

NEURAL NETWORK-BASED METAMODELLING  
APPROACH FOR ESTIMATION OF AIR POLLUTANT  
PROFILES

HERMAN WAHID

DOCTOR OF PHILOSOPHY

UNIVERSITY OF TECHNOLOGY, SYDNEY

2013

# **NEURAL NETWORK-BASED METAMODELLING APPROACH FOR ESTIMATION OF AIR POLLUTANT PROFILES**

By  
**HERMAN WAHID**

A thesis submitted in fulfilment of  
the requirements for the degree of  
Doctor of Philosophy



Faculty of Engineering  
University of Technology, Sydney  
February 2013

## **CERTIFICATE OF AUTHORSHIP/ ORIGINALITY**

I certify that the work in this thesis has not previously been submitted for a similar degree nor has it been submitted as part of requirements for any other degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are referenced in the thesis.

February 2013

Production Note:  
Signature removed prior to publication.  
.....

Herman Wahid

*This thesis is especially dedicated to my dearest father, mother, wife and family for their love, blessing and encouragement ...*

## **ACKNOWLEDGMENT**

First of all, I am greatly indebted to God in His mercy and blessing for making this project successful. This research project was carried out in the Faculty of Engineering, University of Technology Sydney between August 2009 and September 2012. I gratefully acknowledge the financial support for this research by the Malaysia Government and Universiti Teknologi Malaysia (UTM) Scholarship.

I would like to express my deepest gratitude and appreciation to my honourable supervisor, Associate Professor Dr. Quang Ha for his continuous guidance, encouragement, committed support, timely reviews and invaluable advice throughout the duration of this research project.

I would like to thank two atmospheric scientists, Dr. Hiep Duc from Office of Environment and Heritage New South Wales, and Dr. Merched Azzi from CSIRO Energy Technology New South Wales, for their assistance with air quality modelling related theories and their constructive comments with respect to this research project. Also thanks to Dr. Hiep Duc and his division team members for guiding and providing me with the valuable knowledge of air quality measurement methodologies for the collection of the research dataset.

My gratitude is also extended to Dr. Guenter A Plum for editing my thesis draft. Dr. Plum is an Australian academic thesis editor and proofreader, and has more than ten years experience in academic thesis editing, academic thesis proofreading and academic dissertation editing and proofing with clients in Australia, Asia, and around the world. Dr. Plum has worked in universities in Australia and Hong Kong as well as in the computer industry. Dr. Plum's helpful suggestions and attention to detail have been integral to a timely completion of my thesis.

I would also like to extend my thanks and best wishes to all students in my research group, especially to Vahid Vakiloroya and Raja M Taufika Raja Ismail for some

technical supports during my study. Their presence and companionship made my study experience most comfortable and enjoyable.

Last but not least, I would like to express my sincere appreciation and gratitude to my parents, Wahid Moktar and Mazenah Sidik, for their patience, encouragement and support throughout the years of my study. I would like to dedicate this research to my beloved wife, Nazariah Jadon. Her constant empathy and consistent moral support have eased the tough times during my research.

## ABSTRACT

The air quality system is a system characterised by non-linear, complex relationships. Among existing air pollutants, the ozone ( $O_3$ ), known as a secondary pollutant gas, involves the most complex chemical reactions in its formation, whereby a number of factors can affect its concentration level. To assess the ozone concentration in a region, a measurement method can be implemented, albeit only at certain points in the region. Thus, a more complicated task is to define the spatial distribution of the ozone level across the region, in which the deterministic air quality model is often used by the authority. Nevertheless, simulation by using a deterministic model typically needs high computational requirements due to the nonlinear nature of chemical reactions involved in the model formulation, which is also subject to uncertainties. In the context of ozone as an air pollutant, the determination of the background ozone level (BOL), independent from human activities, is also important as it could represent one of reliable references to human health risk assessment. The concept of BOL may be easily understood, but practically, it is hard to distinguish between natural and anthropogenic effects. Apart from existing approaches to the BOL determination, a new quantisation method is presented in this work, by evaluating the relationship of ozone versus nitric oxide ( $O_3$ -NO) to estimate the BOL value, mainly by using night-time and early morning measurement data collected at the monitoring stations.

In this thesis, to deal with the challenging problem of air pollutant profile estimation, a metamodel approach is suggested to adequately approximate intrinsically non-linear and complex input-output relationships with significantly less computation. The intrinsic characteristics of the underlying physics are not assumed to be known, while the system's input and output behaviours remain essential. A considerable number of metamodels approach have been proposed in the literature, e.g. splines, neural networks, kriging and support vector machine. Here, the radial basis function neural network (RBFNN) is concerned as it is known to offer good estimation performance on accuracy, robustness, versatility, sample size, efficiency, and

simplicity as compared to other stochastic approaches. The development requirements are that the proposed metamodels should be capable of estimating the ozone profiles and its background level temporally and spatially with reasonably good accuracies, subject to satisfying some statistical criteria.

Academic contributions of this thesis include in a number of performance enhancements of the RBFNN algorithms. Generally, three difficulties involved in the network training, selection of radial basis centres, selection of the basis function variance (i.e. spread parameter), and training of network weights. The selection of those parameters is very crucial, as they directly affect the number of hidden neurons used and also the network overall performance. In this research, some improvements of the typical RBFNN algorithm (i.e. orthogonal least squares) are achieved. First, an adaptively-tuned spread parameter and a pruning algorithm to optimise the network's size are proposed. Next, a new approach for training the RBFNN is presented, which involves the forward selection method for selecting the radial basis centres. Also, a method for training the network output weights is developed, including some suggestions for estimation of the best possible values of the network parameters by considering the cross-validation approach. For applications, results show that the combination of the proposed paradigm could offer a sub-optimal solution of metamodeling development in the generic sense (by avoiding the iteration process) for a faster computation, which is essential in air pollutant profile estimation.



## PUBLICATIONS

### Journal Articles

1. Hiep Duc, Merched Azzi, Herman Wahid, Q.P. Ha, “Background ozone level in the Sydney basin: Assessment and trend analysis”, *Journal of Climatology*, in Press, doi: 10.1002/joc.3595.
2. H. Wahid, Q.P. Ha, H. Duc, “New sampling scheme for neural network-based metamodelling with application to air pollutant estimation,” *Gerontechnology*, vol. 11(2), 2012, pp. 336, doi:10.4017/gt.2012.11.02.325.00.
3. H. Wahid, Q.P. Ha, H. Duc, M. Azzi, “Meta-modelling approach for estimating the spatial distribution of air pollutant levels”, *Journal of Applied Soft Computing* (revised and re-submitted).
4. Q.P. Ha, H. Wahid, H. Duc, M. Azzi, “Radial basis function neural network-based metamodelling for ozone spatial estimation,” *IEEE Transactions on Neural Network and Learning Systems* (submitted).

### Conference Proceedings

1. H. Wahid, Q.P. Ha, H. Duc, “New sampling scheme for neural network-based metamodelling with application to air pollutant estimation,” *Proc. 8<sup>th</sup> World Conference of Gerontechnology & 29<sup>th</sup> International Symposium for Automation and Robotics in Construction (ISG-ISARC 2012)*, Eindhoven, Netherlands, June 26-29, 2012.
2. Herman Wahid, Q.P. Ha, Hiep Duc, Merched Azzi, “Estimation of background ozone temporal profiles using neural networks,” *Proc. 3<sup>rd</sup> IEEE Int. Conf. on Intelligence Computing and Intelligent Systems (ICIS 2011)*, Guangzhou, China, 18-20 Nov 2011, pp. 292-297.
3. Herman Wahid, Q.P. Ha, Hiep Nguyen Duc, “Computational intelligence estimation of natural background ozone level and its distribution for air quality modelling and emission control,” *Pro. 28<sup>th</sup> International Symposium on*

*Automation and Robotics in Construction (ISARC 2011)*, Seoul, Korea, 29 Jun-2 Jul 2011, pp. 551-557.

4. Herman Wahid, Quang P. Ha, and Hiep Nguyen Duc, "A Metamodel for Background Ozone Level Using Radial Basis Function Neural Networks," *Proc. 11<sup>th</sup> International Symposium on Control, Automation, Robotics and Vision (ICARCV 2010)*, Singapore, 7-10 Dec 2010, pp. 958-963.
5. H. Wahid, Q. P. Ha, and H. Duc, "Adaptive Neural Network Metamodel for Short-term Prediction of Natural Ozone Level in an Urban Area," *Proc. Int. Conf. on Computing and Communication Technologies (2010 IEEE-RIVF)*, Hanoi, Vietnam, 1-4 Nov 2010, pp. 250-253.

### **Related publications**

1. Herman Wahid, Hiep Nguyen Duc, and Quang P. Ha, "Radial Basis Function Neural Network Metamodelling for 2-D Resistivity Mapping," *Proc. 27<sup>th</sup> International Symposium on Automation and Robotics in Construction (ISARC 2010)*, Bratislava, Slovakia, 25-27 Jun 2010, pp. 364-373.
2. H. Wahid, Q.P. Ha, and M.S. Mohamed Ali, "Optimally-Tuned Cascaded PID Control using Radial Basis Function Neural Network Metamodeling," *Proc. of the 3<sup>rd</sup> International Workshop on Artificial Intelligence in Science and Technology (AISAT'09)*, Hobart, Australia, 23-24 November 2009, paper S01.2.

## TABLE OF CONTENTS

CERTIFICATE OF AUTHORSHIP/ ORIGINALITY.....	i
DEDICATION.....	ii
ACKNOWLEDGMENT.....	iii
ABSTRACT.....	v
PUBLICATIONS.....	vii
TABLE OF CONTENTS.....	ix
LIST OF TABLES.....	xvi
LIST OF FIGURES.....	xvii
LIST OF ABBREVIATIONS.....	xxiii
 Chapter 1 INTRODUCTION .....	 1
1.1. General introduction .....	1
1.2. Problem statement .....	3
1.3. Research objectives .....	6
1.4. Significance of the work and contributions.....	7
1.5. Thesis structure .....	10
 Chapter 2 LITERATURE REVIEW .....	 13
2.1. Introduction .....	13
2.2. Overview of air quality assessment .....	14
2.2.1. Three approaches in air quality assessment .....	14
2.2.1.1. Air quality monitoring .....	14
2.2.1.2. Emission inventory and assessment .....	16
2.2.1.3. Air quality modelling.....	18
2.2.2. Air quality standards and air quality index .....	19
2.2.3. Air quality model review .....	21
2.2.3.1. Deterministic model .....	21
2.2.3.2. Statistical / Empirical model .....	26

2.3.	Review of ozone and background ozone level.....	28
2.3.1.	Ozone and its determination.....	28
2.3.2.	Background ozone level and its determination.....	30
2.4.	Review of metamodeling techniques.....	33
2.4.1.	Types of metamodel .....	33
2.4.2.	Trade-offs between metamodels.....	36
2.4.3.	Sampling techniques.....	38
2.5.	Review of radial basis function neural networks .....	40
2.5.1.	RBFNN general function and its architecture .....	40
2.5.2.	Type of basis functions.....	42
2.5.3.	Learning strategies .....	44
2.5.3.1.	General review .....	44
2.5.3.2.	Fully-supervised training.....	47
2.5.3.3.	Two-stage training .....	48
2.5.4.	Regularised and generalised RBFNN.....	50
2.6.	Chapter conclusion .....	52
Chapter 3	BUILDING A NEURAL NETWORK-BASED METAMODEL .....	54
3.1.	Introduction .....	54
3.2.	The metamodeling process .....	54
3.2.1.	Data preparation and sampling .....	55
3.2.1.1.	Variable importance .....	55
3.2.1.2.	Data sampling .....	57
3.2.1.3.	Data set classification .....	57
3.2.2.	Model training .....	58
3.2.3.	Model testing (validation).....	59
3.3.	Metamodel management .....	60
3.4.	A new sampling scheme for a NN based metamodel .....	62
3.4.1.	Introduction .....	62
3.4.2.	Methodology .....	63
3.4.3.	Test function.....	65

3.4.4.	Numerical analysis .....	66
3.4.5.	Discussion .....	72
3.5.	Chapter conclusion .....	73
Chapter 4	ORTHOGONAL LEAST SQUARES ALGORITHM .....	74
4.1.	Introduction .....	74
4.2.	Orthogonal Least Squares (OLS) learning algorithm .....	75
4.3.	Some improvements of OLS algorithm .....	79
4.3.1.	Adaptively-tuned spread parameter .....	79
4.3.1.1.	Introduction .....	79
4.3.1.2.	Methodology.....	79
4.3.2.	A pruning algorithm for RBFNN.....	84
4.3.2.1.	Introduction .....	84
4.3.2.2.	Methodology.....	84
4.3.3.	Numerical analysis .....	86
4.3.3.1.	The OLS performance.....	86
4.3.3.2.	The OLS with adaptive tuned spread parameter .....	89
4.3.3.3.	The OLS with adaptive spread and pruning algorithm .....	93
4.4.	Chapter conclusion .....	94
Chapter 5	PARAMETER DETERMINATION FOR RBFNN LEARNING ALGORITHM	
	95	
5.1.	Introduction .....	95
5.2.	A forward selection method for centres determination.....	96
5.2.1.	Using regularised least squares to form a cost function .....	96
5.2.2.	Solution for the regularised and weighted least squares.....	97
5.2.3.	Formulation of the error function .....	98
5.2.4.	Forward selection with regularised and weighted least squares .....	99
5.3.	Training the network output weights.....	104
5.4.	Selection of $H$ , $\lambda$ and $\sigma$ parameters .....	105
5.4.1.	Estimation of the $H$ parameter .....	105
5.4.2.	Optimal selection of $\lambda$ using cross-validation.....	105

5.4.3.	Estimation of the spread parameter, $\sigma$ .....	107
5.5.	Implementation of the proposed algorithm .....	108
5.6.	Chapter conclusion .....	111
Chapter 6 COLLECTION METHODS FOR SYDNEY BASIN AIR QUALITY DATA		
	112	
6.1.	Introduction .....	112
6.2.	The application domain: Sydney basin .....	113
6.2.1.	New South Wales Greater Metropolitan Region .....	113
6.3.	Data sets from the measurements .....	115
6.3.1.	Air pollutants measurement .....	115
6.3.2.	Instrumentation and its operation.....	116
6.3.3.	Calibration of instrumentation.....	119
6.4.	Data sets from the emissions inventory database.....	120
6.4.1.	Air emissions inventory in New South Wales.....	120
6.4.2.	Air emissions inventory database system .....	121
6.5.	Data sets from the air quality model: TAPM-CTM.....	124
6.5.1.	Overview of TAPM-CTM model .....	124
6.5.2.	Data display system for the air quality model .....	127
6.6.	Chapter conclusion .....	128
Chapter 7 BACKGROUND OZONE LEVEL DETERMINATION IN THE SYDNEY BASIN		
	129	
7.1.	Introduction .....	129
7.2.	The night-time BOL determination .....	130
7.2.1.	Theoretical background .....	130
7.2.2.	Methodology .....	131
7.2.3.	Case study: Analysis of night-time BOL in Sydney .....	132
7.2.3.1.	Night-time BOL statistical results .....	132
7.2.3.2.	Night-time BOL temporal trend .....	133
7.2.3.3.	Discussion.....	136
7.3.	The daytime BOL determination.....	136

7.3.1.	Methodology .....	136
7.3.2.	Case study: Analysis of daytime BOL in Sydney .....	137
7.3.2.1.	Analysis of local background oxidant level .....	137
7.3.2.2.	Background oxidant level temporal trend.....	143
7.3.2.3.	Discussion.....	144
7.4.	Quantisation methods to refine the night-time BOL definition .....	145
7.4.1.	A method to deal with unavailable pollutant data.....	145
7.4.2.	Night-time BOL based on O <sub>3</sub> -NO <sub>x</sub> relationship .....	147
7.4.2.1.	Methodology.....	148
7.4.2.2.	Validity of the proposed method .....	149
7.4.3.	A generic method to determine the duration time for night-time BOL .....	151
7.4.3.1.	Methodology.....	151
7.5.	Chapter conclusion .....	153
Chapter 8	METAMODEL APPLICATION IN AIR QUALITY MODELLING.....	155
8.1.	Introduction .....	155
8.2.	Short-term temporal prediction model of BOL.....	156
8.2.1.	Model development.....	156
8.2.2.	Analysis and discussion .....	158
8.2.2.1.	Prediction results .....	158
8.2.2.2.	Performance analysis and discussion.....	161
8.3.	Long-term estimated model of BOL.....	163
8.3.1.	Model development.....	163
8.3.2.	Analysis and discussion .....	164
8.3.3.	Extended model for long-term estimation of BOL .....	168
8.3.3.1.	Model input-output .....	168
8.3.3.2.	Results and analysis .....	170
8.3.4.	Discussion .....	174
8.4.	A metamodel approach for air pollutant spatial estimation.....	176
8.4.1.	Model development for ozone distribution .....	176
8.4.1.1.	Input-output parameters of the model.....	176

8.4.1.2.	NO <sub>x</sub> emission distribution .....	178
8.4.1.3.	Training and verification of the model .....	180
8.4.2.	Study case: results and analysis .....	181
8.4.2.1.	The application domain .....	181
8.4.2.2.	Neural network model implementation .....	181
8.4.2.3.	Model performance .....	185
8.4.2.4.	Performance comparison .....	188
8.4.3.	Discussion .....	191
8.5.	Chapter conclusion .....	192
Chapter 9	CONCLUSION AND FUTURE WORK .....	194
9.1.	Conclusion .....	194
9.2.	Direction for future work .....	197
BIBLIOGRAPHY	.....	199
APPENDICES	.....	215
Appendix A:	Matlab codes for three DOE methods .....	215
A-1	Matlab codes for Weighted Clustering Design (WCD) sampling method..	215
A-2	Matlab codes for Latin Hypercube Design (LHD) sampling method. ....	216
A-3	Matlab codes for n-level Full Factorial Design (n-FFD) sampling method.	218
A-4	Codes to generate large data points from a known function .....	220
Appendix B:	Matlab codes for some improvement of OLS algorithm .....	221
B-1	Codes for RBFNN with adaptively tuned spread parameter .....	221
B-2	Pruning algorithm codes (additional codes of Appendix B-1) .....	225
Appendix C:	The proposed algorithm codes for RBFNN .....	226
C-1	RBFNN with GRFSWLS strategy .....	226
Appendix D:	Related documents for the data collection methods .....	232
D-1	Coordinate locations of the monitoring stations in NSW, Australia .....	232
D-2	Typical instrument used in the monitoring stations in NSW .....	234
Appendix E:	Codes for RBFNN metamodel application .....	235
E-1	Metamodel codes for hourly prediction of BOL .....	235



E-2	A generic modelling codes for the estimation of BOL.....	237
E-3	The m-codes for the spatial estimation of ozone concentration .....	237

## LIST OF TABLES

<b>Table</b>	<b>Page</b>
Table 1.1 Six basic pollutants found in Australia and their characteristics.....	3
Table 2.1 A comparison of air quality reference methods used in U.S. and Australia monitoring stations. ....	15
Table 2.2 Table comparing U.S., EU and Australia air quality standards for criteria pollutants. ....	20
Table 2.3 Typical domain dimensions for different scale models.....	22
Table 2.4 Atmospheric Stability classes (Turner, 1970). ....	25
Table 3.1 Metamodel comparison results for the test function. ....	66
Table 4.1 The performance of OLS using different spread parameters (i.e. $\sigma = \sigma_c$ ). .....	89
Table 4.2 The performance of improved OLS using different $sp_c$ parameters ( $MSE$ goal=0.001).....	91
Table 4.3 The comparison of training performance using three methods, at different spc values and MSE goals. ....	93
Table 7.1 Statistics of non-photochemical night-time BOL at monitoring sites in the Sydney basin.....	132
Table 7.2 Comparison between episode and non-episode background oxidant levels at Sydney basin sites in 1999.....	140
Table 7.3 Comparison between episode and non-episode background oxidant levels at Sydney basin sites in 2000.....	141
Table 7.4 Episode and non-episode background oxidant levels at Sydney basin sites from 2001 to 2005. ....	142
Table 7.5 Background ozone level determination (in ppb) at St Marys using several methods.....	145
Table 7.6 Comparison of pollutant concentrations and the estimated BOL at two sites in Sydney. ....	147

Table 7.7 The suggested start (evening time) and end (next morning time) to determine the non-photochemical night-time BOL.....	153
Table 8.1 The RBFNN model setting for short-term prediction of BOL.....	158
Table 8.2 Performance indexes on the test set for the simulation performed by different combination of inputs (at Randwick site).....	165
Table 8.3 Performance indices on the test set for the simulation performed by different combination of inputs (at Blacktown site).....	168
Table 8.4 Yearly night-time BOL mean values at various sites around Sydney from 2005 to 2010. ....	173
Table 8.5 The coefficient values for calculating the $\sigma_y$ and $\sigma_{\hat{y}}$ . ....	180
Table 8.6 Comparison of the training performance between three methods for different <i>MSE</i> goals. ....	190

## LIST OF FIGURES

Figures	Page
Fig. 1.1 Air pollutant sources: (a) industrial emissions, (b) motor vehicles, and (c) natural sources. ....	2
Fig. 2.1 An example of air quality monitoring station located in Sydney, Australia. (Photo courtesy of Department of Environment and Climate Change, DECC). ....	15
Fig. 2.2 Air quality categories based on the AQI values. ....	21
Fig. 2.3 Schematic representation of Gaussian plume model. ....	24
Fig. 2.4 Full factorial design for two-dimensional problem. ....	39
Fig. 2.5 An example of Latin hypercube design for two-dimensional problem with the number of sample points being 10. ....	40
Fig. 2.6 A fundamental architecture of radial basis function neural network. ....	41
Fig. 2.7 A comparison of one-dimensional Gaussian, multiquadratic and inverse multiquadratic functions with $c = 0$ and $\sigma = 1$ . ....	43
Fig. 2.8 A one-dimensional thin plate spline function with $c = 0$ . ....	44
Fig. 2.9 A generalised radial basis function neural network. ....	52
Fig. 3.1 A metamodel map between $k$ input variables and an output, with $n$ number of patterns. ....	55
Fig. 3.2 A metamodeling flowchart by sequential sampling process. ....	61
Fig. 3.3 The proposed metamodeling flowchart. ....	62
Fig. 3.4 A distribution of data points in three-dimensional space. ....	64
Fig. 3.5 The $R^2$ performance index against the sample number (in percentages). ....	67
Fig. 3.6 The $RMSE$ performance index against the sample number (in percentages). ....	68
Fig. 3.7 The $d_2$ performance index against the sample number (in percentages). ....	68
Fig. 3.8 The MAE performance index against the sample number (in percentages). ....	69
Fig. 3.9 The comparison of the training performance between three sampling methods, with several sample sizes. ....	71

Fig. 3.10 The estimation output for 100 test data of problem 1 using $N=1000$ (i.e. case no. 3 in Table 3.1).....	72
Fig. 4.1 RMSE value versus the spread parameter with various neuron numbers from a case study. ....	81
Fig. 4.2 RBFNN with adaptively tuned spread parameter algorithm flowchart. ....	83
Fig. 4.3 RBFNN with an additional pruning algorithm.....	86
Fig. 4.4 The evolution of OLS learning process using different <i>MSE</i> goals.....	88
Fig. 4.5 The variations of spread parameters of the hidden units using different threshold values ( $\varepsilon = 0.001, 0.0001, 0.00001, 0.000001$ ), and $sp_c = 6.0$ for the centres selection.....	90
Fig. 4.6 The training performance of OLS and adaptive-OLS using different $sp_c$ values, at a threshold value, $\varepsilon$ of 0.00001 and <i>MSE</i> goal of 0.001.....	92
Fig. 5.1 Generalised radial basis function network scheme with regularised forward selection and weighted least square (GRFSWLS).....	110
Fig. 6.1 Greater Metropolitan Region in New South Wales (NSW), Australia and its monitoring sites. (Map source: DECC, 2007a) .....	113
Fig. 6.2 (a) Typical monitoring station layout; (b) some air pollutants analysers in the instrument rack; (c) TEOM, devices for measuring fine particles; (d) internal view of TEOM device; and (e) data acquisition software for collecting the measurements data. (Photos courtesy of OEH, NSW). .....	115
Fig. 6.3 Examples of ozone analyser models used by EPA: (a) ML8810, (b) TE49C, and (c) EC9810. ....	117
Fig. 6.4 Ozone analyser flow schematic for model TE49C.....	117
Fig. 6.5 Oxides of nitrogen analyser flow schematic for model TE42C.....	119
Fig. 6.6 (a) devices to produce pollutants gas standards; (b) a device to keep the ozone gas standard; (c) a multi-gas calibrator used in the instrument calibration. (Photos courtesy of OEH, NSW) .....	120
Fig. 6.7 (a) EDMS splash screen; (b) the main form where a user can choose the function to be performed. ....	122
Fig. 6.8 EDMS output file structure for a TAPM-CTM area source emission file. ....	122
Fig. 6.9 Creating emissions files for TAPM-CTM model.....	123
Fig. 6.10 The three processes involved in the simulation of TAPM-CTM model. ....	125
Fig. 6.11 Main GUI of the TAPM software. ....	126

Fig. 6.12 Air pollutant level of $\text{NO}_2$ at 4:00pm as displayed by CTM-DDS system.	127
Fig. 7.1 Background trends from night-time to early morning at St Marys' station when nitrogen oxide (NO) is zero, based on ozone concentration.....	134
Fig. 7.2 Background trends from night-time to early morning at St Marys' station when nitrogen oxide (NO) is zero, based on oxidant ( $\text{Ozone} + \text{NO}_2$ ) concentration. ....	135
Fig. 7.3 Daytime oxidant level versus $\text{NO}_x$ in Sydney West in 1998: broken lines for episode, and solid lines for non-episode days.....	139
Fig. 7.4 Daytime background oxidant trend for non-episode days at several Sydney sites from 1998 to 2005. ....	143
Fig. 7.5 Daytime background oxidant trend for episode days at several Sydney sites from 1998 to 2005. ....	144
Fig. 7.6 Correlation of $\text{O}_3$ and NO to estimate the background ozone level at an absence point. ....	146
Fig. 7.7 Daily average of non-photochemical night-time BOL shown by intercept of the linear regression line.....	148
Fig. 7.8 Hourly average of non-photochemical night-time BOL shown by the intercept of the dashed lines. ....	149
Fig. 7.9 Comparison of background ozone level between the proposed method and mean concentration at a semi-pristine site.....	150
Fig. 7.10 Hourly data at Randwick station showing BOL equal to ozone concentration when $\text{NO}_x = 0$ . (The right axis represents $\text{NO}_x$ scale in ppb). .....	150
Fig. 7.11 Diurnal distributions of $\text{O}_3$ and $\text{NO}_x$ at the Bringelly site during summer 2003. ....	152
Fig. 7.12 Diurnal distributions of $\text{O}_3$ and $\text{NO}_x$ at the Bringelly site during winter 2003. ....	152
Fig. 8.1 Inputs and outputs of RBFNN model for short-term prediction of BOL...	157
Fig. 8.2 Short-term prediction results of BOL at Blacktown site for: a) 1-hour; b) 2- hour; c) 3-hour; d) 6-hour; e) 18-hour; and f) 24-hour, respectively. ....	159
Fig. 8.3 Short-term prediction results of BOL at Vineyard for: a) 1-hour; b) 6-hour, respectively.....	160

Fig. 8.4 The models performance for hourly BOL prediction for 24 hours at Blacktown site: a) performance based on $R^2$ index; b) performance based on $MAE$ index. ....	162
Fig. 8.5 The models performance for hourly BOL prediction for 24 hours at Vineyard site: a) performance based on $R^2$ index; b) performance based on $MAE$ index. ....	162
Fig. 8.6 Predicted background ozone level and the observed data at Randwick station over night-time and early morning hourly data. ....	166
Fig. 8.7 Predicted background ozone level and the observed data at Blacktown station over night-time and early morning hourly data. ....	167
Fig. 8.8 A generic metamodel structure to estimate the BOL in the evaluated domain. ....	169
Fig. 8.9 Comparison between night-time BOL estimated by metamodel, and expected values derived by the measured data at St Marys site. ....	171
Fig. 8.10 Comparison between night-time BOL estimated by metamodel, and expected values derived by the measured data at Bringelly site. ....	172
Fig. 8.11 BOL profiles in the Sydney basin over a six year period. ....	174
Fig. 8.12 Comparison between night-time BOL estimated by the metamodel and the expected values derived by the measured data at two sites that were not used in the training data set, namely Macarthur and Bargo. ....	175
Fig. 8.13 Inputs and output for training the RBFNN model for spatial estimation of ozone concentration. ....	179
Fig. 8.14 Regression analysis for determination of the correlation ratio between simulated and observed ozone level at a monitoring site. ....	183
Fig. 8.15 Daily NO <sub>x</sub> distribution for a day in summer: (a) post-process by TAPM-CTM from the emission inventory, (b) added with the calculated emission. ....	184
Fig. 8.16 Scatter plot to illustrate the performance of validation phase. ....	185
Fig. 8.17 Spatial distribution for 8-hour maximum average of ozone by using RBFNN and TAPM-CTM model (Note: the bullet dots show the location of the monitoring stations). ....	187
Fig. 8.18 Performance comparison between RBFNN and TAPM-CTM predictions for eight-hour maximum average of ozone at 10 sites in the Sydney basin. ....	189

Fig. 8.19 The comparison of the training performance between GRFSWLS, OLS and FS methods. ....	190
Fig. 8.20 The results of using higher decimal places for finding the regularisation parameter value in the FS method. ....	191



## LIST OF ABBREVIATIONS

$\sigma$	- RBF spread parameter
$\sigma_c$	- Isotropic spread parameter for RBF centres selection
$\lambda$	- Regularisation parameter
AQM	- Air quality management
BOL	- Background ozone level
CGS	- Classical Gram-Schmidt
$d_2$	- Index of agreement
DAQM	- Deterministic air quality model
DOE	- Design of experiments
EPA	- Environment Protection Authority (in Australia)
FFD	- Full factorial design
FS	- Forward selection
GCV	- Generalised cross-validation
GLS	- Generalised least squares
GMR	- Greater Metropolitan Region
LHD	- Latin hypercube design
LOO-CV	- Leave one out cross-validation
LS	- Least squares
<i>MAE</i>	- Mean absolute errors
<i>MSE</i>	- Mean squared errors
NAAQS	- National Ambient Air Quality Standards
NEPM	- National Environment Protection Measures
NO	- Nitric oxide
NO <sub>2</sub>	- Nitrogen dioxide
NO <sub>x</sub>	- Nitrogen oxides
O <sub>3</sub>	- Ozone
OLS	- Orthogonal least squares
PM	- Particulate matter
<i>ppb</i>	- parts per billion

<i>pphm</i>	- parts per hundred million
$R^2$	- Determination coefficient
RBFNN	- Radial basis function neural network
<i>RMSE</i>	- Root mean square errors
<i>SSE</i>	- Sum of squared errors
TEMP	- Ambient temperature
US-EPA	- U.S. Environmental Protection Agency
VOCs	- Volatile organic compounds
WCD	- Weighted clustering design
WDR	- Wind direction
WLS	- Weighted least squares
WSP	- Wind speed