Department of Chemical Engineering

An Integrated Approach to Artificial Neural Network based Process Modelling

Agus Saptoro

This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University of Technology

March 2010

Statement of Originality

To the best of my knowledge and belief, this thesis contains no material previously published by any other person except where due acknowledgement has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature :

Date : 26th March 2010

SUMMARY

ANN technology exploded into the world of process modelling and control in the late 1980's. The technology shows great promise and is seen as a technology that could provide models for most systems without the need to understand the fundamental behaviour or relationships among the process variables. Today, ANN applications have been applied successfully in a number of areas of process modelling and control, with the best-established applications being in the area of inferential measurements or soft sensors.

Unfortunately, 'the free lunch did not have much meat'. Overtime, people focused more on the true capabilities and power of ANN, the ability to model nonlinear relationships in data without having to define the form of the nonlinearity. However, there is often a tendency to merely plug in the data, turn the ANN training software on, and blindly accept the results. This is probably inevitable since, to date, there are no textbooks or scientific journal papers providing an integrated and systematic approach for ANN model development addressing pre-modelling, training and post-modelling stages. Therefore, addressing issues in those three phases of ANN model development is essential to support and to improve further applications of ANN technology in the area of process modelling and control.

The model development issues in pre-modelling and training phases were addressed by reviewing current practice and existing techniques. For each issue, a novel method was proposed to improve the performance of ANN models. The new approaches were tested in a variety of benchmarking studies using artificial samples and coal property datasets from power station boilers.

The research work in the post-modelling stage analysis which emphasises on taking the lid off black box model, proposes a novel technique to extract knowledge from the models and simultaneously obtain better understanding of the process. Post-modelling phase issues were addressed thoroughly including construction of prediction limit, sensitivity analysis and development of mathematical representation of the trained ANN model.

Confidence intervals of the ANN models were analysed to construct the prediction boundary of the model. This analysis provides useful information related to interpolation and extrapolation of the model. It also highlighted how good the ANN models can be used for extrapolation purposes.

An effort based on sensitivity analysis of hidden layers is also proposed to understand the behaviours of the ANN models. Using this technique, knowledge and information are retrieved from the developed models. A comparative study of the proposed techniques and the current practice was also presented.

The last topic addressed in this thesis is knowledge extraction of ANN models using mathematical analysis of the hidden layers. The proposed analysis is applied in order to open the black box of the ANN models and is implemented to simulated and real historical plant data so that useful information from those data and better understanding of the process are obtained.

All in all, efforts have been made in this thesis to minimise the use of abstract mathematical language and in some cases, simplify the language so that ANN modelling theory can be understood by a wider range of audience, especially the new practitioners in ANN based modelling and control. It is hoped that the insight provided in the dissertation will provide an integrated approach to pre-modelling, training and post-modelling stages of ANN models. This 'new guideline' of ANN model development is unique and beneficial, providing a systematic framework for the preparation, design, evaluation and implementation of ANN models in process modelling and control in particular and prediction / forecasting tool in general.

Keywords: artificial neural networks, pre-modelling, modelling, post-modelling, opening the black box, an integrated approach.

Brief Biography of the Author

Agus Saptoro received the Bachelor of Engineering in Chemical Engineering with academic distinctions from the Faculty of Engineering Gadjah Mada University, Indonesia in 2002. Prior to joining Department of Chemical Engineering National Institute of Technology Indonesia in 2003, he worked as a process control engineer at a Polymer Fibre Plant in Semarang, Indonesia. He received Master degree scholarship from Asian Development Bank through TPSDP project to pursue Master of Engineering at Department of Chemical Engineering Curtin University of Technology which commenced in March 2005. In September 2006, his master program was converted into PhD degree and his research project was supported by Curtin International Research Tuition Scholarship from March 2007 to December 2008. He currently holds the position of lecturer at Department of Chemical Engineering Curtin University Technology, Sarawak, Malaysia.

Journal Publications arising from this Thesis

Saptoro, A., Vuthaluru, H.B. and Tadé, M.O., 2009, "Relation analysis among coal properties: the use of grey superior analysis and artificial neural networks", Environmental Modelling and Software (submitted).

Saptoro, A., Yao, H.M, Tadé, M.O. and Vuthaluru, H.B., 2008, "Prediction of coal hydrogen content for combustion control in power utility using neural network approach", Chemometrics and Intelligent Laboratory Systems, 94 (2), 149-159.

Refereed Conference Publications arising from this Thesis

Saptoro, A., Vuthaluru, H.B. and Tadé, M.O., 2008, "Prediction of coal properties using neural networks: a novel approach based on small training data set", CHEMECA, Newcastle, Australia.

Saptoro, A., Vuthaluru, H.B. and Tadé, M.O., 2008, "Optimising artificial neural networks topology using hierarchical two-population genetic algortihm", 12th APCChE Congress, Dalian, China.

Saptoro, A., Vuthaluru, H.B. and Tadé, M.O., 2006, "Partial Least Squares and Neural Networks based Model of Coal-Fired Power Plant Boiler", CHEMECA, Auckland, New Zealand.

Saptoro, A., Vuthaluru, H.B. and Tadé, M.O., 2006, "A Comparative Study of Prediction of Elemental Composition of Coal using Empirical Modelling", IFAC Symposium on Advanced Control of Chemical Processes, Gramado, Brazil.

Saptoro, A., Vuthaluru, H.B. and Tadé, M.O., 2006, "Design of Feed-Forward Neural Networks Architecture for Coal Elemental Composition Prediction", International Conference on Modelling and Simulation, Kuala Lumpur, Malaysia.

Saptoro, A., Vuthaluru, H.B. and Tadé, M.O., 2006, "Prediction and Monitoring of Unburnt Carbon in Fly Ash in Coal-Fired Power Plant", International Conference on Modelling and Simulation, Kuala Lumpur, Malaysia.

ACKNOWLEDGEMENTS

It would be impossible to mention all of the individuals who are supportive, very encouraging and helpful during the research work and the writing of this thesis. There are many people who have contributed to this work in different ways and who deserved my deep gratitude.

First of all, I am extremely grateful to my supervisors, Prof Moses Oludayo Tadé and A/Prof Hari Babu Vuthaluru, for directing the research programme and their continuous encouragement. Their challenging idea and imaginative feedbacks greatly benefited the work.

Many thanks to all research students in the process systems computations group for useful discussion and inputs that make this work stronger and better. Specially, I am deeply indebted to Dr. Hongmei Yao who helped me to grasp the basic concepts of artificial neural networks. I would also like to thank Mr Brad Huppatz from Muja Power Station, Collie, Western Australia for providing any necessary data used in this research.

I will also never forget Brenda Tipping, Jann Bolton and all staffs at Curtin library and international office. Their kind help is greatly appreciated.

Most importantly I wish to thank my father, my father and mother in law and all my brothers and sisters for their enormous supports and encouragement. This work is also dedicated to my late mum, a caring and loving mother, who passed away in the middle of the project. Alison and Eifion Jones, my "aussie mum and dad" will also not be forgotten for their sincere friendship and support during my stay in Perth.

Finally, I will mention the most important person in my life and who was always there for me. Even though, she could not make it until the end, I would like to thank my wife Pipit for her endless support and understanding. I could not have done it without her.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 5

LIST OF TABLES

<u>CHAPTER 4</u>

<u>CHAPTER 6</u>

LIST OF ACRONYMS

ACO	Ant Colony Optimisation
ANN	Artificial Neural Networks
ARD	Automatic Relevance Determination
BPCA	Bayesian Principal Component Analysis
C	Carbon
CSTR	Continuous Stirred Tank Reactor
CV	Calorific Value
DCS	Distributed Control System
E-ANN	Ensemble Artificial Neural Networks
EM	Expectation - Maximisation
ED	Eucledian Distance
FC	Fixed Carbon
FCC	Fluid Catalytic Converter
GA	Genetic Algorithm
H	Hydrogen
initnw	Nguyen-Widrow layer initialisation function
ITA	Information Theoretic Approach
KS	Kennard-Stone
Kurt	Kurtosis
LOGMSE	Mean Squared Error based on Cauchy distribution
logsig	logarithmic sigmoid transfer function
M	Moisture
MCD	Minimum Covariance Determinant
MD	Mahalanobis Distance
MDKS	Kennard-Stone based on Mahalanobis Distance Method
MF	Membership Function
MI	Mutual Information
ML	Maximum Likelihood
MLP	Multilayer Perceptron
MR	Multiple Regression
MSE	Mean Squared Error

MTD	Mega Trend Diffusion
MTD – ANN	Hybrid Mega Trend Diffusion and Artificial Neural Networks
MVE	Minimum Volume Estimator
MVT	Multivariate Trimming
NLPCA	Nonlinear Principal Component Analysis
O	Oxygen
PC	Principal Component
PCA	Principle Component Analysis
PCR	Principle Component Regression
PLS	Partial Least Square
PMI	Partial Mutual Information
PPCA	Probabilistic Principal Component Analysis
PSD	Pseudo Standard Deviation
PSO	Particle Swarm Optimisation
purelin	linear transfer function
RAND	uniformly distributed random number generator
RANDN	normally distributed random number generator
RANDS	symmetric distributed random number generator
RANDNC	normally distributed random number generator that generates a random weight matrix columns of which are normalised to the value of one
RANDNR	normally distributed random number generator that generates a random weight matrix rows of which are normalised to the value of one
RD	Robust Distance
RS	Random Selection
S	Sulphur
SD	Standard Deviation
Skew	Skewness
SOM	Kohonen Self Organising Map
SVD	Singular Value Decomposition
SPXY	Sample Set Partitioning based on Joint X – Y Distances
tansig	hyperbolic tangent transfer function
VM	Volatile Matter

Chapter 1
INTRODUCTION

1.1 Artificial Neural Networks based soft sensors and models of chemical processes

Empirical modelling is a method for the development of models based on process / experimental data. Since physical modelling is not always obtainable and very time consuming, empirical modelling is a more popular method for gaining insights into the overall (input-output) process behaviour. The developed models are usually used for prediction of future process values using historical data or estimation of unmeasured variables using easily measurable variables.

Most chemical and industrial processes exhibit nonlinear behaviour, therefore, empirical nonlinear models are required instead of the linear ones. In this regard, ANN as an empirical nonlinear modelling technique has been used extensively in recent years to model a wide range of physical and chemical phenomena. ANN is very attractive whenever it is necessary to model complex or less understood processes with large input and output datasets, as well as to replace models that are too complicated to solve in real time [1, 2].

ANN technology exploded into the world of process modelling and control in the early 1990's. The technology shows great promise and is seen as a technology that could provide models for most of our systems without the need to understand the fundamental behaviour or relationships of the process. Today, ANN applications have been applied successfully in a number areas of process modelling and control, with the best-established applications being in the area of inferential measurements or soft sensors.

Unfortunately, the free lunch did not have much meat. With time, researchers shifted their focus more to the true capabilities and power of ANN, the ability to model nonlinear relationships in data without having to define the form of the nonlinearity. However, researchers soon realised that blindly applying black box modelling technique may provide a satisfactory fit for historical data but often leads to poor performance for using online in a closed-loop control application, or on time-variant processes. The ANN model is a more black-box approach, making it difficult to extract process understanding from the model, and it is only valid in the operating region over which the data was modelled.

Regardless of the power of ANN to solve problem effectively, there are also many issues associated with a systematic approach of ANN model development. There is often a tendency to merely plug in the data, turn the ANN training software on, and blindly accept the results. On other hand, several questions as listed below arose regarding pre-modelling phases, modelling stages and post-modelling stages.

How do we select representative input variables?

How to deal with outliers/noises, eliminate them or develop robust ANN model against them?

How do we build and apply the ANN model in scenarios where some measurements are missing and the data is incomplete?

How to handle small size of datasets as building a good ANN model requires sufficient historical/experimental data?

Given such available historical/experimental data, how do we divide these datasets into training and testing/validation data?

To ensure consistency among different magnitudes of input variables of varying scales, which data transformation should be used?

To resume the ANN training, which weight initialisation algorithm should be used?

What activation functions should be used for hidden and output layers?

As there are many training algorithms, which one will give faster and better result?

What learning criteria should be used?

How many hidden layers should be used?

Table 1.1 Issues in ANN based modelling.

Issues on ANN based modelling		
Pre-modelling phase	Modelling / Training stage	Post-modelling phase
Data collection and Input Variable Selection	*Weight initialisation*	*Prediction limit / confidence interval*
Outliers /noise	*Choice of activation function*	*Sensitivity analysis*
Incomplete / missing data	*Choice of training algorithm*	*Extracting knowledge from ANN model*
Small sample dataset	*Choice of learning criteria*	
Data splitting	*Number of hidden layers*	
Data transformation		

Unfortunately, to date, there are no textbooks or scientific journal papers providing an integrated and systematic approach of ANN model development addressing all the above questions. Moreover, current ANN technology applications mostly lack post-modelling phase analysis as listed in Table 1.1. Therefore, addressing issues in the three phases of ANN model development is essential to support and to improve further applications of ANN technology in the area of process modelling and control.

1.2 Motivations for this study

ANN models have been traditionally designed without an integrated and systematic approach. Building ANN models sometimes is also facing undesirable conditions such as missing measurement/incomplete data and small sample dataset. Although the model users are satisfied with their predictive performance, understanding of the ANN function in its hidden layer is still poor. Hence the black-box nature results in difficulties to extract process knowledge from the data/models.

The first motivation for this research is to provide an integrated and systematic approach of ANN model development and propose novel / improved approaches to the existing analysis and methods. A second motivation, which drives many ANN researchers and users, is to propose techniques to open the black box of the ANN and simultaneously to extract process knowledge and understanding from the models.

1.3 Objectives and contributions

1.3.1 Objectives

The main objectives of this research are:

- To propose a systematic post-modelling phase analysis of the ANN models. The emphasise, in particular, is on opening the black-box of the ANN model in order to extract its process knowledge and to obtain mathematical relationships inside the hidden layers.
- To highlight integrated approaches of pre-modelling and modelling phases to provide guidance for ANN model development. Some issues in these two stages are thoroughly addressed using existing techniques and novel methods.
- To demonstrate how the proposed approaches are able to enhance the performance of the black box models.

1.3.2 Contributions

This research work in theoretical and sensitivity analysis of hidden layer is a direct contribution to ANN theory, design and its applications. Previously most of the work in ANN only focused on how to blindly develop ANN models and to implement them to predict output variables using given input variables. An integrated approach to pre-modelling, modelling and post-modelling staged of ANN models is also unique and beneficial, providing a systematic framework of the preparation, design, evaluation and implementation of ANN models. This framework will provide an integrated guidance for building ANN models. In addition, the post-modelling stage analysis which emphasise on taking the lid off black box model, will provide a novel technique to extract knowledge from the models and simultaneously to obtain better understanding of the process.

This research and development work has also a great significance with regards to its application in black box modelling approach and ANN predictive control. Having better performance and well understood models will simultaneously benefit the applications of black-box models and model predictive control.

1.4 Scope of the study

The overall scope of this study entails developing a framework for understanding ANN behaviour and extracting knowledge from ANN models. An integrated model development scheme of ANN models is also presented covering existing techniques and novel approaches. The FFANN structure is selected for this research as this structure has been used almost exclusively in many areas of prediction and modelling. Case studies on synthetic problems, coal properties prediction and modelling of coal-fired power plant were used to describe how all of the approaches work.

1.5 Layout of the thesis

This thesis is divided into seven chapters. In chapter two a literature review of the use of ANN, specifically in coal-power stations, is presented. This review also comprehensively covers the research gaps in current status of research on ANN and the roadmap for further research need. The background and motivations for this work are exclusively outlined in this chapter.

Chapter three deals with the issues in pre-modelling phase of ANN models. It provides an integrated scheme of input variable selection and data preparation for model development. Observations to the existing techniques and their novel approaches are performed and their influences on ANN model performance are studied.

Chapter four mainly discusses issues during training stages. This section attempts to explore the effects of internal model parameters and structure and also learning algorithms on the performance of ANN model. Novel techniques are proposed to improve the performance of the existing methods.

Chapters five to six presents post-modelling phase of ANN model development. Topics in these chapters, to date, are the widest gaps in the ANN research and applications, leaving the users and practitioners to only apply ANN without confidently knowing the prediction limits of their models and understanding what information / knowledge can be extracted from the model / hidden layers of ANN.

Chapter five provides theoretical analysis to construct the confidence intervals / prediction limits of the developed models. It also highlighted how good the ANN models can be used for extrapolation purposes. Chapter six addresses knowledge extraction of trained ANN model. Specifically, the focuses are to open the black box of the ANN model and to understand the behaviour of the developed model.

There are two approaches of knowledge extraction of ANN model discussed and proposed in this chapter. First approach is mainly based on sensitivity analysis of hidden layer and the other approach is a proposed technique to retrieve mathematical representation of the black box model. Using these two techniques, useful information from the data and better understanding of the process can be obtained.

In chapter seven, the conclusions from this study and the recommendations and future directions for research are presented. Appendices provide supplementary materials and information and related programme used in various chapters and sections.

To make the thesis clearer and more compact, while chapter two provides an overall literature review on current ANN research and applications and their research gaps and research need, other chapters also present a comprehensive literature review of related issues in each sections. The organisation of the thesis is described in Figure 1-1.

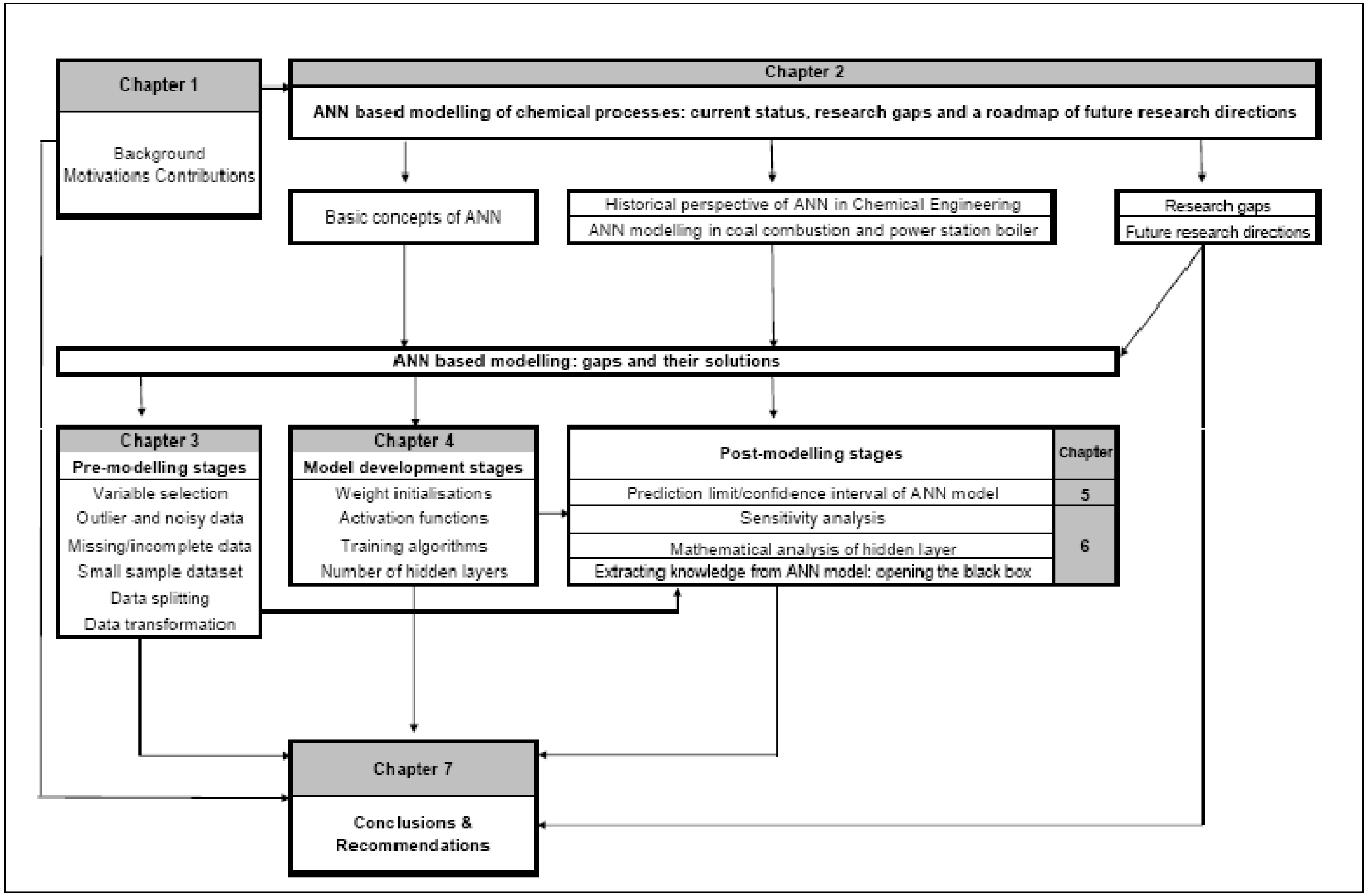

Figure 1.1 Schematic diagram of thesis layout.

Chapter 2
RESEARCH BACKGROUND AND LITERATURE REVIEW

This chapter is unable to be reproduced here due to a journal article (based on it) not yet in print.

Chapter 3

AN INTEGRATED DATA PREPARATION PROCEDURE FOR DEVELOPING ARTIFICIAL NEURAL NETWORK MODEL

This chapter is unable to be reproduced here due to a journal article (based on it) not yet in print.

CHAPTER 4
MODELLING STAGES OF ANN BASED MODELS

This chapter is unable to be reproduced here due to a journal article (based on it) not yet in print.

Chapter 5
Prediction Interval of the ANN Model

This chapter is unable to be reproduced here due to a journal article (based on it) not yet in print.

Chapter 6
Opening the Black Box of ANN Models

This chapter is unable to be reproduced here due to a journal article (based on it) not yet in print.

CHAPTER 7
CONCLUSIONS AND RECOMMENDATIONS

The aim of this research work was to address pre-modelling, training and post-modelling phases of the ANN predictive model since to date there is no systematic and integrated approach into ANN model design covering all the modelling issues above. This chapter summarised the main findings made during this research into three main sections, which are pre-modelling stages, training phases and post-modelling stages. The chapter concludes with recommendations for future work to further enhance the ANN modelling especially the post modelling part.

7.1 CONCLUSIONS

7.1.1 Pre-modelling phases

Several issues exist before the dataset is trained to develop an ANN process model. Being able to select appropriate input variables and examining and treating the data are essential to have a good predictive model. The main features of the proposed methods and findings in this part are:

- Proposed a grey superior analysis based approach in order to select optimum set of input variables. Using this approach, preliminary knowledge related to the degree of importance of each variable is also obtained. Compared with popular partial mutual information based technique, this proposed method was demonstrated to be superior.
- Treatment of outliers in the dataset is an important issue. Most ANN users tend to have outlier elimination to have 'a good quality of dataset'. However, deleting suspicious outliers can eliminate important information inside the datasets. Hence, based on the preliminary work from Liano, in this robustification of the network by modifying its error measure was carried out to reduce the influences of the outliers into the model performance.
- Current ANN technology is not able to deal with missing information in its training set and testing set. The most popular way to handle missing data is zero or mean substitution or deletion procedures. A new hybrid techniques BPCA-ANN was proposed to deal with missing input variables in both

training and testing sets and it is evident that the proposed approach perform very well in this regard.

- To be able to train the ANN successfully, the number of the training set should be as much as possible. However, in many cases, this does not always happen. Database of experiments or several operational parameters is small, on the other, ANN model is required to be built using this small sample dataset. In this work, MTD based ANN and ANN ensemble were applied and the results show that using one of this technique could improve the performance of the model in the case of only small training set available.
- The way the dataset was divided into the training and testing sets is believed to have a great impact on the predictive capability of the network. However, there is no guidance in how these datasets should be splitted. A proposed data division method based on Kennard-Stone algorithm using mahalanobis distance criterion proved to be superior to the standard Kennard Stone and SPXY algorithms
- Data transformation is also a crucial issue during pre-modelling stages of the ANN model. To determine, which data scaling technique should be used, however, is not an easy task. The finding from this work indicated however, that the choice of data transformation technique depends solely on the type of activation functions in the hidden and output nodes. Among other transformation techniques, the log based transformation is comparably recommended for all types of transfer functions

7.1.2 Training stages

- The choice of number of hidden layers and hidden nodes is paramount since it will influence not only the complexity of the model but also its predictive capability and training time. Many "rules-of-thumb' have been proposed to estimate optimum number hidden nodes. This research however indicated that these general guidelines mostly do not perform well to find optimum hidden nodes numbers and they can only be used as preliminary prediction of lower and upper limit of the trial-error-method. In this work, combination of this two approach, trial-error and empirical formula proved to be a quicker process rather than tedious trial-error method without any boundary. One hidden layer was also found to be sufficient as function approximator since multi hidden layer add complexity and do not show much better results.

- Weight initialisation will direct the network optimisation into its optimal solution. Random number generation is the most widely used weight and bias initialisation for ANN training. Among several random number generation approaches, technique based on Nguyen Widrow algorithm was found to be consistent to achieve convergence and its optimal solution
- Different configurations of transfer functions are possible to be used during ANN training. This research however, indicated that the use of tansig, logsig and radbas functions are recommended for hidden nodes since they provide sufficient nonlinearity to capture input-output variable relationship.
- MSE criterion is the most popular objective function for training and optimising the network. But since MSE is highly sensitive to the presence of the outlier, a bounded function based on logmse error measure was tested and the results show that this type of error measure tends to lead the training process having better convergence and performances.
- The last 'design variable' which must be selected during training phases is training algorithm. Several variants of back propagation algorithms exist and evolutionary based algorithms are currently popular. A comparative study of the different training algorithms were carried out and the results indicated that in terms of generalisation capability, the global optimisation like differential evolution is able to enhance the performance of traditional algorithms such as trainlm or trainbr. However, these hybrid algorithms have a limitation in terms of computational time where it was found to be nearly 12 fold. From this study, the use of trainlm is still recommended as it gives a faster convergence and comparably good results.

7.1.3 Post-modelling phases

Compared with other two phases, the post-modelling stages, namely prediction interval construction and knowledge extraction are rarely to discuss and present. This work presents a preliminary study of the post modelling phases of ANN modelling and lays a foundation for further research and applications. From this study, two conclusions can be drawn:

- Constructing prediction interval is necessary to examine how confidence the prediction is and how good the model when it is used for extrapolation. Since, linear regression based confidence interval does not take into account the distribution of the data, a Kernel density estimator based prediction interval was proposed. The results show that this technique is able to build a prediction interval in various degree of confidence.
- Research on knowledge extraction of the ANN model should be directed on opening the black box and obtaining mathematical representation of the overall ANN models. Our preliminary study show that it is possible to extract the simple models from the trained ANN model

7.2 RECOMMENDATIONS FOR FUTURE RESEARCH DIRECTIONS

Further research should be directed into the following areas to enhance the applicability of the proposed method:

- The use of tunable transfer function, that is, adaptive transfer function to comply with the dynamic nonlinearity of the input-output data
- The use of multiple error criteria to ensure that the optimisation process does not get stuck into local minima and also to guarantee that the optimum network is obtained using more stringent criterion
- Application of the proposed method in the last chapter in the grey box modelling. The obtained mathematical representation which is simpler and easy to understand surely will benefit the grey box application and process control in general.
- Further refinement of the approach to extract both qualitative and quantitative information from the trained ANN model. Possibly, more efforts should be directed to the analysing behaviour of the interconnected transfer functions toward network weights and biases to obtain a global function representing these interconnected functions.

References

1. Ramchadran, S. and R.R. Rhinehart, *Do Neural Networks Offer Something For You?* InTech, 1995. **42**(11): p. 59.
2. Su, H.T. and T.J. McAvoy, *Artificial Neural Networks for Nonlinear Process Identification and Control*, in *Nonlinear Process Control*, M.A. Henson and D.E. Seborg, Editors. 1997, Prentice Hall: Upper Saddle River, NJ.
3. Haykin, S., *Neural Network A Comprehensive Foundation*. 2nd ed. 1999, New Jersey: Prentice Hall. pp.1-255.
4. Kalogirou, S.A., *Artificial Intelligence for the Modeling and Control of Combustion Processes: A Review.* Progress in Energy and Combustion Science, 2003. **29**(6): p. 515-566.
5. Rodvold, D.M., D.G. McLeod, J.M. Brandt, P.B. Snow, and G.P. Murphy, *Introduction to Artificial Neural Networks for Physicians: Taking the Lid Off the Black Box.* The Prostate, 2001. **46**: p. 39-44.
6 Zhang, G., B. Eddy Patuwo, and M. Y. Hu, *Forecasting with artificial neural networks: The state of the art.* International Journal of Forecasting, 1998. **14**(1): p. 35-62.
7. Maier, H.R. and G.C. Dandy, *Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications.* Environmental Modelling & Software, 2000. **15**: p. 101-124.
8. Maier, H.R. and G.C. Dandy, *Neural network based modelling of environmental variables: a systematic approach.* Mathematical and Computer Modelling, 2001. **33**: p. 669-682.
9. Saptoro, A., H.B. Vuthaluru, and M.O. Tade. *A comparative study of prediction of elemental composition of coal using empirical modelling*. in *Proceedings of the IFAC Symposium on Advanced Control of Chemical Processes*. 2006. Gramado, Brazil.
10. Hoskins, J.C. and D.M. Himmelblau, *Artificial neural network models of knowledge representation in chemical engineering.* Computers & Chemical Engineering, 1988. **12**(9-10): p. 881-890.
11. Venkatasubramanian, V. and K. Chan, *A neural network methodology for process fault diagnosis.* AIChE Journal, 1989. **35**: p. 1993.
12. Bhagat, P., *AN introduction to neural nets.* Chemical ENgineering Progress, 1990. **55**.
13. Samdani, G., *Neural nets.* Chemical Engineering, 1990. **97**: p. 37.
14. Bhat, N.V. and T.J. McAvoy, *Use of neural nets for dynamic modelling and control of chemical processes.* Computers and Chemical Engineering, 1990. **14**: p. 573.
15. Hussain, M.A., *Review of the applications of neural networks in chemical process control - simulation and online implementation.* Artificial Intelligence in Engineering, 1999. **13**: p. 55-68.
16. Irwin, G., M. Brown, B. Hogg, and E. Swidenbank, *Neural network modelling of a 200 MW boiler system.* Control Theory and Applications, IEE Proceedings-, 1995. **142**(6): p. 529-536.
17. Lu, S. and B.W. Hogg, *Dynamic Nonlinear Modelling of Power Plant by Physical Principles and Neural Networks.* International Journal of Electrical Power & Energy Systems, 2000. **22**(1): p. 67-78.

18. Zhou, H., K. Cen, and J. Fan, *Multi-Objective Optimization of the Coal Combustion Performance with Artificial Neural Network and Genetic Algorithms.* International Journal of Energy Research, 2005. **29**(6): p. 499-510.
19. Pallares, J., I. Arauzo, and E. Teruel, *Development of an engineering system for unburned carbon prediction.* Fuel, 2009. **88**: p. 187-194.
20. Azlan Hussain, M., *Review of the applications of neural networks in chemical process control -- simulation and online implementation.* Artificial Intelligence in Engineering, 1999. **13**(1): p. 55-68.
21. Johannesmeyer , M.C., A. Singhal, and D.E. Seborg, *Pattern matching in historical data.* AIChE Journal, 2002. **48**(9): p. 2022-2028.
22. Yin, C., Z. Luo, J. Zhou, and K. Cen, *A novel non-linear programming-based coal blending technology for power plants.* Trans IChemE, 2000. **78**(Part A): p. 118-124.
23. Patel, S.U., B.J. Kumar, Y.P. Badhe, B.K. Sharma, S. Saha, S. Biswas, A. Chaudhury, S.S. Tambe, and B.D. Kulkarni, *Estimation of gross calorific value of coals using artificial neural networks.* Fuel, 2007. **86**: p. 334-344.
24. Chelgani, S.C., J.C. Hower, E. Jorjani, S. Mesroghli, and A.H. Bagherieh, *Prediction of coal grindability based on petrography, proximate and ultimate analysis using multiple regression and artificial neural network models.* Fuel Processing Technology, 2008. **89**(1): p. 13-20.
25. Teruel, E., C. Cortes, L. Ignacio Diez, and I. Arauzo, *Monitoring and prediction of fouling in coal-fired utility boilers using neural networks.* Chemical Engineering Science, 2005. **60**(18): p. 5035-5048.
26. Liu, Y.P., M.G. WU, and J.X. Qian, *Predicting ash fusion temperature based on its chemical composition using ACO-BP neural network.* Thermochimica Acta, 2007. **454**: p. 64-68.
27. Ozbayoglu, G., A.M. Ozbayoglu, and M.E. Ozbayoglu, *Estimation of Hardgrove grindability index of Turkish coals by neural networks.* International Journal of Mineral Processing, 2008. **85**: p. 93-100.
28. Bagherieh, A.H., J.C. Hower, A.R. Bagherieh, and E. Jorjani, *Studies of the relationship between petrography and grindability for Kentucky coals using artificial neural network.* International Journal of Coal Geology, 2008. **73**: p. 130-138.
29. Saptoro, A., H.M. Yao, M.O. Tade, and H.B. Vuthaluru, *Prediction of coal hydrogen content for combustion control in power utility using neural network approach.* Chemometrics and Intelligent Laboratory Systems, 2008. **94**: p. 149-159.
30. Yao, H.M., H.B. Vuthaluru, M.O. Tade, and D. Djukanovic, *Artificial neural network-based prediction of hydrogen content of coal in power station boilers.* Fuel, 2005. **84**(12-13): p. 1535-1542.
31. Zhu, Q., J.M. Jones, A. Williams, and K.M. Thomas, *The predictions of coal/char combustion rate using an artificial neural network approach.* Fuel, 1999. **78**(14): p. 1755-1762.
32. Nougues, J.M., Y.G. Pan, E. Velo, and L. Puigjaner, *Identification of a pilot scale fluidised-bed coal gasification unit by using neural networks.* Applied Thermal Engineering, 2000. **20**(15-16): p. 1561-1575.
33. Saptoro, A., H.B. Vuthaluru, and M.O. Tade. *Prediction and monitoring of unburnt carbon in fly ash in coal-fired power plant.* in *Proceeding of the*

International Conference on Modeling and Simulation. 2006. Kualalumpur, Malaysia.
34. Tronci, S., R. Baratti, and A. Servida, *Monitoring Pollutant Emissions in a 4.8 MW Power Plant through Neural Network.* Neurocomputing, 2002. **43**(1-4): p. 3-15.
35. Jensen, R.R., S. Karki, and H. Salehfar, *Artificial neural network-based estimation of mercury speciation in combustion flue gases.* Fuel Processing Technology, 2004. **85**(6-7): p. 451-462.
36. Calisto, H., N. Martins, and N. Afgan, *Diagnostic system for boilers and furnaces using CFD and neural networks.* Expert Systems with Applications, 2008. **35**(4): p. 17801787.
37. Yin, C., Z. Luo, M. Ni, and K. Cen, *Predicting coal ash fusion temperature with a back-propagation neural network model.* Fuel, 1998. **77**(15): p. 1777-1782.
38. Peisheng, L., X. Youhui, Y. Dunxi, and S. Xuexin, *Prediction of grindability with multivariable regression and neural network in Chinese coal.* Fuel, 2005. **84**(18): p. 2384-2388.
39. Jorjani, E., S.C. Chelgani, and S. Mesroghli, *Application of artificial neural networks to predict chemical desulfurization of Tabas coal.* Fuel, 2008. **87**: p. 2727-2734.
40. Guo, B., D. Li, C. Cheng, Z.-a. Lu, and Y. Shen, *Simulation of biomass gasification with a hybrid neural network model.* Bioresource Technology, 2001. **76**(2): p. 77-83.
41. Carsky, M. and D.K. Kuwornoo, *Neural network modelling of coal pyrolysis.* Fuel, 2001. **80**(7): p. 1021-1027.
42. Chong, A.Z.S., S.J. Wilcox, and J. Ward, *Prediction of gaseous emissions from a chain grate stoker boiler using neural networks of ARX structure.* Science, Measurement and Technology, IEE Proceedings-, 2001. **148**(3): p. 95-102.
43. Dong, C., B. Jin, Z. Zhong, and J. Lan, *Tests on co-firing of municipal solid waste and coal in a circulating fluidized bed.* Energy Conversion and Management, 2002. **43**: p. 2189-2199.
44. Romeo, L.M. and R. Gareta, *Hybrid system for fouling control in biomass boilers.* Engineering Applications of Artificial Intelligence, 2006. **19**: p. 915-925.
45. Makkonen, P. *Neural network for prediction of superheater fireside corrosion in CFB boilers*. in *Proceedings of the 6th International Conference on Fluidized Beds*. 1999. Wurzburg, Germany.
46. Allen, M.G., C.T. Butler, S.A. Johnson, E.Y. Lo, and F. Russo, *An imaging neural network combustion control system for utility boiler applications.* Combustion and Flame, 1993. **94**(1-2): p. 205-214.
47. Xu, L., Y. Yan, S. Cornwell, and R. Gerry, *Online fuel tracking by combining principal component analysis and neural network techniques.* IEEE Transactions on Instrumentation and Measurement, 2005. **54**: p. 1640-1645.
48. Hancke, G.P. and R. Malan, *A modal analysis technique for on-line particle size measurement of pneumatically conveyed pulverized coal.* IEEE Transactions on Instrumentation and Measurement, 1998. **47**(1): p. 114-122.
49. Chandok, J.S., I.N. Kar, and S. Tuli, *Estimation of furnace exit gas temperature (FEGT) using optimised radial basis and back-propagation*

neural networks. Energy Conversion and Management, 2008. **49**: p. 1989-1998.

50. Guglielmi, G., T. Parisini, and G. Rossi, *Keynote paper: Fault diagnosis and neural networks: A power plant application.* Control Engineering Practice, 1995. **3**(5): p. 601-620.
51. Prasad, G., E. Swidenbank, and B.W. Hogg, *A local model networks based multivariable long-range predictive control strategy for thermal power plant.* Automatica, 1998. **34**(10): p. 1185.
52. Hassoun, M.H., *Fundamentals of Artificial Neural Networks*. 1995, Cambridge, MA: MIT Press.
53. Quantrille, T.E. and Y.A. Liu, *Artificial Intellegence in Chemical Engineering*. 1991, California: Academic Press, Inc. pp. 442 - 483.
54. Hagan, M.T., H.B. Demuth, and O.D. Jesus, *An Introduction to the Use of Neural Networks in Control Systems.* International Journal Robust Nonlinear Control, 2002. **12**: p. 959-985.
55. Bowden, G.J., G.C. Dandy, and H.R. Maier, *Input determination for neural network models in water resources applications. Part 1 - background and methodology.* Journal of Hydrology, 2005. **301**: p. 75-92.
56. Bowden, G.J., H.R. Maier, and G.C. Dandy, *Optimal data division for neural network models in water resources applciations.* Water Resources Research, 2002. **38**(2): p. 2-1 - 2-11.
57. Bowden, G.J., G.C. Dandy, and H.G. Maier, *Data transformation for neural network models.* Journal of Hydroinformatics, 2003. **5**(4): p. 245-258.
58. Shanker, M., M.Y. Hu, and M.S. Hung, *Effect of data standardization on neural network training.* Omega, 1996. **24**(4): p. 385-397.
59. Shahin, M.A., H.R. Maier, and M.B. Jaksa, *Data division for developing neural networks applied to geotechnical engineering.* Journal of Computer in Civil Engineering, 2004. **18**(2): p. 105-114.
60. Sattler, K.U. and E. Schallehn. *A data preparation framework based on a multidatabase language*. in *Proceeding of International Symposium on Database Engineering and Applications*. 2001.
61. Yu, L., S. Wang, and K.K. Lai, *An integrated data preparation scheme for neural network data analysis.* IEEE Transactions on Knowledge and Data Engineering, 2006. **18**(2): p. 217-230.
62. Judd, S., *On the complexity of Loading Shallow Neural Networks.* Journal of Complexity, 1988. **4**: p. 177-192.
63. Geman, S., E. Bienenstock, and R. Doursat, *Neural Networks and the bias/variance dilemma.* Neural Networks, 1994. **4**(1): p. 123-32.
64. Fernando, T.M.K.G., H.R. Maier, and G.C. Dandy, *Selection of input variables for data driven models: An averag shifted histogram partial mutual information estimator approach.* Journal of Hydrology, 2008. ***In press***.
65. Kambhatla, N. and T.K. Leen, *Dimension reduction by local principal component analysis.* Neural Computation, 1997. **9**(7): p. 1493-1516.
66. Yang, J., D. Zhang, A.F. Frangi, and J. Yang, *TwoDimensional PCA: A new Approach to Appearance-Based Face Representation and Recognition.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004. **26**(1): p. 131-137.
67. Holcomb, T.R. and M. Morari, *PLS/neural networks.* Computers & Chemical Engineering, 1992. **16**(4): p. 393-411.

68. Kramer, M.A., *Nonlinear Principal Component Analysis Using Autoassociative Neural Networks.* AIChE Journal, 1991. **37**(2): p. 233-243.
69. Bowden, G.J., H.R. Maier, and G.C. Dandy, *Input determination for neural network models in water resources applications. Part 2. Case study: forecasting salinity in a river.* Journal of Hydrology, 2005. **301**(1-4): p. 93-107.
70. Sarle, W.S., *Neural Network FAQ.* URL:ftp://ftp.sas.com/pub/neural/FAQ.html, 2007.
71. Hasegawa, K., Y. Miyashita, and K. FUnatsu, *GA strategy for variable selection in QSAR studies:GA-based PLS analysis of calcium channel antagonists.* 1997.
72. Tu, C.J., L.Y. Chuang, J.Y. Chang, and C.H. Yang, *Feature Selection using PSO-SVM.* IAENG International Journal of Computer Science, 2007. **33**(1): p. 1-6.
73. Shen, Q., J.H. Jiang, C.X. Jiao, G.L. Shen, and R.Q. Yu, *Modified particle swarm optimization algortihm for variable selection in MLR and PLS modeling: QSAR studies of antagonism of angiotensin II antagonists.* European Journal of Pharmaceutical Sciences, 2004. **22**(2-3): p. 145-152.
74. Shen, Q., J.H. Jiang, J.C. Tao, G.L. Shen, and R.Q. Yu, *Modified Ant Colony Optimization Algorithm for Variable Selection in QSAR Modeling: QSAR Studies of Cyclooxygenase Inhibitors.* Journal of Chemical Information and Modeling, 1005. **45**(4): p. 1024-1029.
75. Sharma, A., *Seasonal to interannual rainfall probablistic forecasts for improved water supply management: Part 1-a strategy for system predictor identification.* Journal of Hydrology, 2000. **239**: p. 232-239.
76. Bowden, G.J., G.C. Dandy, and H.R. Maier, *Input determination for neural network models in water resources applications. Part 1--background and methodology.* Journal of Hydrology, 2005. **301**(1-4): p. 75-92.
77. May, R.J., H.R. Maier, G.C. Dandy, and T.M.K.G. Fernando, *Non-linear variable selection for artificial neural networks using partial mutual information.* Environmental Modelling & Software, 2008. **23**: p. 1312-1326.
78. Fernando, T.M.K.G., H.R. maier, and G.C. Dandy, *Selection of input variables for data driven models: An average shifted histogram partial mutual information estimator approach.* Journal of Hydrology, 2008. **In Press**.
79. Hamilton, J.D., *Analysis of Time Series Subject to Changes in Regime.* Journal of Econometrics, 1990. **45**: p. 39-70.
80. Bahattacharyya, S. and P.C. Pendharkar, *Inductive, evolutionary and neural techniques for discrimination: A comparative study.* Decision Sciences, 1998. **29**: p. 871-900.
81. Pendharkar, P.C., *A computational study on the performance of artificial neural networks under changing structural design and data distribution.* European Journal of Operational Research, 2002. **138**(1): p. 155-177.
82. Pendharkar, P.C. and J.A. Rodger, *An empirical study of impact of crossover operators on the performance of non-binary genetic algorithm based neural approaches for classification.* Computers and Operations Research, 2004. **31**(4): p. 481-498.
83. Altun, H., A. Bilgil, and B.C. Fildan, *Treatment of skewed muti-dimensional training data to facilitate the task of engineering neural models.* Expert Systems with Applications, 2007. **33**: p. 978-983.

84. SubbaNarasimha, P.N., B. Arinze, and M. Anandarajan, *The predictive accuracy of artificial neural networks and multiple regression in the case of skewed data: exploration of some issues.* Expert Systems with Applications, 2000. **19**(2): p. 117-123.
85. Kumar, U.A., *Comparison of neural networks and regression analysis: A new insight.* Expert Systems with Applications, 2005. **29**(2): p. 424-430.
86. Hampel, F.R., E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel, *Robust statistics, the approach based of influence functions*. 1986, New York: John Wiley & Sons.
87. Liano, K., *Robust error measure for supervised neural network learning with outliers.* IEEE Transactions on Neural Networks, 1996. **7**(1): p. 246-250.
88. Zhao, W., D. Chen, and S. Hu, *Detection of outlier and a robust BP algorithm against outlier.* Computers & Chemical Engineering, 2004. **28**(8): p. 1403-1408.
89. Pernia-Espinoza, A.V., J.B. Ordieres-Mere, and F.J. Martinez-de-Pison, *TAO-robust backpropagation learning algorithm.* Neural Networks, 2005. **18**: p. 191-204.
90. Bloch, G., P. Thomas, and D. Theilliol, *Accomodation to outliers in identification of non linear SISO systems with neural networks.* Neurocomputing, 1997. **14**: p. 85-99.
91. Roth, P.L., F.S. Switzer, and D.M. Switzer, *Missing data in multiple item scales: a Monte Carlo analysis of missing data techniques.* Organizational Research Methods, 1999. **2**(3): p. 211-232.
92. Verma, R. and J.C. Goodale, *Statistical power in operations management research.* Journal of Operations Management, 1995. **13**: p. 139-152.
93. Kim, J.O. and J. Curry, *The treatment of missing data in multi- variate analysis.* Sociological Methods and Research, 1977. **6**: p. 215-241.
94. Quinten, A. and W. Raaijmakers, *Effectiveness of different missing data treatments in surveys with Likert-type data: intro- ducing the relative mean substitution approach.* Educational and Psychological Measurement, 1999. **59**(5): p. 725-748.
95. Tsikriktsis, N., *A review of techniques for treating missing data in OM survey research.* Journal of Operations Management, 2005. **24**(1): p. 53-62.
96. Kros, J.F., M. Lin, and M.L. Brown, *Effects of the neural network s-Sigmoid function on KDD in the presence of imprecise data.* Computers & Operations Research, 2006. **33**: p. 3136-3149.
97. Brouwer, R.K. and W. Pedrycz, *Training a feed-forward network with incomplete data due to missing input variables.* Applied Soft Computing, 2003. **3**(1): p. 23-36.
98. Markey, M.K., G.D. Tourassi, M. Margolis, and D.M. DeLong, *Impact of missing data in evaluating artificial neural networks trained in complete data.* Computers in Biology and Medicine, 2006. **36**: p. 516-525.
99. Cybenko, G., *Approximation by superpositions of a sigmoidal function.* Mathematics of Control, Signal, and Systems, 1989. **2**: p. 303-314.
100. Hornik, K., M. Stinchcombe, and H. White, *Multilayer feedforward networks are universal approximators.* Neural Networks, 1989. **2**: p. 359-366.
101. Wray, J. and G.G.R. Green, *Neural networks, approximation theory, and finite precision computation.* Neural Networks, 1995. **8**: p. 31-37.

102. Lanouette, R., J. Thibault, and J.L. Valade, *Process modeling with neural networks using small experimental datasets.* Computers & Chemical Engineering, 1999. **23**(9): p. 1167-1176.
103. Huang, C. and C. Moraga, *A diffusion-neural-network for learning from small samples.* International Journal of Approximate Reasoning, 2004. **35**: p. 137-161.
104. Li, D.C., C.W. Yeh, T.I. Tsai, Y.H. Fang, and S.C. Hu, *Acquiring knowledge with limited experience.* Expert Systems, 2007. **24**(3): p. 162-170.
105. Li, D.C., C.S. Wu, T.I. Tsai, and Y.S. Lina, *Using mega-trend-diffusion and artificial samples in small data set learning for early flexible manufacturing system scheduling knowledge.* Computers & Operations Research, 2007. **34**: p. 966-982.
106. Li, D.C., H.C. Hsu, T.I. Tsai, T.J. Lu, and S.C. Hu, *A new method to help diagnose cancers for small size.* Expert Systems with Applications, 2007. **33**: p. 420-424.
107. Kim, Y.S., *Comparison of the decision tree, artificial neural network, and linear regression methods based on the number and types of independent variables and sample size.* Expert Systems with Applications, 2008. **34**: p. 1227-1234.
108. Li, D.C., L.S. Chen, and Y.S. Lin, *Using Functional Virtual Population as assistance to learn scheduling knowledge in dynamic manufacturing environments.* International Journal of Production Research, 2003. **41**(17): p. 4011-4024.
109. Li, D.C., C.S. Wu, T.I. Tsai, and F.M. Chang, *Using mega-fuzzification and data trend estimation in small data set learning for early FMS scheduling knowledge.* Computers & Operations Research, 2006. **33**(6): p. 1857-1869.
110. Opitz, D.W. and J.W. Shalik, *Actively searching for an Effective Neural-Network Ensemble.* Connection Science, 1996. **8**(3-4): p. 1-17.
111. Opitz, D.W. and J.W. Shavlik, *A Genetic Algorithm Approach for Creating Neural_Network Ensemble*, in *Combining Artificial Neural Nets*, A. Sharkey, Editor. 1999, Springer-Verlag: London. p. 79-97.
112. Zhang, J., B. Martin, A.J. Morris, and C. Kiparissides, *Inferential Estimation of Polymer Quality Using Stacked Neural Networks.* Computers & Chemical Engineering, 1997. **21**: p. S1025-S1030.
113. Eikens, B. and M.N. Karim, *Process identification with multiple neural network models.* International Journal of Control, 1999. **72**(7/8): p. 576-590.
114. Zhang, J., *Improved on-line process fault diagnostic through information fusion in multiple neural networks.* Computers & Chemical Engineering, 2006. **30**: p. 558-571.
115. Wolpert, D.H., *Stacked Generalization.* Neural Networks, 1992. **5**: p. 241-259.
116. Flood, I. and N. Kartam, *Neural networks in civil engineering. I:Principles and understanding.* Journal of Computing in Civil Engineering, 1994. **8**(2): p. 131-148.
117. Galvao, R.K.H., M.C.U. Araujo, G.E. Jose, M.J.C. Pontes, E.C. Silva, and T.C.B. Saldanha, *A method for calibration and validation subset partitioning.* Talanta, 2005. **67**: p. 736-740.
118. Daszykowski, M., K. Kaczmarek, Y.V. Heyden, and B. Walczak, *Robust statistics in data analysis - a review Basic concepts.* Chemometrics and Intelligent Laboratory Systems, 2007. **85**: p. 203-219.

119. Kim, D., *Normalization methods for input and output vectors in backpropagation neural networks.* International Journal of Computer Mathematics, 1999. **71**: p. 161-171.
120. Shi, J.J., *Reducing prediction errors by transforming input data for neural networks.* Journal of Computers in Civil Engineering, 2000. **14**(2): p. 109-116.
121. Demuth, H., M. Beale, and M. Hagan, *Neural Network Toolbox 6 User's Guide*. 2008, MA: The MathWorks, Inc.
122. Faraway, J. and C. Chatfield, *Time series forecasting with neural networks: a comparative study using the airline data.* Applied Statistics, 1998. **47**(2): p. 231-250.
123. Maier, H.R. and G.C. Dandy, *The effects of internal parameters and geometry on the performance of back-propagation neural networks: an empirical study.* Environtmental Modelling & Software, 1998. **13**: p. 193-209.
124. Lawrence, S., C.L. Giles, and A.C. Tsoi, *What Size Neural Networks Gives Optimal Generalization? Convergence Properties of Backpropagation*, in *In Technical Report UMIACS-TR-96-22 and CS-TR-3617.* 1996, Institute for Advanced Computer Studies, Univ. of Maryland.
125. Caruana, R., S. Lawrence, and C.L. Giles, *Overfitting in Neural Networks: Backpropagation, Conjugate Gradient, and Early Stopping*, in *Advances in Neural Information Processing Systems*, T.K. Leen, T.G. Dietterich, and V. Tresp, Editors. 2001, MIT Press. p. 402-408.
126. Bartlett, P.L., *For valid generalization the size of the weights is more important than the size of the network*, in *Advances in Neural Information Processing Systems*, M.C. Mozer, M.I. Jordan, and T. Petsche, Editors. 1997, MIT Press. p. 134-140.
127. Bebis, G. and M. Georgiopoulos, *Feed-forward neural networks: why networks size is so important.* IEEE Potentials October, 1994: p. 27-31.
128. de Villiers, J. and E. Barnard, *Backpropagation neural nets with one and two hidden layers.* Neural Networks, IEEE Transactions on, 1993. **4**(1): p. 136-141.
129. Blum, A., *Neural Networks in C++*. 1992, New York: John Wiley & Sons.
130. Swingler, K., *Applying Neural Networks: A Practical Guide*. 1996, London: Academic Press.
131. Berry, M.J.A. and G. Linoff, *Data Mining Techniques*. 1997, New York: John Wiley & Sons.
132. Boger, Z. and H. Guterman. *Knowledge extraction from artificial neural networks models*. in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*. 1997. Orlando, FL.
133. Maren, A., C. Harston, and R. Pap, *Handbook of Neural Computing Applications*. 1991, San Diego, CA: Academic Press.
134. Hecht-Nielsen, R. *Kolmogorovs mapping neural networks existence theorem.* in *Proceesings of the First International Joint Conference on Neural Networks*. 1987. San Diego, CA.
135. Rogers, L.L. and F.U. Dowla, *Optimization of the groundwater remediation using artificial neural networks with parallel solute transport modeling.* Water Resources Research, 1994. **30**(2): p. 457-481.

136. Hsiao, T.C.R., C.W. Lin, and H.K. Chiang, *Partial least squares algorithm for weights initialization of backpropagation network.* Neurocomputing, 2004. **50**: p. 237-247.
137. Sit, C.W., *Application of Artificial Neural Network-Genetic Algorithm In Inferential Estimation and Control of A Distillation Column*, in *Faculty of Chemical and Natural Resources Engineering*. 2005, Universiti Teknologi Malaysia.
138. Lahiri, S.K. and K.C. Ghanta, *Development of an artificial neural network correlation for prediction of hold-up of slurry transport in pipelines.* Chemical Engineering Science, 2008. **63**: p. 1497-1509.
139. Puig, M.C., X.M. Berbel, M.A.A. Lomillo, F.J.M. Pascual, and M. del Valle, *EIS multianalyte sensing with an automated SIA system-An electronic tongue employing the impedimetric signal.* Talanta, 2007. **72**: p. 774-779.
140. Fildes, R., *The evaluation of extrapolative forecasting methods.* International Journal of Forecasting, 1992. **8**: p. 81-98.
141. Amstrong, J.S. and F. Collopy, *Error measure for generalizing about forecating methods: Empirical Comparisons.* International Journal of Forecasting, 1992. **8**: p. 69-80.
142. Chatfield, C., *Apples.oranges and mean square error.* International Journal of Forecasting, 1988. **4**: p. 515-518.
143. Ilonen, J., J.K. Kamarainen, and J. Lampinen, *Differential Evolution Training Algorithm for Feed-Forward Neural Networks.* Neural Processing Letter, 2003. **17**: p. 93-105.
144. Subudhi, B. and D. Jena, *Differential Evolution and Levenberg-Marquardt Trained Neural Networks Scheme for Nonlinear System Identification.* Neural Processing Letter, 2008. **27**: p. 285-296.
145. Tickle, A.B., R. Andrews, M. Golea, and J. DIederich, *The truth will come to light: Directions and challeges in extracting knowledge embedded within trained artificial neural network.* IEEE Transactions on Neural Networks, 1998. **9**(6): p. 1057-1068.
146. Benitez, J.M., J.L. Castro, and I. Requena, *Are artificial neural networks blac boxes?* IEEE Transactions on Neural Networks, 1997. **8**(5): p. 1156-1164.
147. Castro, J.L., C.J. Mantas, and J.M. Benitez, *Interpretation of artificial neural networks by means of fuzzy rules.* IEEE Transactions on Neural Networks, 2002. **13**(1): p. 101-116.
148. Lozowski, A., T.J. Cholewo, and J.M. Zurada. *Crisp rule extraction from perceptron network classifiers*. in *Proceedings of the IEEE International Conference on Neural Networks*. 1996. Washinton, D.C.
149. Sudheer, K.P., *Knowledge Extraction from Trained Neural Network River Flow Models.* Journal of Hydrologic Engineering, 2005. **10**(4): p. 264-269.
150. Jaynes, E.T., *Information Theory and Statistical Mechanics.* Physical Review, 1957. **106**(4): p. 620-630.
151. Chernick, M.R., *Boostrap Methods: A Practitioner's Guide*. 1999, New York: John Wiley & Sons.
152. Davies, L. and U. Gather, *The identification of mutiple outliers.* Journal of the American Statistical Association, 1993. **88**(423): p. 782-792.
153. Pearson, R.K., *Exploring process data.* Journal of Process Control, 2001. **11**(2): p. 179-194.
154. Deng, J.L., *The introduction of grey system.* The Journal of Grey System, 1989. **1**(1): p. 1-24.

155. Yin, X. and W. Yu, *The virtual manufacturing model of the worsted yarn based on artificial neural networks and grey theory.* Appllied Mathematics and Computation, 2007. **185**: p. 322.
156. Yin, X.G. and W. Yu, *Selection and evaluation of input parameters of neural networks using grey superior analysis.* Textile Research Journal, 2007. **77**: p. 375-386.
157. Li, G.D., D. Yamaguchi, and M. Nagai, *A grey-based decision-making approach to the supplier selection problem.* Mathematical and Computer Modelling, 2007. **46**: p. 573-581.
158. Hair, J.F., R.E. Anderson, R.L. Tatham, and W.C. Black, *Multivariate Data Analysis*. 1998, Upper Saddle River, New Jersey: Prentice-Hall International, Inc.
159. De Maesschalck, R., D. Jouan-Rimbaud, and D.L. Massart, *The Mahalanobis distance.* Chemometrics and Intelligent Laboratory Systems, 2000. **50**(1): p. 1-18.
160. Rousseeuw, P.J., *Least median of square regression.* Journal of the American Statistical Association, 1984(79).
161. Croux, C. and G. Haesboeck, *Influence function and efficiency of the minimum covariance determinant scatter matrix estimator.* Journal of Multivariate Analysis, 1999. **71**: p. 161-190.
162. Rousseeuw, P.J. and K. van Driessen, *A fast algorithm for the minimum covariance determinant estimator.* Technometrics, 1999. **41**(3): p. 212-223.
163. Bishop, C.M. *Variational principal components*. in *Proceeding of the IEE Conference Publication on Artificial Neural Networks*. 1999.
164. Oba, S., M. Sato, I. Takemasa, M. Monden, K. Matsubara, and S. Ishii, *A Bayesian missing value estimation method for gene expression profile data.* Bioinformatics, 2003. **19**(16): p. 2088-2096.
165. Tipping, M.E. and C.M. Bishop, *Mixtures of probabilistic principal component analysers.* Neural Computation, 1999. **11**: p. 443-482.
166. *Dynamic models for fossil fueled steam units in power system studies.* Power Systems, IEEE Transactions on, 1991. **6**(2): p. 753-761.
167. Attias, H. *Inferring parameters and structure of latent variable models by variational bayes*. in *Proceeding of the Uncertainty in Artificial Intelligence*. 1999.
168. Sato, M., *On-line model selection based on the variational Bayes.* Neural Computation, 2001. **13**: p. 1649-1681.
169. Funahashi, K., *On the approximate realization of continuous mappings ny neural networks.* Neural Networks, 1989. **2**: p. 183-192.
170. Zhang, J., *Batch-to-batch optimal control of batch polymerisation process based on stacked neural networks models.* Computers & Chemical Engineering, 2008. **63**: p. 1273-1281.
171. Lee, H., S. Hong, and E. Kim, *Neural network ensemble with probabilistic fusion and its application to gait recognition.* Neurocomputing, 2008. ***In press***.
172. Master, T., *Practical Neural Network Recipes in C++*. 1993, San Diego: Academic Press.
173. Friedman, J., *Multivariate adaptive regression splines (with discussion).* Annal of Statistics, 1991. **9**(1): p. 1-141.
174. Roffel, B. and B. Betlem, *Process dynamics and control: modeling for control and prediction*. 2006, Chichester: John Wiley & Sons, Ltd.

175. Speight, J.G., *Handbook of Coal Analysis*. 2005, New Jersey: John Wiley & Sons, Inc.
176. Saptoro, A., H.B. Vuthaluru, and M.O. Tade. *Design of feed-forward neural networks architecture for coal elemental composition prediction*. in *Proceedings of the International Conference on Modeling and Simulation*. 2006. Kualalumpur, Malaysia.
177. Saptoro, A., H.B. Vuthaluru, and M.O. Tade. *Prediction of coal properties using neural networks: a novel approach based on small training dataset*. in *Proceedings of the CHEMECA*. 2008. Newcastle, Australia.
178. Yin, C., L. Rosendahl, and Z. Luo, *Methods to improve prediction performance of ANN models.* Simulation Modelling Practice and Theory, 2003. **11**: p. 211-222.
179. Sheng, C.D. and J.L.T. Azevedo, *Estimating the higher heating value of biomass fuels from basic analysis data.* Bimass and Bioenergy, 2005. **28**: p. 499-507.
180. Parikh, J., S.A. Channiwala, and G.K. Ghosal, *A correlation for calculating HHV from proximate analysis of solid fuels.* Fuel, 2005. **84**(5): p. 487-494.
181. Mazumdar, B.K., *Theoretical oxygen requirement for coal combustion: relationship with its calorific value.* Fuel, 2000. **79**(11): p. 1413-1419.
182. Singh, K.P. and M.C. Kakati, *Comprehensive models for computation of the specific energy of coals.* The Chemical Engineering Journal and the Biochemical Engineering Journal, 1995. **57**(3): p. 295-298.
183. Nguyen, D. and B. Widrow. *Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights*. in *Proceeding of the International Joint Conference on Neural Neworks*. 1990. San Diego, CA.
184. Yam, Y.F. and T.W.S. Chow, *A new method in determining initial weights of feed forward neural networks.* Neurocomputing, 1997. **16**: p. 23-32.
185. Yam, Y.F., C.T. Leung, P.K.S. Tam, and W.C. Siu, *An independent component analysis based weight initialization method for multilayer perceptrons.* Neurocomputing, 2002. **48**: p. 807-818.
186. Maier, H.R., *Empirical comparisons of various methods for training feed-forward neural networks for salinity forecasting.* Water Resources Research, 1999. **35**(8): p. 2591-2596.
187. Plumb, A.P., R.C. Rowe, P. York, and M. Brown, *Optimisation of the predictive ability of artificial neural network (ANN) models: a comparison of three ANN programs and four classes of training algorithm.* European Journal of Pharmaceutical Sciences, 2005. **25**: p. 395-405.
188. Hagan, M.T., H.B. Demuth, and M.H. Beale, *Neural Network Design*. 1996, Boston: PWS.
189. MacKay, D.J.C., *Bayesian interpolation.* Neural Computation, 1992. **4**: p. 415-447.
190. Storn, R. and K. Price, *DIfferential evolution - a simple and efficient heuristic for global optimization over continuous spaces*. 1995, International Computer Science Institute: Berkeley, CA.
191. Masters, T. and W. Land. *A new training algorithm for the general regression neural network.* in *Proceedings of the IEEE Conference on Systems, Man and Cybernetics, Computational Cybernetics and Simulation, 1990-1994*. 1997.
192. Hecht-Nielsen, R., *Neurocomputing*. 1990, Reading, MA: Addison-Wesley.

193. Chryssolouris, G., M. Lee, and A. Ramsey, *Confidence Interval Prediction for Neural Network Models.* IEEE Transactions on Neural Networks, 1996. **7**(1): p. 229-232.
194. Rivals, I. and L. Personnaz, *Construction of confidence intervals for neural networks based on least squares estimation.* Neural Networks, 2000. **13**: p. 463-484.
195. da Silva, A.P.A. and L.S. Moulin, *Confidence Intervals for Neural Network Based Short-Term Load Forecasting.* IEEE Transactions on Power Systems, 2000. **15**(4): p. 1191-1196.
196. Hwang, J.T.G. and A.A. Ding, *Prediction Intervals for Artificial Neural Networks.* Journal of the American Statistical Association, 1997. **92**(438): p. 748-757.
197. Baffi, G., E.B. Martin, and J. Morris, *Prediction intervals for non-linear projection to latent structures regression models.* Chemometrics and Intelligent Laboratory Systems, 2000. **61**: p. 151-165.
198. Jiahui, L. and L. Chen. *The Modified PNN Prediction Interval for Spacecraft Data*. in *Proceedings of the IEEE Congress on Image and SIgnal Processing*. 2008.
199. Shao, R., E.B. Martin, J. Zhang, and A.J. Morris, *Confidence bounds for neural network representations.* Computers & Chemical Engineering, 1997. **21 (suppl.)**: p. 748-757.
200. Safavi, A.A., J. Chen, and J.A. Romagnoli, *Wavelet-Based Density Estimation and Application to Process Monitoring.* AIChE Journal, 1997. **43**(5): p. 1227-1241.
201. Garson, G.D., *Interpreting neural network connection weights.* Artificial Intelligence Expert, 1991. **6**: p. 47-51.
202. Goh, A.T.C., *Back-propagation neural networks for modeling complex systems.* Artificial Intelligence in Engineering, 1995. **9**(3): p. 143-151.
203. Gevrey, M., I. Dimopoulos, and S. Lek, *Review and comparison od methods to study the contribution of variables in artificial neural network models.* Ecological Modelling, 2003. **160**: p. 249-264.
204. Tchaban, T., J.P. Griffin, and M.J. Taylor, *A comparison between single and combined backpropagation neural networks in the prediction of turnover.* Engineering Applications of Artificial Intelligence, 1998. **11**(1): p. 41-47.
205. Kingston, G.B., H.R. maier, and M.F. Lambert, *A probabilistic method for assisting knowledge extraction from artificial neural networks used for hydrological prediction.* Mathematical and Computer Modelling, 2006. **44**: p. 499-512.
206. Gaweda, A.E., R. Setiono, and J.M. Zurada. *Rule extraction from Feedforward Neural Network for Function Approximation*. in *Proceedings of the 5th National Conference on Neural Networks and Soft Computing*. 2000.