3 (raw)	1.73	-1.32	-2.05	56.04	50.18	39.86	26.60	22.26	22.87
M8	3 (raw)	5.40	5.44	7.93	42.73	42.17	35.29	27.03	22.78	23.04
M9	5 (features)	8.25	8.06	8.35	65.52	64.07	64.07	28.37	21.97	21.69
M10	4 (features)	7.60	7.99	7.65	65.24	64.94	64.94	28.46	21.85	21.51
M11	4 (features)	7.42	7.42	0.03	63.16	60.80	60.80	28.79	21.87	21.12
M12	3 (features)	5.48	8.80	7.15	68.62	65.38	64.49	29.22	22.23	21.60
M13	3 (features)	3.65	3.65	3.71	81.04	64.32	64.32	31.92	22.08	23.25
M14	3 (features)	8.69	8.83	8.54	65.28	61.40	61.40	28.45	21.50	21.69
M15	3 (features)	1.66	8.13	6.31	65.88	58.19	58.19	28.42	21.35	21.68
M16	3 (features)	8.27	8.27	8.27	64.02	50.51	50.79	28.79	21.09	21.58

To solve the problem, we found that the Kohonen self-organizing map (or Kohonen features map) has the power to extract features from noisy data and at the same time eliminate the effect of missing values that the ANFIS cannot deal with. Therefore, The KSOM was applied to extract the most relevant features of the raw data records. These features more closely represent the natural structures in data and were therefore used to drive the ANFIS as illustrated in Figure 8.1. This constitutes the hybrid KSOM-ANFIS model.

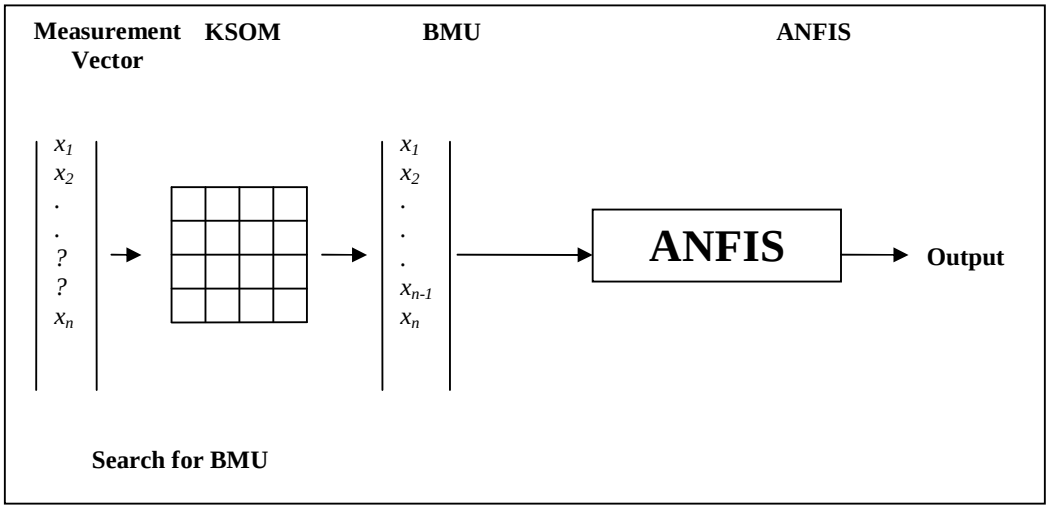


Figure 8.*1 Illustration of the integrated KSOM-ANFIS modelling using BMU features*

The hybrid KSOM-ANFIS modelling strategy was applied to several combination structures of ANFIS, models (9-16) in Tables 8.1 and 8.2. The performance of these models was evaluated using the usual criteria and the results are presented in as can be seen in Tables 8.3 and 8.4. It can be seen that for the same structure, the performance of the model has improved using the features of the measurement vectors than the measurement vectors themselves. For example, model number 2 and model number 10 have the same structure, 4 inputs and three membership functions associated with each input. However, the performance of model number

10 was much better than model number 2, with the correlation coefficient in the testing data set jumping from 53.05% in model number 2 to 84.67% in model number 10 in case of BOD_5. The same conclusion can be inferred for the rest of the models in the three evaluation criteria as can be seen in Tables 8.3 and 8.4. Furthermore, the new modelling strategy generates no negative values in all the combination structures as can be seen in Tables 8.5 and 8.6.

The architecture that performs best was chosen as the final model for predicting the effluent BOD_5 and SS concentrations. Model number 10 generates the best modelling performance in the two cases. Therefore, further discussions were only done on this model. This model has 4 inputs variables: BOD Load, RAS-MLSS, F/M, and temperature with 3 membership functions associated with each input as illustrated in Figure 8.2. Figure 8.3 show the Gaussian membership functions on the operating range. Tables 8.7 shows the parameters of the Gaussian membership functions associated with the input variables, where c is the centre of the corresponding membership function and b is the width. Both models (BOD_5 and SS) contain 81 rules and the total number of modified parameters is 429, composing 24 premise parameters and 405 consequent parameters.

Tables 8.8 and 8.9 present the optimized fuzzy rules generated using the modelling strategy developed in this study for model number 10. The fuzzy neural network model consists of a selection of the 81rules describing the relationship between the input variables and the output variable. The number of rules was 81 and the aggregation process is illustrated in Figure 8.4. Each rule listed in the table consists of an IF and THEN part. The IF part specifies a set of conditions and the THEN part specifies the conclusion or the action. For example, rule 10 in Table 8.8 can be read as:

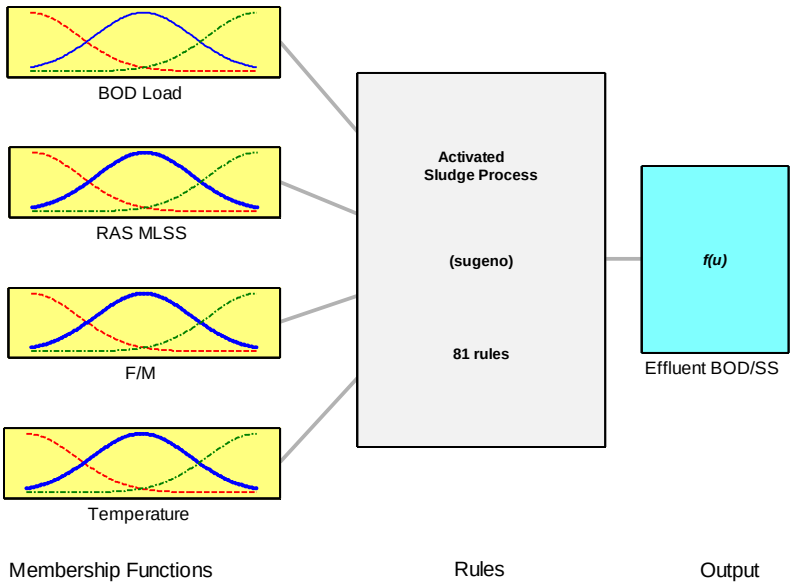


Figure 8.2 *The structure of model number 10*

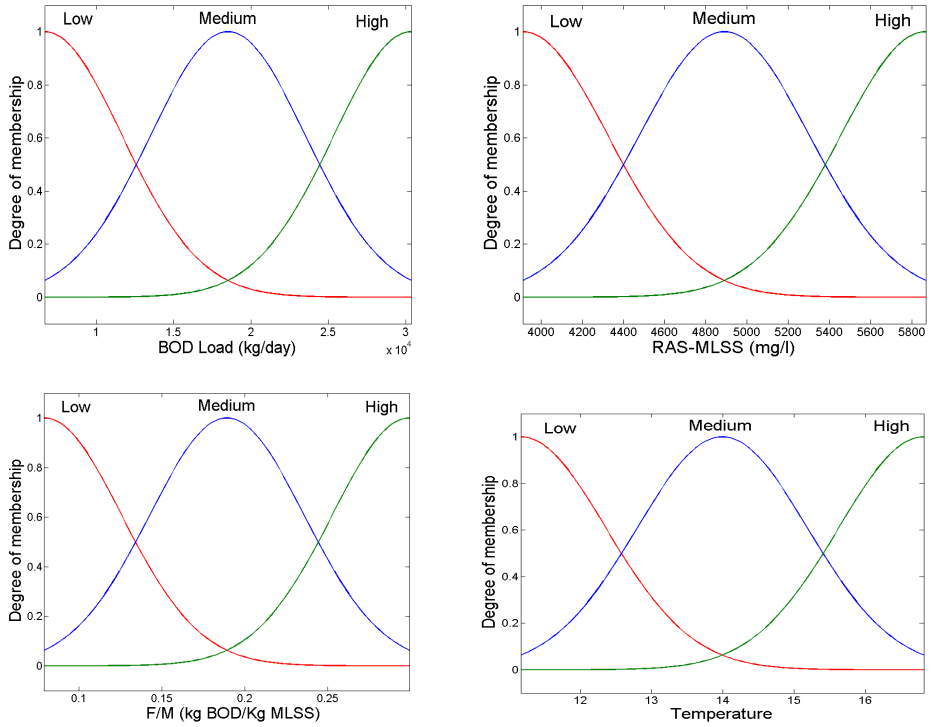


Figure 8.3 *Fuzzy membership functions in the input space.*

Table8.7 *The parameters of Gaussian membership functions associated with the input variables*

Inputs (BOD Model)	Membership Function	b (Width)	C (Centre)
BOD Load	Low	5030	6670
	Medium	5030	18510
	High	5030	30360
RAS-MLSS	Low	414.9	3912
	Medium	414.9	4889
	High	414.9	5866
F/M	Low	0.04678	0.07901
	Medium	0.04678	0.1892
	High	0.04678	0.2993
Temperature	Low	1.203	11.16
	Medium	1.203	13.99
	High	1.203	16.82

IF (BOD Load) is Low and (RAS-MLSS) is Medium and (F/M) is Low and Tem. Is Low, **THEN** (effluent BOD) is -85.1- (0.05327*BOD Load) + (0.08111* RAS-MLSS) - (5.697*F/M) – (18.55*Tem.).

The trained model was tested using data that had not been employed for training. The time series plots of the observed and predicted effluent BOD_5 and Effluent SS during training, validation, and testing data sets are shown in Figure 8.5 and 8.6 respectively. The residuals analysis of the model during training, validation and testing are illustrated in Figure (8.7 and 8.8). The scatter plot of modelled versus observed during training, validation and testing are shown in Figure (8.9 and 8.10).

In the conventional fuzzy inference system, the number of rules is decided by an expert who is familiar with the system to be modelled. The expert uses heuristic knowledge gathered over years of experience to generate these rules. In our case, no expert is needed and the number of membership functions assigned to each input variable is chosen empirically by trial and error. The proposed model was

able to determine the state of the process using the values of several variables together. Furthermore, the proposed model is much more robust with respect to degradation of any associated sensor or measurement equipment. In addition, fuzzy logic model is much easier to design than deterministic model.

These aspects of the proposed modelling strategy are due to the merging of several techniques. Each of these techniques, NN, KSOM, and FLS appears to be extremely effective at handling dynamic, non-linear and noisy data. However, when utilized together, the strengths of each technique can be exploited in a good manner for the development of hybrid systems.

However, different from BP-ANN models, the developed models do not only just give modelling accuracy, but also extract knowledge from the data, for example the adjustment of membership functions and fuzzy rules. All the knowledge is presented in human understandable forms. This is important in order to understand the data and explain how the results were obtained. The extracted knowledge also let an operator know what an expert would say about the state of the process through the knowledge extracted from the linguistic rules. Each rule is a linguistic expression of human expert knowledge establishes relationships between variables, which lead to a diagnosis output.

8.4 Conclusion

In this chapter, the use of ANFIS for modelling wastewater treatment plants was demonstrated. The ANFIS allows fuzzy rules to be extracted and the ANN enabled optimised fuzzy membership functions to be determined. The methodology was applied to activated sludge data for Seafield wastewater treatment plant in order to predict the effluent BOD_5 and SS. Initially, a set of measured, raw, data were used to train and test the ANFIS, but the resulting model did not perform well. This was attributed to the existing of noise in the raw

data. To overcome this, features of the data or their BMU were extracted with the KSOM and used for training and testing a new set of models that gave better performance. The results indicate that the KSOM-ANFIS not only outperforms the basic ANFIS model in modelling capability with different number of inputs and different number of fuzzy membership functions, it is unhindered by missing values or gaps in the data.

However, although the extracted features of KSOM enhanced the performance of ANFIS, the input data cannot propagate directly from the input of the KSOM to an output of the ANFIS as the task were performed separately in this work. Therefore, developing a single model that incorporates the features and the ANN or ANFIS is one of the major open research issues. Moreover, even if satisfactory results have been obtained based on integrated KSOM-ANFIS to deal with noise and missing values, the robustness of the developed models has not yet been analysed mathematically, for example, the effect of the number of missing values in each vector. Such an analysis together with further experimentation would be worth pursuing to prove conclusively the general validity of the proposed schemes.

Table 8.8 *Optimised fuzzy rules generated using the modelling strategy developed in this study for model number 10 (BOD Model)*

Rule Number	Rule Antecedent (If)				Then	Consequent Parameters				
	BOD-Load	RAS-MLSS	F/M	Temp.		a_1	a_2	a_3	a_4	a_0
1	Low	Low	Low	Low		-1.464	0.6413	2.152	507.8	11.91
2	Low	Low	Low	Medium		0.2801	0.9969	23.75	-356.1	34.67
3	Low	Low	Low	High		0.02804	0.3626	-76.3	-80.99	-102.9
4	Low	Low	Medium	Low		27.95	-61.98	-0.1368	-19.02	-1.702
5	Low	Low	Medium	Medium		-1.463	-2.554	-11.29	1329	32.73
6	Low	Low	Medium	High		0.3338	-0.9537	-3.959	-0.8736	-104.8
7	Low	Low	High	Low		-434.7	-114.4	-0.006909	-0.4417	-0.01234
8	Low	Low	High	Medium		53.01	-116.3	-0.1005	-5.05	-1.549
9	Low	Low	High	High		-41.12	100.1	0.007273	-26.11	-0.7746
10	Low	Medium	Low	Low		-0.05327	0.08111	-5.697	-18.55	-85.1
11	Low	Medium	Low	Medium		-0.26	-0.481	66.87	235.6	168.5
12	Low	Medium	Low	High		0.3912	-0.594	-2.843	-98.61	6.734
13	Low	Medium	Medium	Low		2.437	-4.422	-3.428	491.2	2.402
14	Low	Medium	Medium	Medium		0.03674	-1.874	-12.9	628.4	32.39
15	Low	Medium	Medium	High		-0.8339	0.3398	-0.7227	483.2	17.3
16	Low	Medium	High	Low		64.32	-303.8	0.04543	0.9669	-0.04233
17	Low	Medium	High	Medium		-80.16	230.3	0.3187	45.79	0.6505
18	Low	Medium	High	High		9.89	-48.85	0.1341	-21.57	0.05542
19	Low	High	Low	Low		0.329	-0.1468	3.803	-182.3	-46.6
20	Low	High	Low	Medium		-2.684	5.404	4.49	-619.4	-18.23
21	Low	High	Low	High		-103.5	213.8	0.1728	6.23	0.1674
22	Low	High	Medium	Low		-0.6909	1.5	-3.579	137	7.24
23	Low	High	Medium	Medium		4.641	-10.72	-3.688	-198.4	-2.424
24	Low	High	Medium	High		101.6	-201.7	-0.1835	-19.62	-0.2552
25	Low	High	High	Low		155.5	-299.7	0.0369	-0.7076	-0.05671
26	Low	High	High	Medium		30.35	42.34	0.05174	-1.566	-0.05357
27	Low	High	High	High		-102.8	-125.2	-0.001678	-0.6948	-0.02658
28	Medium	Low	Low	Low		-26.1	68.72	0.07732	27.95	-0.9504

Table 8.8 *Continue*

Rule Number	Rule Antecedent (If)				Then	Consequent Parameters				
	BOD-Load	RAS-MLSS	F/M	Temp.		a_1	a_2	a_3	a_4	a_0
29	Medium	Low	Low	Medium		4.11	-10.64	1.76	204.5	14.46
30	Medium	Low	Low	High		-1.137	1.08	-6.471	281.7	2.76
31	Medium	Low	Medium	Low		22.74	-64.54	0.2399	96.05	3.981
32	Medium	Low	Medium	Medium		-2.099	6.206	-0.7487	360.9	23.64
33	Medium	Low	Medium	High		0.05308	-0.3521	-2.706	166	3.143
34	Medium	Low	High	Low		74.22	-328.3	0.03933	-0.5536	-0.05439
35	Medium	Low	High	Medium		8.44	-43.11	0.2981	41.34	-1.15
36	Medium	Low	High	High		-11.41	47.81	0.2932	-110.2	0.3064
37	Medium	Medium	Low	Low		-1.685	2.148	7.804	419.1	-3.966
38	Medium	Medium	Low	Medium		-0.946	4.292	26.61	-303.5	56.58
39	Medium	Medium	Low	High		1.011	-4.937	2.772	837.3	1.516
40	Medium	Medium	Medium	Low		1.807	-6.858	-8.68	437.6	-11.29
41	Medium	Medium	Medium	Medium		0.8092	-2.723	-29.89	152.4	-48.23
42	Medium	Medium	Medium	High		0.1212	-1.806	-2.909	368.6	3.74
43	Medium	Medium	High	Low		3.757	-2.811	0.17	12.25	0.598
44	Medium	Medium	High	Medium		-3.106	12.89	1.026	65.06	-0.008005
45	Medium	Medium	High	High		-6.956	35.79	0.232	-56.41	0.3373
46	Medium	High	Low	Low		0.4053	-1.34	9.16	20.45	10.54
47	Medium	High	Low	Medium		-3.665	10.76	7.984	-96.8	3.429
48	Medium	High	Low	High		35.4	-40.18	0.1678	3.029	0.03406
49	Medium	High	Medium	Low		-0.4842	1.674	-6.858	-200.5	13.02
50	Medium	High	Medium	Medium		0.9006	-2.767	-9.852	82.36	8.144
51	Medium	High	Medium	High		32.97	-143.9	-0.121	-19.74	0.1601
52	Medium	High	High	Low		17.38	-64.66	0.06569	-11.93	-0.2243
53	Medium	High	High	Medium		30.73	-124.5	0.268	6.056	-0.069
54	Medium	High	High	High		-46.98	413.6	0.01382	0.1548	0.06961
55	High	Low	Low	Low		-509.9	-193	-0.004156	-0.1879	-0.03211
56	High	Low	Low	Medium		250	-1244	0.01438	-7.833	-0.1029

Table 8.8 *Continue*

Rule Number	Rule Antecedent (If)				Then	Consequent Parameters				
	BOD-Load	RAS-MLSS	F/M	Temp.		a_1	a_2	a_3	a_4	a_0
57	High	Low	Low	High		86.59	-177.2	0.00813	8.405	0.0405
58	High	Low	Medium	Low		49.84	-454.1	-0.02964	-2.152	-0.1469
59	High	Low	Medium	Medium		-19.62	109.6	-0.1506	10.73	0.07485
60	High	Low	Medium	High		33.69	-155.3	-0.05348	-20.51	-0.2145
61	High	Low	High	Low		-20.91	83.51	-0.1133	5.447	0.0592
62	High	Low	High	Medium		3.586	-20.62	-0.4702	-19.08	-0.2712
63	High	Low	High	High		-84.27	501.6	-0.02081	0.622	0.03551
64	High	Medium	Low	Low		-189.2	699.2	0.09335	2.254	0.0933
65	High	Medium	Low	Medium		23.39	-85.35	0.2616	-3.211	-0.1131
66	High	Medium	Low	High		35.57	-211.9	0.01069	4.471	-0.121
67	High	Medium	Medium	Low		23.5	-88.46	0.005508	3.541	0.9487
68	High	Medium	Medium	Medium		8.804	-44.27	-0.7718	47.75	0.3608
69	High	Medium	Medium	High		1.795	-12.54	-0.1307	-9.892	0.2305
70	High	Medium	High	Low		41.96	-247.2	-0.2783	1.822	-0.3011
71	High	Medium	High	Medium		-1.042	5.463	-0.6585	-23.39	0.1907
72	High	Medium	High	High		20.66	-127.2	0.06481	6.189	0.2732
73	High	High	Low	Low		17.77	-48.53	0.07674	-4.994	0.05322
74	High	High	Low	Medium		35.01	-99.64	0.1146	2.731	0.07672
75	High	High	Low	High		216.2	30.32	0.002732	0.07731	0.007275
76	High	High	Medium	Low		-3.71	13.47	-0.03734	15.14	1.342
77	High	High	Medium	Medium		-8.635	28.76	-0.1598	12.29	-0.07895
78	High	High	Medium	High		-55.91	180.4	-0.01187	-0.6627	0.01127
79	High	High	High	Low		-2.27	14.08	0.0991	31.68	0.8121
80	High	High	High	Medium		-16.01	91.43	0.02866	4.486	0.1876
81	High	High	High	High		6.632	-13.57	-0.00177	-0.7383	-0.02826

Table 8.9 *The optimised fuzzy rules generated using the modelling strategy developed in this study for model number 10 (SS Model)*

Rule Number	Rule Antecedent (If)				Then	Consequent Parameters				
	BOD-Load	RAS-MLSS	F/M	Temp.		a_1	a_2	a_3	a_4	a_0
1	Low	Low	Low	Low		8.353	-18.46	5.569	1877	22.53
2	Low	Low	Low	Medium		0.9542	2.475	69.99	-993	119.2
3	Low	Low	Low	High		0.2811	1.08	-153.7	-279.6	-429.7
4	Low	Low	Medium	Low		75.82	-191	-0.2211	25.81	-5.111
5	Low	Low	Medium	Medium		-3.02	-11.33	-22.9	4271	120.8
6	Low	Low	Medium	High		0.8393	-5.881	-15.33	678.6	-320.1
7	Low	Low	High	Low		-509.7	-121.8	-0.01206	0.0611	0.04464
8	Low	Low	High	Medium		142.8	-354.6	-0.36	-22.47	-5.323
9	Low	Low	High	High		-88.69	264.3	0.2473	-50.52	-1.166
10	Low	Medium	Low	Low		-0.05756	0.3658	-9.819	-168.9	-235
11	Low	Medium	Low	Medium		-1.523	-1.28	169.9	957.3	722.6
12	Low	Medium	Low	High		1.339	-4.206	11.65	259.7	0.9296
13	Low	Medium	Medium	Low		1.984	-1.547	-8.724	1128	-2.023
14	Low	Medium	Medium	Medium		-0.06445	-7.299	-18.16	2640	128.4
15	Low	Medium	Medium	High		-3.665	5.211	-2.172	1370	71.47
16	Low	Medium	High	Low		190.3	-576.7	0.1592	2.334	-0.1674
17	Low	Medium	High	Medium		-153.8	428.8	1.051	182.3	2.709
18	Low	Medium	High	High		187.3	-505.3	0.3548	-85.27	-0.1046
19	Low	High	Low	Low		1.406	0.2621	5.936	-1219	-116.9
20	Low	High	Low	Medium		-9.543	17.55	7.509	-1372	-38.58
21	Low	High	Low	High		-319.6	625.7	0.3734	9.19	0.3032
22	Low	High	Medium	Low		-6.587	13.93	-8.545	458.4	23.72
23	Low	High	Medium	Medium		19.84	-45.26	-9.712	-548.3	-8.606
24	Low	High	Medium	High		243.9	-473.9	-0.4158	-75.74	-0.9188
25	Low	High	High	Low		602.2	-1326	0.1159	-3.235	-0.2693
26	Low	High	High	Medium		175.3	-310.4	0.1873	-6.43	-0.2758

Table 8.9 *Continue*

Rule Number	Rule Antecedent (If)				Then	Consequent Parameters				
	BOD-Load	RAS-MLSS	F/M	Temp.		a_1	a_2	a_3	a_4	a_0
27	Low	High	High	High		102.2	-303.3	-0.002048	-2.241	-0.06128
28	Medium	Low	Low	Low		-91.39	235	0.004699	103.2	-3.811
29	Medium	Low	Low	Medium		10.71	-29.9	2.278	956.1	45.44
30	Medium	Low	Low	High		-1.933	0.7618	-11.44	325.1	-41.33
31	Medium	Low	Medium	Low		34.4	-82.51	0.5056	315.7	12.15
32	Medium	Low	Medium	Medium		-5.334	18.68	-3.503	7.981	44.39
33	Medium	Low	Medium	High		-0.5462	3.719	-10.39	-39.71	7.438
34	Medium	Low	High	Low		340.1	-1540	0.146	1.28	-0.1943
35	Medium	Low	High	Medium		22.96	-110.7	0.9042	101.2	-5.582
36	Medium	Low	High	High		39.83	-147.6	0.9208	-413.9	1.215
37	Medium	Medium	Low	Low		-2.936	2.697	19.19	770.3	-19.71
38	Medium	Medium	Low	Medium		-4.99	18.29	52.65	-535.7	158.7
39	Medium	Medium	Low	High		3.035	-13.1	6.011	2001	11.73
40	Medium	Medium	Medium	Low		3.823	-14.36	-21.14	1009	-18.97
41	Medium	Medium	Medium	Medium		2.195	-7.084	-64.76	361.4	-177.2
42	Medium	Medium	Medium	High		0.5308	-5.216	-0.8529	580.7	4.666
43	Medium	Medium	High	Low		48.9	-182.8	0.6469	41.19	1.957
44	Medium	Medium	High	Medium		-3.638	19.11	3.225	253.1	-0.2107
45	Medium	Medium	High	High		-14.67	31.42	0.4245	-236.1	0.6424
46	Medium	High	Low	Low		3.743	-10.62	21.51	495.4	40.64
47	Medium	High	Low	Medium		-17.15	47.46	21.67	-359.4	7.952
48	Medium	High	Low	High		255.8	-479.7	0.3591	1.92	0.1752
49	Medium	High	Medium	Low		-2.645	8.256	-13.92	-400.8	14.08
50	Medium	High	Medium	Medium		2.932	-9.138	-26.59	213.9	12.02
51	Medium	High	Medium	High		91.24	-400.7	-0.0186	-79.48	1.231
52	Medium	High	High	Low		27.51	-119.4	-0.08227	-50.49	-0.981
53	Medium	High	High	Medium		85.98	-319.8	0.9399	41.63	0.1614
54	Medium	High	High	High		-334.6	1565	0.04873	-0.5903	0.2768

Table 8.9 *Continue*

Rule Number	Rule Antecedent (If)				Then	Consequent Parameters				
	BOD-Load	RAS-MLSS	F/M	Temp.		a_1	a_2	a_3	a_4	a_0
55	High	Low	Low	Low		-529.2	-249.5	-0.003354	0.3187	-0.02806
56	High	Low	Low	Medium		620.3	-3042	0.03459	-21.08	-0.2247
57	High	Low	Low	High		133.3	-125.6	0.02509	21.75	0.1826
58	High	Low	Medium	Low		288.6	-1946	-0.1056	-4.567	-0.5193
59	High	Low	Medium	Medium		-60.23	296.8	-0.3838	4.577	-0.2681
60	High	Low	Medium	High		38.45	-218.4	-0.1018	-60.39	-0.5869
61	High	Low	High	Low		114.5	-919.4	-0.3956	18.62	0.001931
62	High	Low	High	Medium		-10.09	66.95	-1.572	-73.31	-0.7882
63	High	Low	High	High		-277	1730	-0.09642	1.498	0.06121
64	High	Medium	Low	Low		-438.1	1618	0.2217	5.585	0.2336
65	High	Medium	Low	Medium		47.71	-169.1	0.5543	-16.35	-0.6082
66	High	Medium	Low	High		-29.82	-319	-0.005186	10.97	-0.3395
67	High	Medium	Medium	Low		41.65	-154.6	-0.1154	2.106	2.108
68	High	Medium	Medium	Medium		30.52	-147.3	-2.364	52.08	-1.395
69	High	Medium	Medium	High		-41.42	208.1	-0.1288	-21.43	0.9559
70	High	Medium	High	Low		84.29	-489.9	-1.092	2.629	-1.629
71	High	Medium	High	Medium		11.55	-77.51	-2.435	-125.7	-0.7623
72	High	Medium	High	High		8.38	-50.28	0.3089	31.38	1.376
73	High	High	Low	Low		133.5	-435.6	0.1543	-5.253	0.08619
74	High	High	Low	Medium		116.3	-205.7	0.3183	3.329	0.1459
75	High	High	Low	High		-155.7	-125.4	0.0004806	-0.3384	-0.01498
76	High	High	Medium	Low		-12.92	52.66	0.2382	125.8	5.846
77	High	High	Medium	Medium		-16.52	41.95	-0.3145	78.84	1.154
78	High	High	Medium	High		-99.72	208.3	-0.03955	-5.604	-0.09403
79	High	High	High	Low		-8.541	51.79	0.4702	228.6	4.584
80	High	High	High	Medium		-44.53	254.6	0.09065	23.1	1.051
81	High	High	High	High		18.32	176.6	-0.01171	-3.323	-0.09995

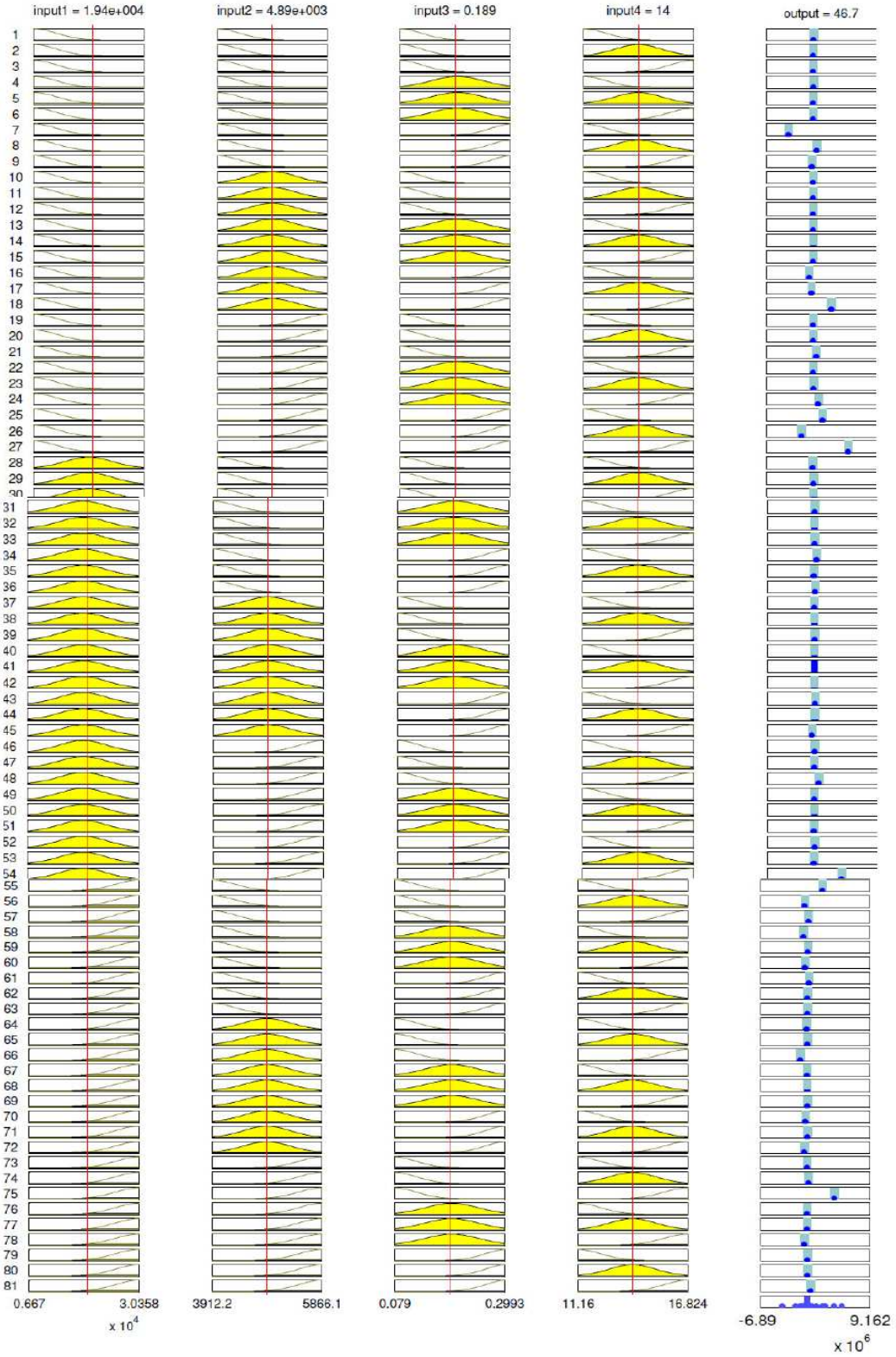


Figure 8.4 *Fuzzy inference diagram for model number 10 predicting effluent BOD_5. the user just need to put the input values to get the output value as seen from the figure.*

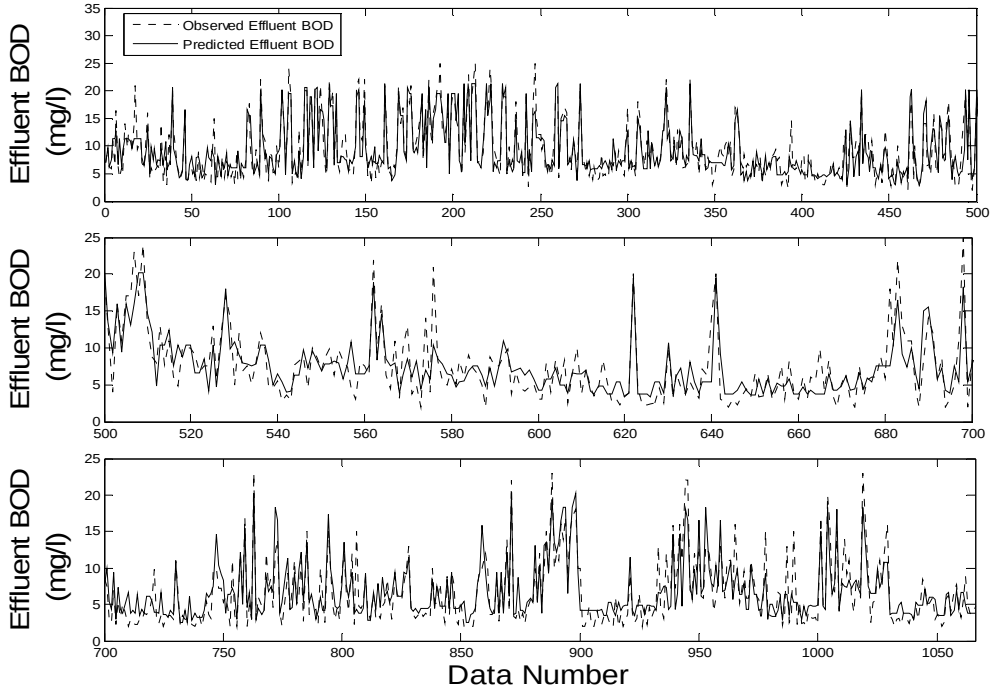


Figure 8.5 *The time series plots of the observed and predicted BOD during training, validation and testing data sets for model number 10.*

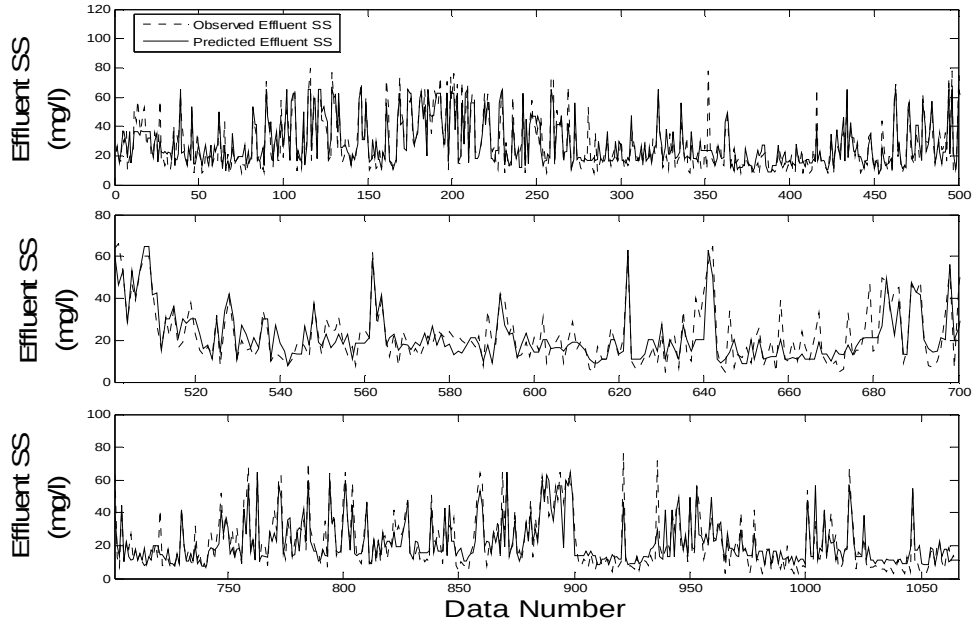


Figure 8.6 *The time series plots of the observed and predicted SS during training, validation and testing data sets for model number 10*

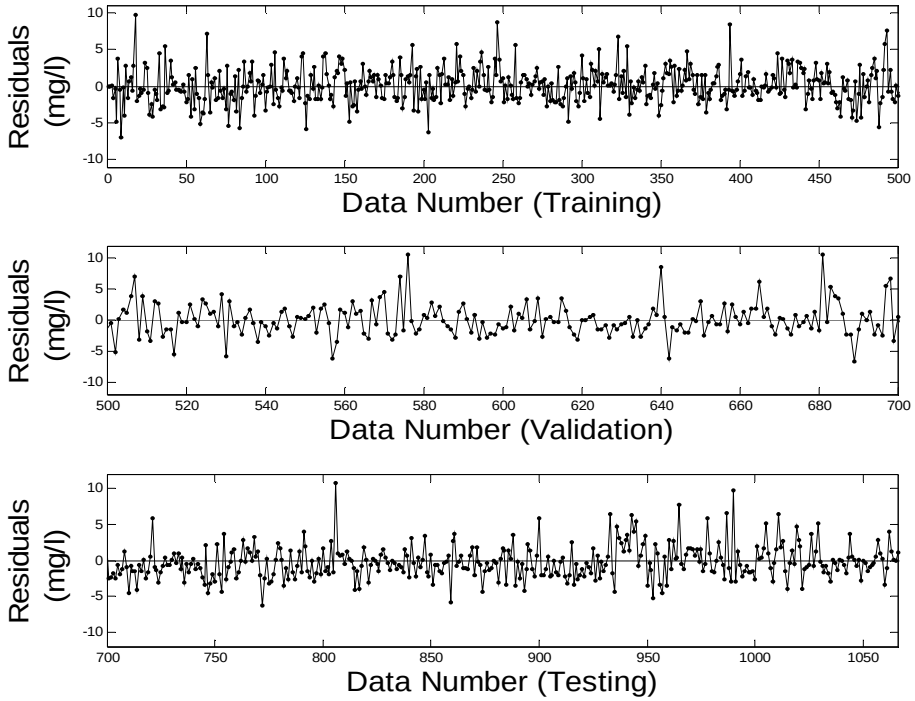


Figure 8.7 *Residuals analysis of the BOD model during training, validation and testing for model number 10*

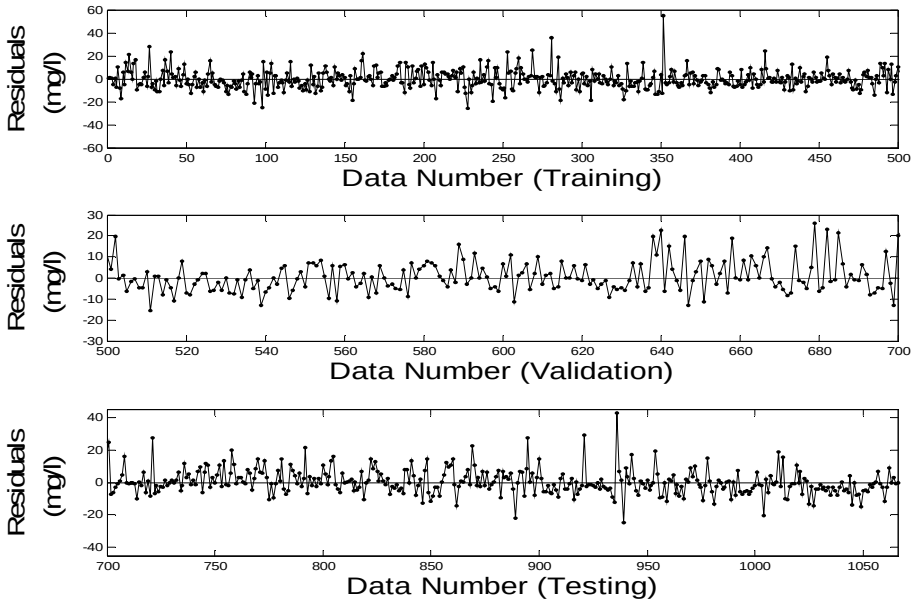


Figure 8.8 *Residuals analysis of the SS model during training, validation and testing for model number 10*

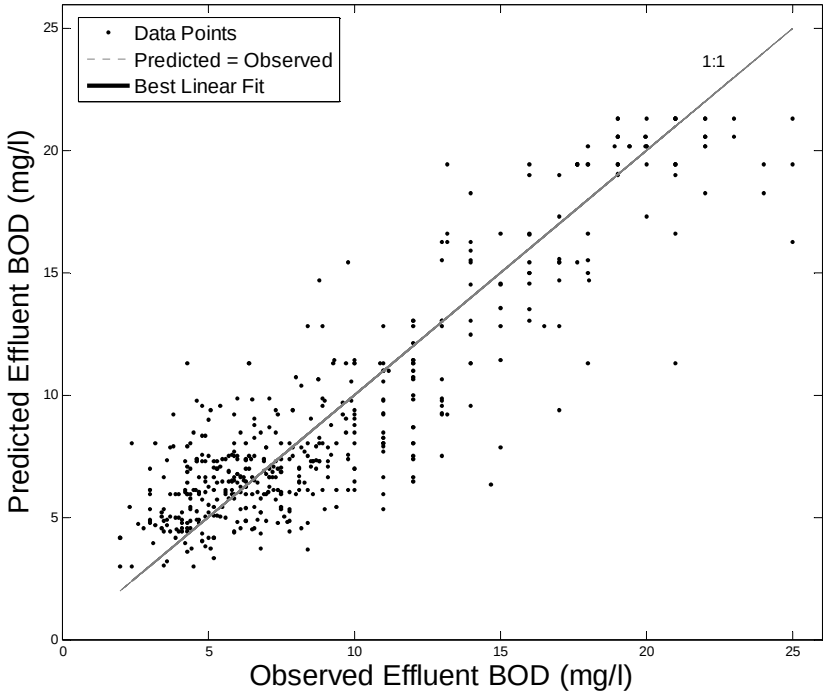


Figure 8.9a *The scatter plot of modelled versus observed during training for model number 10*

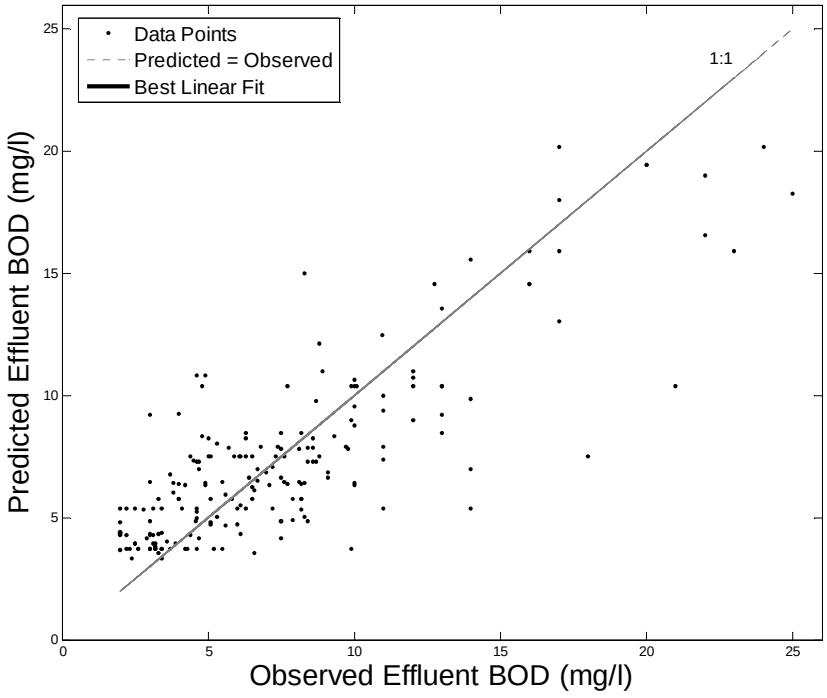


Figure 8.9b *The scatter plot of modelled versus observed during for model number 10*

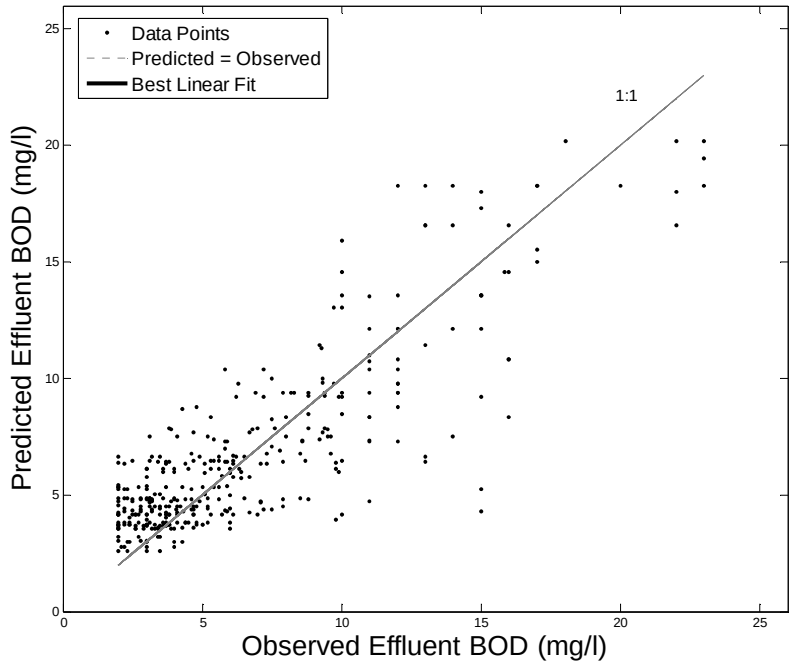


Figure 8.9c *The scatter plot of modelled versus observed during testing for model number 10*

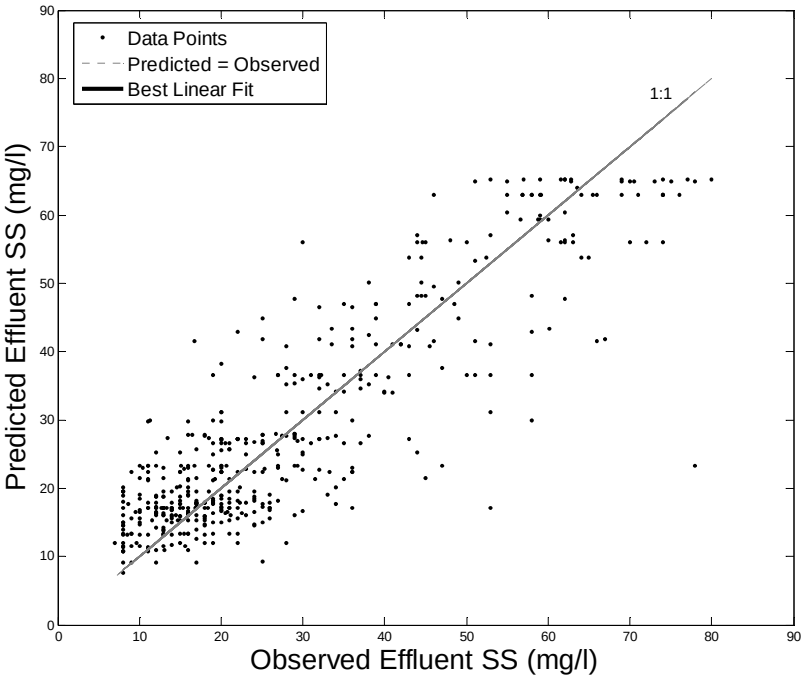


Figure 8.10a *The scatter plot of modelled versus observed during training for model number 10*

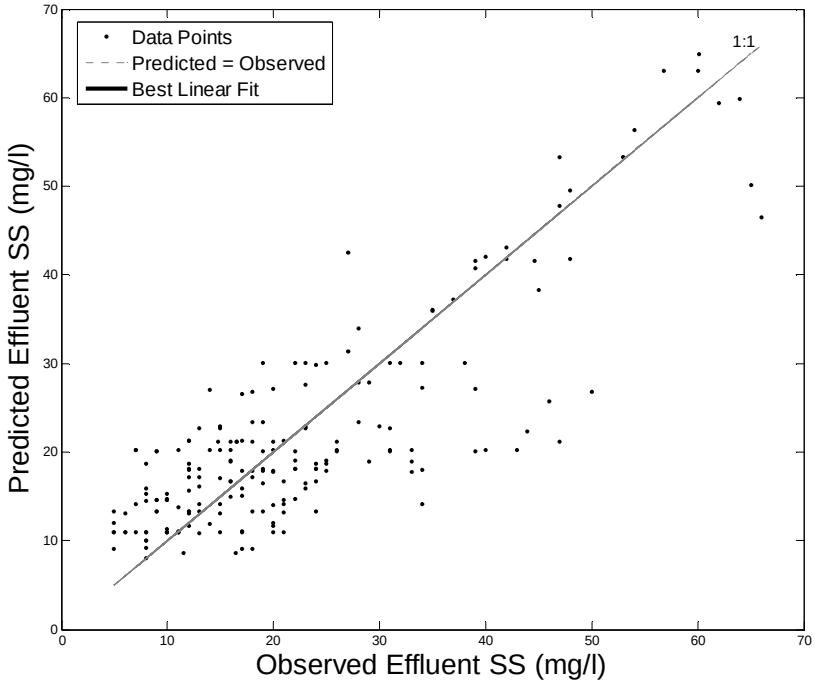


Figure 8.10b *The scatter plot of modelled versus observed during validation for model number 10*

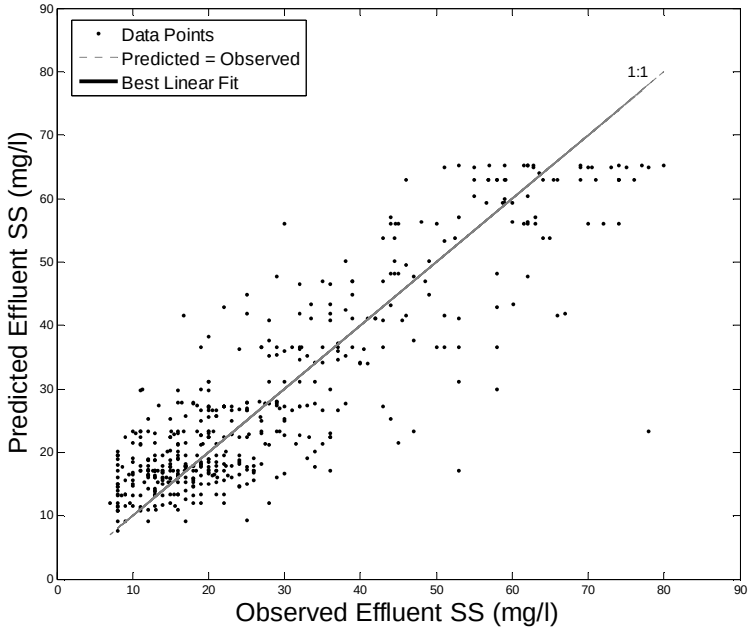


Figure 8.10c *The scatter plot of modelled versus observed during testing for model number 10.*

CHAPTER 9

DISCUSSION AND CONCLUSIONS

9.1 Discussion

Modelling the activated sludge process can improve the performance of wastewater treatment plant and can lead to a better understanding of the system. However, the complexity and uncertainty in the process make the task somewhat complicated using traditional deterministic models. Their exists another set of modelling techniques, known as artificial intelligence or data driven techniques, which require no prior knowledge of the structure or state of the system. However, the quality of these techniques depends strongly on the quality of the data.

The advantages of AI models are that these models are able to predict the effluent concentrations without the previous knowledge of the system. In addition, assumptions about the mathematical relationships between inputs and outputs are not needed. Furthermore, these models are able to recognize the relationships between the inputs and outputs without explicitly considering the physics of process.

The main aim of this research work was to test the hypothesis that AI techniques can be used for modelling the activated sludge wastewater treatment plants. Consequently, the objectives of this study were to investigate the efficiency of KSOM in improving the data driven techniques developed in this study. Other objectives were also inherent in this project, namely preparation of the data to eliminate the effect of noise and missing values and developing a software sensor to predict the biological oxygen demand.

The study has provided a systematic and thorough approach to achieve the highlighted objectives. A variety of intelligent models were developed and applied, and significant differences between KSOM, ANN, and ANFIS were highlighted.

9.1.1 Data preparation

Measurement data collected at real wastewater treatment plants are often distorted by noise, outliers, and missing values. This calls for validation and reconstruction of data prior to any thorough analysis. In application I, the activated sludge data obtained from the Seafield activated sludge wastewater treatment plant Edinburgh, UK during a period of about three years have been modelled to replace outliers and missing values using the Kohonen Self Organising Map (KSOM). After the iterative training of the KSOM, each of the 1066 samples was associated with an output unit known as the best map unit (BMU). The outliers or missing values were then replaced with the corresponding component from the BMU. The results demonstrated that the KSOM is an excellent tool for replacing outliers and missing values in high dimensional data sets. The predicted missing values are plausible and show a trend not dissimilar to that of the observed measurements. These results cannot be obtained from traditional time series models due to the multivariate, time varying and highly non-linear nature of the process. The method is simple, computationally efficient and highly accurate.

9.1.2 Features extraction and data visualization

The most important issue in modelling the wastewater treatment plant using AI techniques is the quality of the data. Therefore, even when missing values and outliers have been dealt with as above, the ensuing data record still contains significant noise that must be filtered. Therefore, features of the data need to be extracted. Furthermore, when it is decided to use data set to build the activated

sludge model, the first logical step was to have a general view of the possible major relationships between variables. For these two purposes, multidimensional features extraction tools had to be developed in order to extract the features and present it in an understandable form.

The KSOM was applied for extracting features from the raw data. Then these features are organised in an understandable way using the component planes. These visualizations enable the human eyes to explore these large amounts of data and discover the complex correlation between process variables for diagnosing the potential cause for upset situations in the activated sludge wastewater treatment plant. The component planes of the KSOM reveal the complex relationship between the process variables without any extra information about the mechanism of this complex system. The results demonstrated the efficiency of KSOM as a tool for the discovery of correlations between large data sets, as well as the visualization of such correlations, thus making it easy to immediately identify cause-effect correlations between process variables.

9.1.3 Software sensor for fast predicting of BOD_5

Application II of the study presented a completely novel methodology based on the use of the Kohonen self-organizing map (KSOM) models to predict five-days @ 20°C biochemical oxygen demand (BOD) concentrations in wastewater, using raw sewage data obtained at three wastewater treatment plants in Scotland. Extensive testing and validation of the model shows that the model is sufficiently general to predict the BOD readily using variables which can be measured within three hours or in real-time using on-line hardware sensors, thus making it possible to estimate BOD very rapidly. This allows for a timely intervention and cost reduction during problem diagnosis.

The proposed BOD software sensor methodology is preferred over BOD biosensors because the BOD can be estimated directly and no costly maintenance is required. The software sensor does not require calibration and cannot be negatively affected by toxins and other inhibitors. Moreover, the software sensor is very dynamic and can be readily updated when additional data become available, thus enhancing its accuracy. The KSOM tool used for the development of the software sensor can readily deal with missing values in one or more of the input variables without significantly impacting on the accuracy of the model.

9.1.4 Modelling ASP using AI paradigms

Application III used a new methodology based on a hybrid supervised-unsupervised artificial neural network to improve the performance of the basic backpropagation neural network method in modelling the activated sludge wastewater treatment plant. Input variables were selected based on their correlation with the effluent BOD and SS, which were the target prediction variables. Several ANN models with different numbers of neurons in the hidden layers were developed. For each model, two types of data were used, the first one is the raw data set, and the second one is the extracted features of the raw data using the Kohonen self-organising map. The results showed that the models using the features were better than those using the raw data.

The findings prove the ability of KSOM to improve the performance of modelling using basic back-propagation neural networks, particularly when the available data are noisy, a common problem with the process data of wastewater treatment plants. Furthermore, the KSOM can readily deal with missing values in one or more of the input variables without significantly negative impacts on the accuracy of the model. Results obtained provide that KSOM-ANN present a versatile tool in modelling ASP and provided an alternative methodology for

predicting the performance of WWTPs. The results also stated that the best ANN structure did not necessarily mean the most number of hidden neurons.

The results indicated that the KSOM extracted features improve the performance of backpropagation multilayer perceptron. The learning algorithms were found to be suitable for modelling ASP. However, the concept or knowledge cannot be clearly expressed in a human understandable way in the form of if-then rules. In contrast, Fuzzy logic systems (FLS) are more favourable, in that their behaviours can be explained using fuzzy rules. In addition, FLSs can easily be interpreted in human understandable terms rather than with numerical quantities. In other words, fuzzy logic models depend not only on black box such as ANN, but are also based on a combination of knowledge of the system and operational experience.

Therefore, due to the power of KSOM in enhancing the performance of ANN, the last application of this work investigates the possibility of integrating the KSOM with ANFIS in order to improve the performance of the ANFIS trained with raw data. The ANFIS allows fuzzy rules to be extracted and the ANN enabled optimised fuzzy membership functions to be determined. The results indicate that the KSOM-ANFIS not only outperforms the basic ANFIS model in modelling capability with different number of inputs and different number of fuzzy membership functions, it is unhindered by missing values or gaps in the data.

A comparison between different modelling strategies is illustrated in Table 9.1 and 9.2. It can be seen that the performance of the models is better using the features than using the raw data. Indeed, the performance of the models can increase by more than 30% based on the correlation coefficient criterion. The results also indicated that KSOM-ANFIS was the best strategy in modelling the activated sludge process as it provides a correlation of more than 80%.

Table 9.1 *Comparison of statistical coefficient for effluent BOD for the best model for each category of models*

Modelling Type	MSE $(mg/l)^2$			AAE (mg/l)			Correlation (R) %		
	Training	Validation	Testing	Training	Validation	Testing	Training	Validation	Testing
ANN (Raw)	22.73	17.01	17.64	3.60	2.88	3.09	49.54	51.71	32.02
KSOM-ANN (Features)	10.34	11.54	11.37	2.32	2.41	2.51	81.80	68.31	65.49
ANFIS (Raw)	17.44	13.41	14.00	3.05	2.57	2.85	62.06	61.99	53.05
KSOM-ANFIS (Features)	5.44	7.07	5.53	1.79	1.96	1.77	89.78	82.32	84.67

Table 9.2 *Comparison of statistical coefficient for effluent SS for the best model with each category of models*

Model Number	MSE $(mg/l)^2$			AAE (mg/l)			Correlation (R) %		
	Training	Validation	Testing	Training	Validation	Testing	Training	Validation	Testing
ANN (Raw)	230.48	134.09	222.94	11.31	8.67	11.51	45.68	58.53	42.25
KSOM-ANN (Features)	121.04	106.20	162.73	7.62	7.51	8.49	78.32	67.53	65.04
ANFIS (Raw)	182.33	121.89	174.17	10.02	8.01	9.91	64.12	60.85	59.71
KSOM-ANFIS (Features)	70.44	56.33	60.52	6.22	5.65	6.54	87.94	84.31	87.34

It must be stressed here that modelling the activated sludge wastewater treatment plant as performed in this work is an alternative to the more widely used approach base on deterministic mathematical modelling. However, it is an alternative, and it is difficult to be compared with the traditional methods as it follows different modelling paradigms. While the traditional methods are useful for the design, the new approach is useful for operation. The results obtained from this study provide useful information for a process engineer who is faced

with improving the performance of the WWTP because the KSOM can reveal the complex dependencies among process variables which can be used to solve the operational problems in WWTP. In addition, the AI models developed in this study can aid the operator to predict the performance of the plant as well as fast predict the BOD values which is the key parameter in the operation of the treatment works. This will enable the operator to make some adjustments to prevent the failure of the works in case of overloading of influent concentrations or flow. Therefore, the operator will have enough time to take action for solving the problems such as manipulating RAS, WAS, etc.

Although the success of the developed models so far is evident, they suffer from limitations which will affect their ability to give correct answers about the behaviour of the system under a new set of circumstances. The models may have been capable of predicting the correct output if the inputs are close to the features of training data; otherwise such cannot be guaranteed. In other words, the models may not able to simulate outputs outside the range of those they were trained with, i.e. they could prove to be poor extrapolation. Therefore, care must be taken when extrapolating these results to other nonlinear systems.

Furthermore, there are problems concerned with identifying the AI elements and parameters. Examples of the elements and parameters include the number of hidden layers, the number of neurons in each layer, training function, and transfer function in case of ANN and the number and shape of fuzzy membership function in case of ANFIS. Therefore, finding the best general model is time consuming as it is highly empirical, and the only way to optimize the models is by searching for the best network parameters through trials. In addition, the model parameters need to be re-calibrated from time to time, in order to make sure that the model maintains an adequate description of the process.

9.2 Conclusions

The specific conclusions of this study are:

9.2.1. The use of AI techniques for modelling the complex nature of activated sludge wastewater treatment has been shown to be feasible. Because these tools are essentially data-driven and do not require *a priori* specification of the mathematical form of the processes, they reduce the numerous uncertainties associated with traditional mechanistic modelling especially with regards to the model identification and parameter estimation of such models.

9.2.2. The AI models developed in this work appear to be unaffected by missing values or outliers; consequently they could be applied to any treatment works with minor modification irrespective of the state or completeness of the available data. This is a major advantage since most process data records have some gaps in them.

9.2.3. The work has developed a software sensor for the BOD_5, the single most important water quality parameter used for assessing wastewater bio-treatability, monitoring effluent quality for the purpose of water pollution control and for assessing the overall performance of wastewater treatment plants. Contrary to the usual 5-days delay of traditional BOD_5 bioassay techniques, the software sensor produces almost instantaneous estimate of BOD_5 using simple and readily available water quality parameters. This is a major development as it makes it possible to use the BOD_5 for real time operational control of wastewater treatment. This aspect of the work has been published in the Journal of Water Environment Research.

9.2.4. The use of the KSOM for the analysis and visualisation of large dimensional dataset has resulted in the prediction of missing values in data sets. The clustering capability of the KSOM means that such missing values maintain the multivariate correlation existing in the data and are hence the most accurate estimate possible for the missing data. The scarcity of water quality data and the huge expenses associated with its collection mean that data vectors with missing elements cannot be discarded; robust estimates for such missing values must be found.

9.2.5. Another significant output of the study is the advantage it has taken of the power of the KSOM to extract the essential features in a huge data base, thus producing "noiseless" data that essentially improve modelling performance. The superiority of modelling with the features-extracted data over raw data was clearly demonstrated in the work, leading to better models of wastewater treatment plants.

9.2.6. The imprecision associated with wastewater process data are better addressed using fuzzy inference modelling techniques. However, while the use of fuzzy logic in engineering applications is numerous, there remains the problem with establishing the membership functions. This difficulty was largely tackled in this work by combining the power of artificial neural networks with the linguistic capability of fuzzy inference system in order to optimise the membership functions and eliminate the traditional trial-and-error procedures. The resulting fuzzy model of the wastewater treatment plant was therefore more accurate but less onerous to develop.

9.2.7. While data driven modelling techniques- ANN, KSOM, fuzzy logic- are relatively adequate in their own right as demonstrated in numerous areas of this thesis, better performance often results when hybrids (or

combinations) of these tools are used. Thus, as shown in the work, combining ANN with KSOM extracted features of the data produced a much more improved model performance than if the ANN was implemented directly on the raw data. A similar observation was observed when ANFIS was combined with the KSOM extracted features. The reason for the enhanced performance when features are used is that the features are essentially "noiseless" data whereas the raw data contain a lot of noise. It is therefore proposed that future modelling using data driven techniques should take advantage of the sort of hybridisation tested in this work.

9.2.8. Because the simulation results are quite encouraging, it is believed that the proposed methodology is generic enough for applying to many other types of WWTP with minor modifications using different set of data. Hence, research will be pursued further in order to test the models in real time operation of activated sludge wastewater treatment plants.

9.3 Recommendation for further work

Despite the success recorded in this work, there are certain aspects which have been identified and would benefit from further investigations. Consequently, the following are suggested as areas for further work:

9.3.1 Although the extracted features of KSOM enhanced the performance of back propagation ANN and ANFIS, the input data cannot propagate directly from the input of the KSOM to an output of the ANN or ANFIS as the task were performed separately in this work. Therefore, developing a single model that incorporates the features and the ANN or ANFIS is one of the major open research issues.

9.3.2. Although satisfactory results have been obtained based on integrated KSOM-ANN and KSOM-ANFIS to deal with noise and missing values, the robustness of the developed models has not yet been analysed mathematically, for example, the effect of the number of missing values in each vector. Such an analysis together with further experimentation would be worth pursuing to prove conclusively the general validity of the proposed schemes.

9.3.2 The trial and error approach developed in this work to optimise the structure of ANN and ANFIS could be replaced by another search technique such as genetic algorithms.

9.3.3 Applying other fault diagnosis techniques could provide more information about the quality of the data and other ways to improve the performance of the developed models.

9.3.4. It seems likely that a relationship between BOD and other relevant information such as conductivity, turbidity, DO, etc. could be used to develop a software sensor and a sensitivity analysis could be performed in order to obtain the best model.

9.3.5. As discussed in the chapter 3, there exist a several types of features extraction techniques, such as PCA. The application of these features extraction on the same data and a comparison with the developed models is a good future research.

9.3.6. Further studies using data from plants with wider ranges of features will ensure that models developed are sufficiently general to be applied to most activated sludge plants. Moreover, new input variables likely to affect the performance of the plant with respect to nitrogen and phosphorus would constitute a logical extension of this work.

9.3.7. Further validation of the model with more unseen data is necessary so as to ensure a rich level of accuracy, consistency and reliability of the models.

9.3.8. The modelled effluent quality in this work was restricted to the BOD_5 and SS. Given other quality parameter are now included in consents, e.g. nutrients, heavy metals, it will be good if models for these are developed as well. This, however, would require dedicated intensive monitoring and huge cost, since some of these variables are not routinely monitored for most treatment works.

9.3.9. Developing the same techniques for different kinds of wastewater treatment plants such as trickling filter, aeration ponds, etc.

9.3.10. Using different neural network structures: the ANNs developed in this study were restricted to just backpropagation neural networks with one hidden layer. While this proved adequate, there are many types of training algorithms possible, e.g. recurrent neural network, radial basis networks, etc., each of which has different advantages and disadvantages. Therefore, it is suggested to use these algorithms in the developed models; if only to test the sensitivity of the results to changes in the model assumptions.

9.3.11. A software sensor for BOD_5 was developed. However, as outlined in the thesis, a number of hardware sensors for BOD_5 have been developed and are being used. A comparative study on the relative efficacy of the software and hardware sensors would also represent an interesting area of further research.

9.3.12. If data-driven modelling techniques are going to have a foothold in wastewater treatment practice, then the data required for calibrating and validating these models must be available. Whilst this study was fortunate to have been supported by an organisation that has invested heavily in short-term, intensive monitoring of its works, this is often an exception and not a rule even in a developed economy such as the UK. Thus another support for further

investment in monitoring wastewater treatment plants so as to make data available for model development is appropriate.

REFERENCES

Abrahart R.J., Kneale P.E., See L.M., 2004. Neural networks for hydrological modelling. Publisher: Balkema.

Adeloye A.J., De Munari A., 2006.Artificial neural network based generalized storage–yield–reliability models using the Levenberg–Marquardt algorithm. Journal of Hydrology, 326(1-4): 215-230.

Alarie R.L., McBean E.A. and Farquhar G.J., 1980. Simulation modeling of primary clarifiers, J. Environ. Eng., 106, 293-309.

Alhoniemi E., Hollmen J., Simula O. and Vesanto J., 1998. Process monitoring and modelling using the self organising map. Helsinki University of Technology, Laboratory of computer and information science, Finland.

Alhoniemi E., Simula O., and Vesanto J., 1997. Analysis of Complex Systems Using the Self Organising Map. Helsinki University of Technology, Laboratory of computer and information science, Finland.

Andrews J.F., 1992. Modeling and Simulation of Wastewater Treatment Systems, Wat. Sci . Tech., 28 (11-12):141-150.

Andrews J.F., 1994. Dynamic control of wastewater treatment plant, Environmental science and technology, 28(9): 434-440.

APHA, 1998. Standard methods for the Examination of Water and Wastewater; 20th ed, American Public Health Association, United Book Press: Maryland, USA.

Arbib M.A, 2003. The handbook of brain theory and neural networks. 2nd ed. MIT.

Astel A., Tsakovski S., Barbieri P., Simeonov V., 2007. Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. Water Research, 41(19): 4566-4578.

Back B., Sere K. and Hanna V., 1998. Managing complexity in large database using self organising map. Accounting Management & Information Technologies, 8:191-210.

Badekas E., Papamarkos N., 2007.Optimal combination of document binarization techniques using a self-organizing map neural network. Engineering Applications of Artificial Intelligence, 20(1): 11-24.

Barnett V. and Lewis T., 1994. Outliers in statistical data. (3rd ed), John Wiley & sons Ltd, Chichester, UK.

Barreto G. A. and Araujo A. F. R., 2004. Identification and control of dynamical systems using the self-organising map. IEEE Transaction on Neural Networks, 15(5):1244-1259.

Basheer A. and Najjar M., 1996. Predicting dynamic response of adsorption columns with neural nets, J. Computing in Civil Engineering, ASCE, 10 (1): 31-39.

Beck M.B., 1984. Modeling and Control studies of the activated sludge process at Norwich Sewage Works. Transactions Institute Measurement and Control, 6(3):117–131.

Beck M.B., 1986. Identification, estimation and control of biological waste-water treatment processes. In: IEE Proc. 133: 254–264.

Béline F., Boursier H., Daumer M.L., Guiziou F. and Paul E., 2007. Modelling of biological processes during aerobic treatment of piggery wastewater aiming at process optimisation. Bioresource Technology. 98(17): 3298-3308.

Beltratti A., Margarita S., Terna, P., 1996. Neural Networks for Economic and Financial Modelling. International Thomson Computer Press, London and Boston.

Bishop C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press, Oxford, UK.

Bistrup S.M., Grady C.P.L., 1988. SSSP-Simulation of single sludge processes. J. Wat. Pollut. Control Fed., 60 (3): 351-361.

Boger Z, 1992. Application of neural networks to water and wastewater treatment plant operation, in Applying Inst. and Aut. in Env .Eng., edited by Alleman, J.E., Sweeney, M.W., and Vaccari, D.A., ISA Transactions, 31,1, 25-33.

Boscolo A., Mangiavacchi C., Drius F., Rongione F., Pavan F., and Cecchi,F., 1993. Fuzzy control of an anaerobic digester for the treatment of the organic fraction of the MSW. Water Sci . and Technol., 27(2), 5768-5776.

Box, G. E. P. and Jenkins, G. M.,1976. Time Series Analysis: Forecasting and Control , Holden Day, San Francisco.

Brydon, D. A.; Frosham, D. A. (2001) A Model-based Approach to Predicting BOD5 in Settled Sewage. Wat. Sci. Tech., 44 (2-3), 9-16.

Capodaglio A. G., Jones H. V., Novotny V. and Feng X., 1991. Sludge bulking analysis and forecasting application of system identification and artificial neural computing technology. Water Res. 25, 1217–1225.

Capodaglio A.G., 1994. Evaluation of modelling techniques for wastewater treatment plant automation. Water Sci Tech 30 (2) pp. 149–156.

Carrasco E. F., Rodriguez J., Punal A., Roca E. C. and Lema J.M ., 2002. Rule-based diagnosis and supervision of a pilot scale wastewater treatment plant using fuzzy logic techniques . Exp. Syst. Appt., 22, pp. 11-20 .

Chang F.J., Chang Y.T., 2006. Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. Advances in Water Resources, 29(1): 1-10.

Chee G.J, Nomura Y. and Karube I., 1999. Biosensor for the estimation of low biochemical oxygen demand, Analytica Chimica Acta, 379(1-2): 185-191.

Chen J, Beck M.B., 1993. Modelling, control and offline estimation of activated sludge bulking. Water science and technology, 28 (11,12), pp. 249-256.

Chen J.C., Chang N.B., Shieh W.K., 2003. Assessing Wastewater reclamation Potential by neural network model . Engineering Applications of Artificial Intelligence 16, 149-157.

Cheng C.Y. and Ribarova I, 1999. Activated sludge system modelling and simulation for improving the effluent water quality. Wat. Sci. Tech, 39(8):93-98.

Chiang Y.M, Chang L.C. and Chang F.J., 2004. Comparison of static feedforward and dynamic feedback neural networks for rainfall-runoff modeling. J. Hydrol. 290 *(3–4): 297–311.*

Choi D-J.; Park H., 2001. A hybrid artificial neural network as a software sensor for optimal control of a wastewater treatment process. Water Research, 35(16), November 2001 , pp. 3959-3967.

Cinar O., 2005. New tool for evaluation of performance of wastewater treatment plant: Artificial neural network. Process Biochemistry 40 pp. 2980-2984.

Cohen A., Janssen G., Brewster S. D., Seeley R., Boogert A. A., Graham A. A., Mardani, M. R., Clarke, N., Kasabov, N. K., 1997. Application of computational intelligence for on-line control of a sequencing batch reactor (SBR) at Morrinsville sewage treatment plant. Wat. Sci. Tech., 35, 63.

Collins M., 1990. Empiricism strikes back: neural networks in biotechnology. Biotechnology 11 , pp. 163–166.

Cooper P.F., 2002. The DESAR Concept for Environmental Protection: Historical aspects of wastewater treatment. In the 'Decentralised Sanitation and Reuse: Concepts, Systems and Implementation'. Water Intelligence Online , April 2002.

Cote M., Grandijean B.P.A., Lessard P.and Yhibault J.,1995. Dynamic modeling of the activated sludge process: improving prediction using neural networks, Water Res. 29 , pp. 995–1004.

Czogala E. and Rawilk T.,1989. Modeling of a fuzzy controller with application to the control of biological process. Fuzzy Sets and Systems 31, pp. 13–22.

Daliakopoulos I.N., Coulibaly P., Tsanis I.K., 2005.Groundwater level forecasting using artificial neural networks. Journal of Hydrology, 309, (1-4): 229-240.

Demuth, H. & Beale, M., 1998.) Neural Network Toolbox, For Use with MATLAB. The Math Works, Inc., User's Guide, Version 4.Natick.

Diamantopoulou M. , Papamichail D. , Antonopoulos V., 2005. The Use of a Neural Network Technique for the Prediction of Water Quality Parameters. International Journal of Operational Research, 5 (1),paper No.9.

Dick, R.I., Young, K.W., 1972. Analysis of thickening performance of final settling tanks. Proceedings of the 27th Industrial Waste Conference, Purdue University, Lafayette, Indiana, USA.

Du Y.G. Tyagi R.D. and Bhamidimarri R., 1999. Use of fuzzy neural-net model for rule generation of activated sludge process. Process Biochemistry, 35(1): 77-83.

Dupont R. and Dahl C.,1995. A one-dimensional model for a secondary settling tank including density current and short-circuiting. Wat Sci Tech, 31(2), 215–224.

Dupont R. and Henze M., 1992. Modelling of the secondary clarifier combined with the activated sludge no.1. Wat. Sci. Technol. 25, pp. 285–300.

EEC, 1991. Directive concerning urban wastewater treatment (91/271/EEC). European Economic Community (EEC); Official Journal L135/40; http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31991L0271:EN:HTML (accessed December 1, 2006).

Ellis, G. W.; Ge, X.; Grasso, D., 1990. Time Series Analysis of Wastewater Quality, Proceedings of the 5th International Association on Water Pollution Research and Control (IAWPRC) Workshop on Instrumentation, Control and Automation of Water and Wastewater Treatment and Transport Systems. July 26-August 3; IAWPRC: Tokyo, Japan, 595-600.

Estaben M., Polite M., and Steyer J.P.,1997. Fuzzy Control For An Anaerobic Digester. Control Eng . Practice , Vol .5 , No. 98, pp. 1303-1303.

Esteves S., 2002. Monitoring and control of biological textile wastewater treatment using artificial neural networks. PhD thesis, university of Glamorgan,UK.

Fallon, A. and Spada, C., Detecting and accommodation of outliers in normally distributed data sets. http://ewr.cee.vt.edu/environmental/teach/smprimer/outlier/outlier.html.

Ferrer J., Rodrigo M.A., Seco A. and Penya-roja J.M., 1998. Energy Saving in the aeration process by fuzzy logic control. Water science and technology, 38 (3),pp. 209-217.

Firat M., Güngör M., 2007. River flow estimation using adaptive neuro fuzzy inference system. Mathematics and Computers in Simulation, 75(3-4): 87-96.

Fu CS, Poch M ., 1995(a). System identification and real-time pattern recognition by neural networks for an activated sludge process. Environment International 21 (1), pp. 57–69.

Fu C.S. and Poch M., 1995(b). Fuzzy modeling pattern recognition for dynamic process and its application for an activated sludge process. Chem Eng Sci 50 (23), pp. 371–383.

Fu C.S. and Poch M., 1998. Fuzzy model and decision of COD control for an activated sludge process. Fuzzy Sets and Systems 41 pp. 281–292.

Fuzzy Logic Toolbox for use with MATLAP, user's Guide, Version 2. June, 2004.

Gagne F.; Blaise C., 1997. Predicting the toxicity of complex mixtures using artificial neural networks, Chemosphere, Volume 35, Number 6, September 1997 , pp. 1343-1363.

Gamal El-Din A., and Smith D.W., 2002. A neural network model to predict the wastewater inflow incorporating rainfall events . Water Research 36 , 1115 – 1126.

Garcia, H. &Gonzalez, L. 2004. Self-organizing map and clustering for wastewater treatment monitoring. Eng. Appl. Artific. Intellig. 17(3), 215–225.

Gerardi M and Wiley A., 2002. Settleability problems and loss of solids in the activated sludge process. Wiley& sons Publication, USA.

Gernaey K.V., Loosdrecht M.C. M., Henze M., Lind M., Jørgensen S.B., 2004. Activated sludge wastewater treatment plant modelling and simulation:state of the art.Environmental Modelling & Software, 19(9): 763-783.

Geselbracht, J.J., Downey Brill, E., Pfeffer, J.T., 1988.Rule Based Model of Design Judgment about Sludge Bulking , Journal of Environmental Engineering, Vol .114, No.1,February ,PP.54 – 73 .

Giokas D. L., Kim Y., Paraskevas P. A., Paleologos E. K., Lekkas T. D., 2002.A simple empirical model for activated sludge thickening in secondary clarifiers. Water Research, 36(13): 3245-3252.

Giraldo-Gomez , E. and Duque , M. , 1998 . Automatica startup of a high rate anaerobic reactor using a fuzzy logic control system . fifth Latin American Workshop Seminar Wastewater Anaerobic Treatment . October, 2730th . Vina del Mar, Chile.

Gleick P. H., 1996. Water resources, In Encyclopedia of Climate and Weather, ed. by S. H. Schneider, Oxford University Press, New York, vol. 2, pp.817-823.

Gonzalez I.M., Garcia H.L., 2006. End-point detection of the aerobic phase in a biological reactor using SOM and clustering algorithms. Engineering Applications of Artificial Intelligence, Volume 19, Issue 1, pp. 19-28.

Gray N.F., 2004. Biology of wastewater treatment, 2nd ed. Imperial College, UK.

Gujer W, Henze M, Mino T and van-Loosdrecht M., 1999. Activated sludge model No. 3. Wat Sci Tech 39(12), 183- 191.

Gujer W., and Henze M., 1991. Activated Sludge Modelling and Simulation. Wat.Sci.Tech., *23: 1011-1023.*

Gurney K., 1997. An introduction to neural networks. Publisher: UCL, 1997.

Guyon I., Elisseeff A., 2003. An Introduction to Variable and Feature Selection. Journal of Machine Learning Research, 3 : 1157-1182.

Hagan, M. T., Demuth, H. B. & Beale, M. 1996. Neural Network Design. PWS Publishing Company, Boston.

Hamed M ., Khalafallah M., and Hassanien E., 2004. Prediction Of Wastewater Treatment Plant Performance using artificial neural networks. Environment Modelling &Software 19 ,919-928.

Han, J. and Kamber, M., 2001. Data Mining: concepts and techniques. , Morgan Kaufmann, Los Atos, CA.

Harremoës, P., Capodaglio, A. G., Hellstrom, B. G., Henze, M., Jensen, K. N., Lynggaaard-Jensen, A., Otterpohl, R., Soeborg, H., 1993. Wastewater treatment plants under transient loading-performance, modeling and control. Wat. Sci. Tech., 27 (12): 71- 79.

Harvey, A. C., 1989. Forecasting structural time series models and Kalman filter. Cambridge University Press, Cambridge, England.

Henze, M., Grady Jr., C.P.L., Gujer, W., Marais, G.V.R., Matsuo, T., 1987. Activated Sludge Model No. 1. IWA scientific and technical report No. 1, London.

Henze, M., Gujer, W., Mino, T. and Loodrecht, M.,2000. Activated Sludge Models: ASM1, ASM2, ASM2D and ASM3. Rep. No. 9, IWA, London.

Henze, M., Gujer, W., Mino, W., Matsuo, T., Wentzel, M.C., Marais, G.v.R., 1995. Activated Sludge Model No. 2. Scientific and Tech. Report No.3,IAWQ,London.

Hill A. and Hare G, 1999. The almond valley and Seafield PFI project risk issues associated with flow prediction. WaPUG Autumn Meeting, 1999.

Hiraoka, M.; Tsumura, K.; Fujita, I.; Kanaya, T. 1990. System Identification and Control of Activated Sludge Process by Use of Autoregressive Model, Instrumentation, Control and Automation of Water and Wastewater Treatment and Transport Systems. Advanced Water Pollution Control; Report No. 10; pp. 121-128; Briggs, R. (ed.); Pergamon; IAWPRC: London, UK.

Hong, Y., Rosen, M., Bhamidimarri, R., 2003. Analysis of a municipal wastewater treatment plant using a neural network-based pattern analysis. Water research 37, 1608-1618.

Hopfield, J. J. ,1982. Neural networks and physical systems with emergent collective computational abilities. In Proceedings of the National Academy of Sciences, pp. 2554-2558. National Academy of Sciences.

Huang, Y. C., Wang , X.Z., 1999. Application of fuzzy causal network to wastewater treatment plants . Chem . Eng. Sci. 54, 2731 – 2738 .

Jang J.S.R., 1993, ANFIS: Adaptive Network Based Fuzzy inference Systems, IEEE Transactions on Systems, Man, and Cybernetics, 23 (3), 665-685.

Jang, J.S.R. & Gulley, N. ,1997. MATLAB/Fuzzy Logic Toolbox, MathWorks. Inc. Natick, MA.

Jang, J.S.R. & Sun, C.T. ,1995. Neurofuzzy modeling and control. Proceedings of the IEEE, 83(3), 378-406.

Jeppson U, 1996. Modelling aspects of wastewater treatment processes, PhD thesis, Lund Institute of Technology, ISBN 91-88934-00-4.

Jeppson U and Diehl S, 1996. an evaluation of a dynamic model for the secondary clarifier. Water science and technology 34(5-6): 19-26.

Kalker T.J.J., Van-Goor C.P., Roeleveld P.J., Ruland M.F. and Babuska R., 1999. Fuzzy Control of Aeration in activated sludge wastewater treatment plant: Design, simulation and evaluation. Water science and technology, 39 (4) pp. 71-78.

Kalteh A.M., Hjorth P., Berndtsson R., 2008. Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application Environmental Modelling & Software, 23(7): 835-845.

Kamarthi S. V., Pittner S., 1999. Accelerating neural network training using weight extrapolations. Neural Networks, 12(9): 1285-1299.

Kangas. J. and Simula, O., 1995. Process monitoring and visualization using self organising map. Chapter 14 in (ed. Bulsari, A. B.) Neural Networks for Chemical Engineers, Elsevier Science Publishers.

Karube, I.; Matsunaga, T.; Mitsuda, S.; Suzuki, S. 1977. Biochemical electrode BOD sensors. Biotechnol. Bioeng. 19 (10), 1535-1547.

Kaski S., Kangas J., and Kohonen T., 1998. Bibliography of Self Organisng Map (SOM) Papers: 1981-1997, Neural Computing Surveys, Vol 1, pp. 102-350.

Keshwani D.R., Jones D.D., Meyer G. E., Brand R.M., 2008. Rule-based Mamdani-type fuzzy modelling of skin permeability. Applied Soft Computing, 8(1): 285-294.

Kim, M. N.; Kwon, H. S., 1999. Biochemical oxygen demand sensor using Serratia marcescens LSY 4. Biosen. Bioelectr., 14 (1), 1-7.

Klir G.J and Folger T.A., 1988. Fuzzy sets, uncertainty and information, Prentice-Hall, 1988.

Kohonen T., 2001. Self-organizing maps, 3rd ed. Berlin, New York: Springer, 2001.

Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. Biological Cybernetics 43, pp. 59–63.

Kohonen, T., Oja, E., Simula, O., Visa, A. and Kangas, J., 1996. Engineering applications of the self-organizing map. Proceedings of the IEEE 84, (10), pp. 1358–1384.

Krebs P., 1995. Success and shortcomings of clarifier modelling. Wat Sci Tech, 31(2): 181–191.

Kristensen N. R., Madsen H., Jørgensen S. B., 2004. A Method for Systematic Improvement of Stochastic Grey-Box Models, Computers and Chemical Engineering, 28(8): 1431-1449.

Kynch G.J., 1952. A theory of sedimentation, Trans. Farad. Soc. 48 : 166--176.

Landeras G., Ortiz-Barredo A., López J.J., 2008.Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain). Agricultural Water Management, Volume 95, Issue 5, May 2008, Pages 553-565.

Langley P., 1994. Selection of relevant features in machine learning. In AAAI Fall Symposium of Relevance, 140-144.

Lee, B.-H.; Scholz, M., 2006a. A Comparative Study: Prediction of Constructed Treatment Wetland Performance with K-nearest Neighbours and Neural Networks. Wat. Air Soil Poll., 174 (1-4), 279-301.

Lee, B.-H.; Scholz M., 2006b. Application of the Self-organizing Map (SOM) to assess the Heavy Metal Removal Performance in Experimental Constructed Wetlands. Wat. Res., 40 (18), 3367-3374.

Lee, H.; Stenstrom, M. K., 2005. Utility of storm water monitoring. Wat. Environ. Res., 77 (3), 219-228.

Lessard P. and Beck M.B., 1991. Dynamic modelling of wastewater treatment processes: its current status. Envir. Sci. Technol. 25 , pp. 30–39.

Lessard P. and Beck M.B., 1993. Dynamic modelling of the activated sludge process: a case study. Water Res 27 , pp. 963–978.

Lessard P. and Beck M.B.,1993. Dynamic modelling of the activated sludge process: a case study. Wat. Res. 27 , pp. 963–978.

Lessard, P. and Beck, M. B., 1988. Dynamic Modelling of Primary Sedimentation. Journal of Environmental Engineering, Division (ASCE), 114, 753-769.

Lighthill J. 1973.Artificial Intelligence: A General Survey in Artificial Intelligence: a paper symposium, Science Research Council.

Liu J. and Mattiasson B., 2002. Microbial BOD sensors for wastewater analysis. Water Research, 36 (15): 3786-3802.

Liu Q. and , Kim S.H., 2008.Evaluation of membrane fouling models based on bench-scale experiments: A comparison between constant flowrate blocking laws and artificial neural network (ANNs) model Journal of Membrane Science, 310 (1-2): 393-401.

Love N.G. and Bott C.B., 2000. A Review and Needs Survey of Upset Early Morning Devices,Alexandria, VA, Water Environment Research Foundation.

MacDonald I. L. and Zucchini W., 1997. Hidden Markov and other models for discrete valued time series. Monographs on Statistics and Applied Probability 70, Chapman & Hall, London.

Mamdani , E.H., 1977. Application of fuzzy logic to approximate reasoning using linguistic systems . Fuzzy Sets Systems, 26, pp. 1182 – 1191 .

Mamdani ,E. H., 1974. Applications of fuzzy algorithms for control of a simple dynamic plant, Proc. Inst. Elelc . Eng., vol . 121, no. 12, pp. 1585-1588.

Mamdani E. H. and Assilian S, 1975. An experiment in linguistic synthesis with a fuzzy logic controller , Int . J . Man Machine Studies , vol. 7 (1) , pp. 1-13.

Manesis S. A., Sapidis D. J. and King R. E., 1998. Intelligent Control of Wastewater treatment plants , Artificial Intelligence in Engineering, 12 (3), pp. 275-281.

Manfred, S., Butler, D. and Beck, M. B., 2002. Modelling, Simulation and Control of Urban Wastewater Systems, 27–29. Springer-Verlag London Limited, 357.

Marsili-Libelli , A. and Muller, A., 1996. Adaptive fuzzy pattern recognition on the anaerobic digestion process . Pattern Recognition Letters, 17, 651.

Maryns F., Bauwens W., 1997. The application of the activated sludge model no.1 to a river environment. Water Science and Technology, 36(5): 201-208.

McBean, E. A. and Rovers, F. A., 1998. Statistical procedures for analysis of environmental monitoring data and risk assessment. Prentice Hall PTA, New Jersey, USA.

McCulloch, W. S. and Pitts, W. H., 1943. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, *5:115-133.*

Metcalf and Eddy, 2003. Wastewater Engineering Treatment and Reuse. Fourth Edition, Mc Graw Hill, New York.

Meyer U. and Popel H.J., 2003. Fuzzy control for improved nitrogen removal and energy saving in WWT plants with predenitrification . Wat.Sci.Technol., 47(11),pp. 69-76 .

Miller A., 2006. An investigation into fuzzy logic for the use in inhabited intelligent environments in particular the use of ANFIS with rich data set. CO600 Group Project Reports 2005/06, Kent university, UK. < http://www.cs.kent.ac.uk/pubs/ug/2006/co620-projects/robofuzz/Final%20Report.pdf>

Mjalli, F., Alasheh S. , and Alfadala .H.E. 2007. Use of artificial neural network black- box modelling for the prediction of wastewater treatment plants performance Journal Of Environment Management 83 (2007) 329-338.

Mohanty, S.; Scholz, M.; Slater, M. J. ,2002. Neural Network Simulation of the Chemical Oxygen Demand Reduction in a Biological Activated Carbon Filter, J. Ch. Inst. Wat. Environm. Managem., 16 (1), 58-64.

Muller A., Marsili-Libelli S., Aivasidis A., Lloyd T., Kroner S., Wandery C.,1998. Fuzzy control of disturbances in a wastewater treatment process. Water Research, 31(12), 3157 – 3167 .

Murnleitner E, Becker T.M. and Degado A., 2002. State detection and control of overloads in the anaerobic wastewater treatment using fuzzy logic. Water research, 36, 201-211.

Nguyen T., Walker E.A., 2006. A first course in fuzzy logic. 3rd ed., Chapman and Hall.

Nikkila J. , 2005. Exploratory Cluster Analysis of Genomic High-Throughput Data Sets and Their Dependencies. PhD thesis, Helsinki University of Technology Publications in Computer and Information Science, Report D11. < http://lib.tkk.fi/Diss/2005/isbn9512279096/isbn9512279096.pdf.

Nokyoo,C., 2002. The use of artificial neural networks in real time forecasting of wastewater treatment plant performance. PhD thesis, University of Newcastle Upon Tyne.

Nowak O., Franz A., Svardal K., Müller V., Kühn V., 1999. Parameter estimation for activated sludge models with the help of mass balances. Water Science and Technology, 39(4): 113-120.

Obu-can, K., Fujimura, K., Tokutaka, H., Ohkita, M., Inui, M. and Ikeda, Y., 2001. Data mining of power transformer database using self-organising map. Tottori University, Dept. Of Electrical and Electronics Engineering, Koyama-Minami, Japan. Www.ele.tottori-u.ac.jp.

Ohki, A.; Shinohara, K.; Ito, O.; Naka, K.; Maeda, S.; Sato, T.; Akano, H.; Kato, N.; Kawamura, Y., 1994. A BOD sensor using Klebsiella oxytoca AS1. Environ. Anal. Chem., 56, 261-269.

Oja M, Kaska S., and Kohonen T., 2003. Bibliography of Self Organisng Map (SOM) Papers: 1998-2001, Neural Computing Surveys, Vol. 3, pp. 1-156.

Olsson G .and Newell R., 1999 . Wastewater Treatment Systems – Modelling, Diagnosis and Control . IWA publishing , London .

Olsson J., Uvo C.B., Jinno K., Kawamura A., Nishiyama K., Koreeda N., Nakashima T. and Morita O.,2004. Neural network for rainfall forecasting by atmospheric downscaling, Journal of Hydrologic Engineering 9 (1):1–12.

Olsson, G., Nielsen, M., Yuan, Z., Jensen, A. L. and J-P Steyer, 2005. Instrumentation, Control and Automation in Wastewater Systems. IWA Publishing, ISBN 1900222833.

Orhan D. and Artan N., 1994. Modelling of Activated Sludge Systems. Technomic Publishing Co., Lancaster, USA.

Otterpohl, R. and Freund, M., 1992. Dynamic models for clarifiers of activated sludge plants with dry and wet weather flows. Wat. Sci. Tech., 26(5–6), 1391–1400.

Paraskevas P. A., Pantelakis I. S., Lekkas T. D., 2003. An advanced integrated expert system for wastewater treatment plants control. Knowledge-Based Systems, 16 (3) 173.

Patry, G. G. and Takács, I. ,1990.. Simulator-based modelling of wastewater treatment plants. In Proceedings of Annual Conference and 1st Biennial Environmental Specialty Conference CSCE, Vol. I, pp. 491-505.

Penn B.S., 2005. Using Self Organizing Maps to Visualize high dimensional Data. Computer and Geoscience 31, pp. 531-544.

Plósz B.G., Weiss M., Printemps C., Essemiani K., Meinhold J., 2007. One-dimensional modelling of the secondary clarifier-factors affecting simulation in the clarification zone and the assessment of the thickening flow dependence.Water Research, 41(15): 3359-3371.

Polit M., Genovesi A., Claudet B.,2001. Fuzzy logic observers for a biological wastewater treatment process. Applied Numerical mathematics 39, pp. 173- 180.

Powell W.B., 2007. Approximate Dynamic Programming: Solving the Curses of Dimensionality. Wiley, ISBN 0470171553.

Poznanski R.R., 2001. Biophysical neural networks: foundations of integrative neuroscience. Publisher: Mary Ann Liebert.

Praet E., Reuter V., Gaillard T., Vasel J.L.,1995. Bioreactors and biomembranes for biochemical oxygen demand estimation . Trends in analytical chemistry, 14 (7).

Pu H. and Hung Y., 1995a. Use of artificial neural networks: predicting trickling filter performance in a municipal wastewater treatment plant, Environ. Manage. Health 6 , pp. 16–27.

Pu, H. and Hung , Y. 1995b. Artificial Neural Networks for Predicting Municipal Activated Sludge Wastewater Treatment Plant Performance , Intl. J. Environmental Studies, 48, 97 – 116.

Pulido-Calvo I., Montesinos P., Roldán J., Ruiz-Navarro F., 2007. Linear regressions and neural approaches to water demand forecasting in irrigation districts with telemetry systems. Biosystems Engineering, 97(2): 283-293.

Pulido-Calvo I. and Gutiérrez-Estrada J.C., 2008.Improved irrigation water demand forecasting using a soft-computing hybrid model Biosystems Engineering, In Press, Corrected Proof, Available online 30 November 2008.

Punal , A., Rodriguez ,J., Carrsco , E. F., Roca, E . and Lema , J. M ., 2002 . Expert system for the online diagnosis of anaerobic wastewater treatment plants. Wat. Sci. Tech., 45(10), 195 – 200 .

Raduly B., Gernaey K.V., Capodaglio A.G., Mikkelsen P.S., and Henze M., 2007. Artificial neural networks for rapid WWTP performance evaluation. Environment Modelling &Software 22,1208- 1216.

Riedel, K.; Renneberg, R.; Kuhn, M.; Scheller, F., 1988. A fast estimation of biochemical oxygen demand using microbial sensors. Appl. Microbiol. Biotechnol., 28, 316-318.

Rivas A., Irizar I., Ayesa E., 2008. Model-based optimisation of Wastewater Treatment Plants design. Environmental Modelling & Software,23(4): 435-450.

Rocha M., Cortez P., Neves J., 2007. Evolution of neural networks for classification and regression. Neurocomputing, 70,(16-18): 2809-2816.

Roche, P.; Al-Jowder, R.; Narayanaswamy, R.; Young, J.; Scully, P. ,2006. A novel luminescent lifetime-based optrode for the detection of gaseous and dissolved oxygen utilising a mixed ormosil matrix containing ruthenium (4,7-diphenyl-1,10-phenanthroline)(3)Cl-2 (Ru.dpp). Analyt. Bioanalyt. Chem., 386, 1245-1257.

Rosen C and Lennox A.J., 2001. Multivariate and multiscale monitoring of wastewater treatment operation. Wat. Res. 35 (14) pp. 3402–3410.

Rosen, C., 1998. Monitoring wastewater treatment system. PhD thesis, Department of industrial Electrical Engineering and Automation, Lund Institute of technology, Lund University, Sweden.

Rosenblatt, F., 1958. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. Psychological Review, Vol 65, pp 386-408. Reprinted in Anderson and Rosenfeld, 1988, pp 92-114.

Ross J.T, 2004. Fuzzy logic with engineering applications, 2nd ed., Wiley.

Rustum R and A.J.Adeloye, 2007(a). Replacing outliers and missing values from activated sludge data using Kohonen Self Organizing Map. Journal of Environmental Engineering, 133 (9): 909-916.

Rustum R and Adeloye A.J., 2006(a). Knowledge Discovery from Activated Sludge Process Using Unsupervised Neural Networks (Kohonen Self-Organizing Map). 3rd international symposium on integrated water resources Management, 26-28 September, Ruhr University, Bochum, GERMANY (Poster presentation).

Rustum R and Adeloye A.J., 2007(b). Modelling activated sludge process using adaptive network-based fuzzy inference system. Presented at the 10th IWA Specialised Conference on Design, Operation and Economics of Large Wastewater Treatment Plants, 9-13 September, Vienna, Austria:101-104.

Rustum R. and Adeloye A. J., 2006(b). Features extraction from primary clarifier data using unsupervised neural networks (Kohonen Self Organising Map). In the 7th International Conference on Hydroinformatics, HIC 2006, Nice, France.

Rustum R. and Adeloye A.J., 2006. Features Extraction from Primary Clarifier data using Kohonen Self Organizing Map, Unsupervised Neural Network. Presented at the 7th International Conference on Hydroinformatics, HIC, September 2006, Nice, FRANCE.

Rustum R., Adeloye A.J., and Scholz M. , 2008a. Applying Kohonen Self-organizing Map as a Software Sensor to Predict the Biochemical Oxygen Demand. Water Environment Research, 80 (1): 32-40.

Rustum R., Adeloye A.J., and Simala A., 2007(c). Kohonen self-organizing map (KSOM) extracted features for enhancing MLP-ANN prediction models of BOD5. Water Quality and Sediment Behavior of the Future: Predictions for the 21st Century, IAHS Publ., Red Book 314, pp. 181-187.

Sadiq R., AL-zahrani M., Sheikh AK., Husain T., Farooq S., 2004. Performance evaluation of slow sand filters using fuzzy rule-based modelling . Environmental Mmodelling&Software 19 , pp. 507- 515.

Sahiner B., Chan H.P., Hadjiiski L., 2008. Classifier performance estimation under the constraint of a finite sample size: Resampling schemes applied to neural network classifiers. Neural Networks, 21 (2-3):476-483.

Sakai, Y.; Abe, N.; Takeuchi, S.; Fermen, J.,1995. BOD sensor using magnetic activated sludge. J. Fermentation Bioeng., 80 (3) 300-303.

Sanchez E.N., Beteau J.F., and Carlos-Hernandez S.,2001. Fuzzy Supervisory For A Wastewater Anaerobic Treatment Plant. Proceeding of the 2001 IEEE International Symposium on Intelligent Control September, 5 – 7 .

Sanchez E.N., Beteau J.F., Vera G., and Cadet C., 2003. Fuzzy Control of An Activated Sludge Wastewater Treatment Plant. Proceeding of the American Control Conference Denver , Colorado June 4-6 , 2003.

Schilling W., 1994. Smart sewer systems improved performance by real time control. Eur. Water Pollution control, 4(5): 24-31.

Scholz, M., 2006. Wetland Systems to Control Urban Runoff; Elsevier: Amsterdam, The Netherlands.

Seeger H., 1999. The history of German waste water treatment. http://www.valt.helsinki.fi/projects/enviro/articles/Seeger.pdf.

SEPA, 2006. THE URBAN WASTE WATER TREATMENT (SCOTLAND) REGULATIONS (www.sepa.org.uk/pdf/guidance/urbanwastewaterregs.pdf)

Shankir, Y., 2000. Fuzzy logic Systems and fuzzy neural networks for dynamic systems modelling and control. PhD thesis, university of Wales Cardiff, intelligent system laboratory, Cardiff school of engineering, UK.

Shen, J.; Shen, W.; Chang, J.; Gong, N. (2006) Fuzzy Neural Network for Flow Estimation in Sewer Systems During Wet Weather. Wat. Environ. Res., 78 (2), 100-109.

Spellman, F. R., 2003. "Hand book of water and wastewater treatment plant operation". Lewise publishers, USA.

Steyer, J.P., Buffiere, P, Rolland , D. and Moletta , R.(1999). Advanced control of anaerobic digestion processes through disturbances monitoring . Wat. Res., 33(9), 2059 – 2068 .

Stokes, A., 1998. Modelling and management of the activated sludge process. University of Birmingham, school of civil enginnering, UK. PhD thesis.

Sugeno, M. & Kang, G.T., 1988. Structure identification of fuzzy model.Fuzzy Sets and Systems, 26, 15–33.

Tailor, 2005, Special communication with Mr Doncan Tailor, operational director at Seafiedl wastewater treatment plant, Edinburgh, UK.

Takacs I., Patry G.G., and Nolasco D.,1991. A Dynamic model of the clarification thickening process. Water Research, 25 (10): 1263-1271.

Takagi T. and Sugeno M., 1985. Fuzzy identification of systems and its application to modeling and control. IEEE Trans Syst, Man, Cybern 15 (1) pp. 116–132.

Tan, T. C.; Li, F.; Neoh, K. G.; Lee, Y. K. ,1992. Microbial membrane-modified dissolved oxygen probe for rapid biochemical oxygen demand measurement. Sens. Actuat., 8, 167–172.

Tasi Y.P., Ouyang C.F., Wu M.Y., Chiang W. L .,1993. Fuzzy control of dynamic activated sludge process for the forecast and control of effluent suspended solid concentration , Wat . Sci . Tech ., Vol. 28, Nr 11 – 12 , pp. 355-367 .

Tay and Zhang, 1999. J. Tay and X. Zhang, Neural fuzzy modelling of anaerobic biological wastewater treatment systems. J. Env. Eng. 125 12 (1999), pp. 1149–1159.

Tchobanoglous G. and Burton F. L., 1991. Waste-water Engineering: Treatment, Disposal and Reuse, eds Staff Mecalf and Eddy, McGraw-Hill Series in Water Resources and Environmental Engineering, New York.

Tomiello M., Perrin E., Roubens M., Crine M .,1999. Fuzzy Control Of An Activated Sludge Process . ECCE2-Second European Congress of Chemical Engineering- Montpellier 1999.

Tong R .M.., Beck M . B., Latten A .,1980. Fuzzy Control of the Activated Sludge Wastewater Treatment Process, Automatica , vol . 16, pp . 695 – 701 .

Torkkola K., 2003. Feature Extraction by Non Parametric Mutual Information Maximization. Journal of Machine Learning Research, 3 :1415-1438.

Traore A. , Grieu S., Thiery F., Polit M., Colpirm J., 2006. Control of sludge height in secondary settler using fuzzy algorithms . Computers and Chemical Engineering 30,1235-1242.

Tsai Y.P., Ouyang C.F., Wu M.Y., Chiang W. L .,1993. Fuzzy control of dynamic activated sludge process for the forecast and control of effluent suspended solid concentration , Wat . Sci . Tech ., Vol. 28, Nr 11 – 12 , pp. 355-367 (1993).

Tsai Y.P., Ouyang C.F., Chiang W.L. and Wu M.Y. , Construction of an on-line fuzzy controller for the dynamic activated sludge process. Water Res 28 4 (1994), pp. 913–921.

Tsai Y.P., Ouyang C.F., Wu M.Y. and . Chiang W.L ,1996. Effluent suspended solid control of activated sludge process by fuzzy control approach. Water Environ Res 68 (6) pp. 1045–1053.

Tyagi R.D. and Du Y.G., 1992. Kinetic model for the effects of heavy metals on activated sludge process using neural networks. Environ Technol 13 , pp. 883–890.

Tyagi R.D., Du G.Y., Sreekrishnan, T.R. and Villeneuve J.P. (1993). Neural Model for the Operational Control of Activated ludge Process. Process Biochemistry, 28, 259-267.

Vanrolleghem P. A. 2001. "The usefulness of models in wastewater engineering". University OF Gent, Department of applied mathematics, biometrics, and process control. Lecture held at the Institute for Urban Water Management, Dresden University of Technology, Dresden, Germany.

Vanrolleghem P., Dries D., Verstreate W., 1990. Rapid determination of biochemical oxygen demand. Proceeding of the 5th Eurobian Congress Biotechnology Copenhagen, Denmark, pp 161-164.

Vanrolleghem P., Van-Der-Schueren D, Krikilion G, Grijspeerdt K, Willems P, and Verstraete W, 1996. Online quantification of settling properties with In-Sensor-Experiments in an automated settlometer. Water Sci. Technol. 33(1) 37-51.

Vesanto J. , 1999. SOM-based data visualization methods. Intelligent Data Analysis, 3 , pp. 111–126.

Vesanto J., Himberg J., Alhoniemi E. and Parhankangas J.,2000. Self Organising Map in Matlab: the SOM Toolbox. Helsinki University of Technology, Laboratory of computer and information sience.Finland. February 18, ISBN 952 22 4951 0.

Vesanto, J., Alhoniemi, E., Himberg, J., Kiviluoto, K., Parviainen, J., 1999. Self-organizing map for data mining in MATLAB: the SOM toolbox. Simulation News Europe, pp. 25–54.

Vos N. J. and Rientjes T. H. M., 2005. Constraints of artificial neural networks for rainfall-runoff modelling: trade-offs in hydrological state representation and model evaluation. Hydrol. Earth Syst. Sci. Discuss., 2, 365–415, 2005, www.copernicus.org/EGU/hess/hessd/2/365/, European Geosciences Union, Hydrology and Earth System Sciences Discussions. http://www.hydrol-earth-syst-sci-discuss.net/2/365/2005/hessd-2-365-2005-print.pdf.

Wange B., McKay R.I., Hussein A.A., Barlow M. 2003. A Comparative Study for Domain Ontology Guided Feature Extraction. Twenty-Fifth Australian Computer Science Conference (ACSC2003), Adelaide, Australia. Conferences in Research and Practice in Information Technology, Vol. 16. Michael Oudshoorn, Ed. http://crpit.com/confpapers/CRPITV16BWang.pdf

Watanabe S., Baba K., Yoda M., Wu W.C., Enbutsu I., Hiraoka M.and Tsumura K., 1993. Intelligent operation support system for activated sludge process. Water Sci Technol 128 11 , pp. 325–332.

Weijers S.R. and Vanrolleghem P.A., 1997. A procedure for selecting best identifiable parameters in calibrating activated sludge model no. 1 to full-scale plant data . water science and technology, 36(5): 69-79.

Wen C.H., Vassiliadis C.A., (1998). Applying hybrid artificial intelligence techniques in wastewater treatment . Engineering Applications of Artificial Intelligence 11 (1998) 685 – 705.

Werbos, P. 1974. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. Ph.D Thesis, Cambridge, MA: Harvard U. Committee on Applied Mathematics.

Yi C, Cao Z, Kandel A., 1990. Application of fuzzy reasoning to the control of an activated sludge plant . Fuzzy Sets and Systems;38:1 – 14 .

Zadeh L.A., 1965. Fuzzy sets. Inf. Control 8 pp. 338–353.

*Zadeh L.A., 2008.*Is there a need for fuzzy logic? *Information Sciences, 178 (13):2751-2779.*

Zhang L. , Scholz M. , Mustafa A., Harrington R, 2009. Application of the self-organizing map as a prediction tool for an integrated constructed wetland agroecosystem treating agricultural runoff, Bioresource Technology, 100(2): 559-565.

Zhao, H., Hao, O.J., Fellow, A.S.C.E., McAvoy, T.J., Change, C.H., 1997. Modeling nutrient dynamics in sequencing batch reactor. J. Environ. Eng .123 (4), 863 – 879 .

RESUME

Rabee Rustum is an Environmental engineer specialized in hydroinformatics with experience in water and wastewater systems. He is registered as a Professional Engineer in Syria and he is listed on an International Marquis who's who in the world, 2009.

He has a Master Degree from Poitiers University, France in Environmental Risk Management, 2003; Post Graduate Diploma (First Class Position), Civil and Environmental Engineering, Damascus University, Syria. 2002; B.eng. Civil and Environmental Engineering, Damascus University, Syria, 2001. He was awarded the prize of Martyr Basel Al_Asad for Studying Mastery, 2000.

He has experience in municipal drinking water, wastewater, hydraulic modelling and researching as well. He is a member of Syrian Body of Engineers, Syrian Body of teachers and Institute of Civil Engineers.