



THE UNIVERSITY  
*of* ADELAIDE

Improving Partial Mutual Information Based  
Input Variable Selection for Data Driven  
Environmental and Water Resources Models

XUYUAN LI

B.E. (Hons)

Thesis submitted in fulfilment of the requirements for the degree  
of Doctor of Philosophy

The University of Adelaide  
Faculty of Engineering, Computer and Mathematical Sciences  
School of Civil, Environmental and Mining Engineering

Copyright© March 2015



# Improving Partial Mutual Information Based Input Variable Selection for Data Driven Environmental and Water Resources Models

By:

Xuyuan Li, *B.E. (Hons)*

Supervised by:

Professor Holger R. Maier, *B.E. (Hons), Ph.D., MIEAust, CPEng (NPER)*  
*Professor of Integrated Water Systems Engineering, Associate Editor, Water Resources Research, Member of Editorial Board, Environmental Modelling and Software, School of Civil, Environmental & Mining Engineering, The University of Adelaide*

Dr. Aaron C. Zecchin, *Ph.D., B.E. (Hons), B.Sc. (Math & Comp. Sci.)*,  
*Senior lecturer, School of Civil, Environmental & Mining Engineering, The University of Adelaide*

Thesis submitted in fulfillment of the requirements for the degree of  
**Doctor of Philosophy**

School of Civil, Environmental & Mining Engineering  
Faculty of Engineering, Computer and Mathematical Sciences  
The University of Adelaide  
North Terrace, Adelaide, SA 5005, Australia  
Phone: +61 8 8313 1575  
Fax : +61 8 8303 4359  
Email: [xli@civeng.adelaide.edu.au](mailto:xli@civeng.adelaide.edu.au), [xliadelaide@gmail.com](mailto:xliadelaide@gmail.com)  
Copyright© Xuyuan Li, March, 2015.



## TABLE OF CONTENT

LIST OF FIGURES .....	VIII
LIST OF TABLES .....	XI
NOMENCLATURE & ABBREVIATIONS.....	XIV
ABSTRACT .....	XX
STATEMENT OF ORIGINALITY .....	XXIV
ACKNOWLEDGEMENTS .....	XXVI
Chapter 1 INTRODUCTION .....	1
1.1 Background .....	1
1.1.1 ANNs in environmental and water resources modelling.....	1
1.1.2 IVS.....	1
1.1.3 PMI IVS .....	3
1.1.4 Bandwidth issue in PMI IVS.....	6
1.1.5 Boundary issue in PMI IVS.....	7
1.2 Objectives .....	13
1.3 Thesis overview.....	15
Chapter 2 JOURNAL PAPER 1 - <i>Selection of Smoothing Parameter</i> <i>Estimators for General Regression Neural Networks - Applications to</i> <i>Hydrological and Water Resources Modelling</i> .....	19
2.1 Introduction .....	23
2.2 GRNNs .....	27
2.3 Methodology .....	29
2.3.1 Procurement of input/output data with different degrees of normality and non-linearity .....	31
2.3.2 Estimation of GRNN smoothing parameters using different estimation methods.....	38
2.3.3 Development of benchmark MLP model .....	44
2.3.4 Model performance assessment.....	44
2.3.5 Test regime .....	46

## TABLE OF CONTENT

---

2.4 Results and discussion .....	47
2.4.1 Synthetic case studies .....	47
2.4.2 Real case studies .....	54
2.5 Summary and conclusions .....	58
2.6 Acknowledgments.....	60
Chapter 3 JOURNAL PAPER 2 - <i>Improved PMI-Based Input Variable Selection Approach for Artificial Neural Network and Other Data Driven Environmental and Water Resource Models</i> .....	
3.1 Introduction.....	65
3.2 PMI IVS .....	67
3.3 Methodology .....	71
3.3.1 Generation of input/output data with different degrees of normality .....	73
3.3.2 Estimation of PDF and MI using different bandwidth estimators .	75
3.3.3 Performance assessment .....	80
3.3.4 Test regime.....	80
3.4 Results and discussion .....	82
3.4.1 Selection accuracy .....	82
3.4.2 Computational efficiency.....	91
3.4.3 Suggested rules and guidelines .....	91
3.5 Testing of proposed rules and guidelines.....	94
3.6 Summary and conclusions .....	102
3.7 Acknowledgments.....	104
Chapter 4 JOURNAL PAPER 3 - <i>Improved Partial Mutual Information-Based Input Variable Selection by Consideration of Boundary Issues Associated With Bandwidth Estimation</i> .....	
4.1 Introduction.....	110
4.2 Background on PMI IVS and Boundary Issues .....	113

## TABLE OF CONTENT

---

4.2.1 PMI IVS .....	113
4.2.2 Boundary issues in PMI IVS .....	115
4.2.3 Potential solutions to solve boundary issues in PMI IVS .....	117
4.3 Methodology .....	120
4.3.1 Generate input/output data with different degrees of normality ..	122
4.3.2 Estimate MI using different boundary correctors and suggested bandwidth estimators.....	124
4.3.3 Estimate residuals using alternative approaches and suggested bandwidth estimators.....	127
4.3.4 Test regime .....	132
4.3.5 Assess performance of IVS over 30 trials .....	134
4.4 Results and Discussion .....	135
4.4.1 Selection accuracy .....	136
4.4.2 Computational efficiency .....	147
4.4.3 Suggested rules and guidelines.....	152
4.5 Validation on Murray Bridge and Kentucky River Basin case studies .....	155
4.5.1 Background .....	155
4.5.2 Experimental Procedure .....	158
4.5.3 Results and discussion.....	158
4.6 Summary and Conclusions .....	160
4.7 Acknowledgments .....	163
Chapter 5 CONCLUSIONS .....	165
5.1 Thesis summary.....	165
5.2 Research contributions .....	167
5.3 Publications .....	172
5.4 Recommendations for future research.....	173
REFERENCES .....	176

## TABLE OF CONTENT

---

APPENDICES .....	192
APPENDIX-A Supplementary Material from Paper 1 (Chapter 2).....	192
APPENDIX-B Supplementary Material from Paper 2 (Chapter 3).....	220
B.1 Mathematical derivations .....	220
B.2 Supplementary figures and tables .....	231
APPENDIX-C Supplementary Material from Paper 3 (Chapter 4).....	243
C.1 Mathematical explanation and derivations.....	243
C.2 Supplementary figures and tables .....	248
APPENDIX-D Copy of Publications.....	256
D.1 Copy of Paper 1 from Chapter 2 (as published).....	256
D.2 Copy of Paper 2 from Chapter 3 (as published).....	283



## TABLE OF CONTENT

---

**LIST OF FIGURES**

Figure 1.1 Framework of thesis ..... 9

Figure 2.1 General architecture of a GRNN ..... 27

Figure 2.2 Overview of proposed assessment approach ..... 30

Figure 2.3 Predictive accuracy for the validation data of MLPs and GRNNs for different synthetic data-generating models and distributions for which optimal parameters have been obtained using different methods ..... 48

Figure 2.4 Computational efficiency of MLPs and GRNNs for different synthetic data-generating models and distributions for which optimal parameters have been obtained using different methods ..... 50

Figure 2.5 Suggested smoothing parameter estimators under different problem situations ..... 53

Figure 2.6 Predictive accuracy of MLPs and GRNNs with different smoothing parameter estimators for the validation data for the real case studies ..... 56

Figure 2.7 Predictive efficiency of MLPs and GRNNs with different smoothing parameters for the validation data for the real case studies ..... 57

Figure 3.1 Procedure of PMI IVS adopted in this study ..... 71

Figure 3.2 Outline of the proposed experimental approach..... 72

Figure 3.3 Correct selection rate of EAR4 model with alternative bandwidth estimators ..... 86

Figure 3.4 Correct selection rate of TEAR10 model with alternative bandwidth estimators ..... 86

Figure 3.5 Correct selection rate of NL model with alternative bandwidth estimators ..... 86

Figure 3.6 KDE accuracy measured by K-S statistics for EAR4 & TEAR10 models ..... 87

Figure 3.7 Residual accuracy measured by CE for EAR4 model..... 89

Figure 3.8 KDE accuracy measured by K-S statistics for NL model ..... 89

Figure 3.9 Residual accuracy measured by CE for NL model ..... 90

Figure 3.10 Computational efficiency of EAR4 model with different bandwidth estimators ..... 90

Figure 3.11 Suggested bandwidth estimators under different distribution scenarios..... 93

## LIST OF FIGURES

---

Figure 3.12 The River Murray in South Australia (Maier and Dandy, 1996).	95
Figure 3.13 Correct selection rate and efficiency of salinity forecast at Murray Bridge with proposed and alternative bandwidth estimators .....	98
Figure 3.14 The Kentucky River Basin in USA (Jain et al., 2004).....	99
Figure 3.15 Correct selection rate and efficiency of flow forecast at Kentucky River Basin with proposed and alternative bandwidth estimators .....	102
Figure 4.1 Graphical representation of the boundary issue in 2D (Hazelton and Marshall, 2009) .....	117
Figure 4.2 Taxonomy of methods for dealing with boundary issues in mutual information and residual estimation .....	120
Figure 4.3 Overview of the proposed analysis for the PMI IVS influenced by bandwidth and boundary issues.....	121
Figure 4.4 Selection accuracy of the PMI with suggested settings for EAR4 models .....	139
Figure 4.5 Relative change of K-S and MI in-between M1 and B3 for EAR4 model.....	140
Figure 4.6 Accuracy of residual estimation with alternative estimators for EAR4 model (3 cases).....	141
Figure 4.7 Selection accuracy of the PMI with suggested settings for TEAR10 models .....	143
Figure 4.8 Selection accuracy of the PMI with suggested settings for NL models .....	144
Figure 4.9 Accuracy of residual estimation with alternative estimators for TEAR10 model (3 cases) .....	145
Figure 4.10 Accuracy of residual estimation with alternative estimators for NL model (3 cases).....	146
Figure 4.11 Selection efficiency of the PMI IVS with tested methods for EAR4 models .....	149
Figure 4.12 Suggested PMI IVS approaches under distinct scenarios.....	153
Figure 4.13 Selection accuracy and efficiency of the PMI IVS with suggested settings for Murray Bridge case .....	159
Figure 4.14 Selection accuracy and efficiency of the PMI IVS with suggested settings for Kentucky River basin case .....	160

## LIST OF FIGURES

---

## LIST OF TABLES

Table 1.1 Review of input variable selection methods for ANNs applied to environmental and water resources problems (developed based on May, 2010) .....	10
Table 1.2 Bandwidth estimators applied within the statistics literature .....	12
Table 1.3 Boundary correctors proposed within the statistics literature .....	12
Table 2.1 Details of the simulated input distributions for the time series models (EAR4, TEAR10) .....	32
Table 2.2 Details of the simulated input distributions for the nonlinear model (NL) .....	32
Table 2.3 Inputs and outputs used to forecast salinity at Murray Bridge 1, 5, & 14 days in advance .....	35
Table 2.4 Inputs and output used to model rainfall-runoff from the Kentucky River basin .....	36
Table 2.5 Selected smoothing parameter estimators with different fitness functions and assumptions of normality and error basis .....	37
Table 3.1 Details of the distributions used to generate values of the exogenous input variables and the statistical properties of the generated data for all time series models (EAR4, TEAR10) .....	73
Table 3.2 Details of the distributions used to generate values of the input variables and the statistical properties of the generated data for the non-linear model (NL) .....	74
Table 3.3 GRNN bandwidth estimation techniques used for residual estimation during the PMI IVS process (based on the guidelines from Li et al. (2014b)) .....	81
Table 3.4 Average ratio of different kernel bandwidths under different distribution scenarios for EAR4 model .....	85
Table 3.5 Candidate inputs and output for the salinity case study .....	96
Table 3.6 Candidate inputs and output used for the rainfall-runoff case study .....	100
Table 4.1 Details of the distributions used to generate values of the exogenous input variables and the statistical properties of the generated data for all time series models (EAR4, TEAR10) .....	122

## LIST OF TABLES

---

Table 4.2 Details of the distributions used to generate values of the input variables and the statistical properties of the generated data for the non-linear model (NL).....	123
Table 4.3 GRNN bandwidth estimation techniques used for residual estimation during the PMI IVS .....	128
Table 4.4 Different approaches used for PMI IVS by considering bandwidth and boundary issues .....	133
Table 4.5 Candidate inputs and output used to forecast salinity at Murray Bridge 14 days in advance .....	156
Table 4.6 Candidate inputs and outputs used to forecast flow at Kentucky River Basin 1 day in advance.....	157

## LIST OF TABLES

---

## NOMENCLATURE & ABBREVIATIONS

### Symbols

$\widehat{m}_{v_i}(X_{i^*}^j)$ : residual estimate of  $v_i$  based on  $X_{i^*}$

$(X^j, y^j)$ : observed pairs of input and output data

$\widehat{\phi}_4(g) = n^{-1} \sum_{i=1}^n \widehat{L}^{(4)}(X^i; g)$ : fourth order integrated squared density derivative

$F_{emp}(X_i^j)$ : empirical CDF of the input variable estimated by a histogram

$F_{est}(X_i^j)$ : estimated kernel based CDF of the input variable

$I_{X_i, y}$ : mutual information

$I_{v_i, u}$ : partial mutual information

$\widehat{ISB}(h)$ : estimation of the integrated squared bias

$K^{(n)}$ :  $n^{th}$  derivative of kernel function  $K$

$\widehat{R}(f'')$ : approximation of the integrated squared second derivative of  $f$

$S_{x_i}^2$ : sample variance of the input  $X_i$

$S_{xy, i}$ : covariance between input  $X_i$  and output  $y$

$S_y^2$ : sample variance of output  $y$

$X_{i^*}$ : selected inputs

$X^{(j)}$ :  $j$ -th data point formed by the interpolated and original data points

$X_S$ : significant input set

$\widehat{f}(\mathbf{X}, y)$ : estimation of the joint probability density function between inputs  $\mathbf{X}$  and output  $y$

$p_{t-n}$ : exogenous input with lag  $n$

$\widehat{y}$ : estimation of the actual output  $y$

$\bar{y}$ : sample mean of the observations

$\mathbf{e}_1$ : vector having 1 in the first entry and 0 elsewhere

$\varepsilon_t$ : introduced error term

$\mu_2(K) = \int x^2 K(x) dx$ : second moment of  $K$

$\mu_k(L) = \int u^k L(u) du$ :  $k$ -th moment of  $L$

$\mu_n(K)$ :  $n^{th}$  moment of kernel function  $K$

$\rho_{xy, i}$ : correlation coefficient between input  $X_i$  and output  $y$



## NOMENCLATURE & ABBREVIATIONS

---

$\sigma_i$ : sample standard deviation of the  $X_i^j$

$*$ : convolution operation

$h$ : kernel bandwidth

$K$ : kernel function

$B(u; h_x)$ : univariate boundary kernel with bandwidth  $h_x$  and variable

$$u = (X_i - X_i^j)/h_x$$

$B(u, v; \mathbf{H})$ : bivariate boundary kernel with bandwidth matrix  $\mathbf{H}$  and vector

$$(u, v) \text{ where } u = (X_i - X_i^j)/h_x \text{ and } v = (y - y^j)/h_y$$

$E[y|\mathbf{X}]$ : conditional expectation of output  $y$  given input  $\mathbf{X}$

$L$ : pilot kernel

$O(h)$ : bias of density function

$P(t)$ : lagged effective rainfall

$Q(t - 1)$ : lagged runoff

$R(K) = \int [K(x)]^2 dx$ : integrated square of kernel function

$a$ : left boundary of kernel density

$c$ : right boundary of kernel density

$e$ : number of effective inputs

$f(\mathbf{X}, y)$ : joint probability density function between inputs  $\mathbf{X}$  and output  $y$

$g$ : pilot bandwidth

$k$ : kurtosis

$k$ : order of pilot kernel  $L$

$m(x)$ : regression function

$m$ : number of inputs

$n$ : number of observations

$r$ : stage number of  $L$

$s$ : skewness

$sup$ : supremum function

$$\mathbf{H} = h_i^2 \begin{bmatrix} S_{x,i}^2 & S_{xy,i} \\ S_{xy,i} & S_y^2 \end{bmatrix}: \text{bivariate bandwidth matrix}$$

$$\mathbf{X} = [X_1 \dots X_m]^T: \text{input variables}$$

$$\mathbf{h} = [h_1 \quad \dots \quad h_n]^T: \text{kernel bandwidth vector}$$

### Abbreviation

ACF: auto-correlation function

AIC: Akaike information criterion

AMISE: asymptotic mean integrated squared error

ANNs: artificial neural networks

BCV: biased cross validation

BCVDPI: a combination of BCV and DPI

BE: backward elimination (pruning)

BJ: Box-Jenkins

BK: boundary kernel

BP: back-propagation algorithm

CE: coefficient of efficiency

CK: conventional kernel

CSR: correct selection rate

CT: computational time

DELSA: distributed evaluation of local sensitivity analysis

DPI: 2-stage direct plug-in

EAR4: exogenous auto-regressive time series model (with time order up to 4)

EMISE: exact mean integrated squared error

ES: exhaustive search

ETC: empirical translation correction

EVT1: extreme value type I distribution

EXP: exponential distribution

FS: forward selection

GAMMA: gamma distribution

GRNN: general regression neural network

GRR: Gaussian reference rule

GSS: golden section search

HS: heuristic search

ICAIVS: hybrid independent component analysis and input variable selection filter

IIS: tree-based iterative input variable selection

IoAd: index of agreement

## NOMENCLATURE & ABBREVIATIONS

---

IVS: input variable selection  
KDEs: kernel density estimates  
K-S: Kolmogorov-Smirnov statistic  
KT: kernel transformation  
L1UL: Lock 1 upper river level  
LBE: local bandwidth (enlarging)  
LBR: local bandwidth (reducing)  
LHOP: local high order polynomial  
LLM: local linear method  
LLP: local linear polynomial  
LOGN: log-normal distribution  
LOGPT3: log-Pearson type III distribution  
LOS: Loxton river salinity  
LQP: local quadratic polynomial  
LSCV: least squared cross validation  
MAE: mean absolute error  
MAS: Mannum river salinity  
MBS: Murray Bridge river salinity  
MCE: modified coefficient of efficiency  
MI: mutual information  
MIoAd: modified index of agreement  
MLPANNs: multi-layer perceptron artificial neural networks  
MLPs: multi-layer perceptrons  
MOS: Morgan river salinity  
MPI: modified persistence index  
MVC: multi-variable calibration  
MVCA: multi-variable calibration with mean absolute error as the objective function  
MVCS: multi-variable calibration with squared error as the objective function  
NL: nonlinear input-output function  
NORM: normal distribution  
NS: normal scale  
OM: optimisation method  
PA: pseudo-data approach

## NOMENCLATURE & ABBREVIATIONS

---

PACF: partial auto-correlation function

PC: partial correlation

PCA: principal component analysis

PDF: probability density function

PI: persistence index

PMI: partial mutual information

PNNs: probabilistic neural networks

PSO: particle swarm optimisation

PT3: Pearson type III distribution

RBFs: radial basis functions

RC: reflection correction

RE: residual estimation

RMSE: root mean squared error

RNNs: recurrent neural networks

RVSDM: recursive variable selection embedded in dynamic emulation models

SCV: smoothed cross validation

SOM-GAGRNN: self-organising map genetic algorithm general regression neural network

SVC: single variable calibration

SVCA: single variable calibration with mean absolute error as the objective function

SVCS: single variable calibration with squared error as the objective function

SVO: single variable optimisation

SVR: single variable regression

TEAR10: threshold exogenous auto-regressive time series model (with time order up to 10)

WAS: Waikerie river salinity

## NOMENCLATURE & ABBREVIATIONS

---

### ABSTRACT

Artificial neural networks (ANNs), as one of the most commonly used data driven models for environmental and water resources problems, have been applied successfully and extensively over the last two decades and are still gaining in popularity. Consideration of the methods used in the steps in the development of ANNs, which consist of data collection, data processing, input variable selection, data division, calibration and validation, are vitally important, as ANN model development is based on data, rather than understanding of the underlying physical processes.

Among these methods, input variable selection (IVS) plays a significant role, as the performance of the developed model can be compromised if inputs having a pronounced relationship with the modelled output are omitted. In contrast, calibration becomes extremely challenging and modelling validation, as well as knowledge extraction, are problematic if redundant or superfluous inputs are included. Given the facts explained above, various techniques have been developed for the sake of more accurate IVS.

Partial mutual information (PMI) is one of the most promising approaches to IVS, as it has a number of desirable properties, such as the ability to account for input relevance, the ability to cater to both linear and non-linear input-output relationships and the ability to check the redundancy of selected inputs. PMI is a stepwise input selection algorithm, which only selects one variable per iteration, as part of which the strength of the relationship between each potential input and the output is quantified using mutual information (MI) and input redundancy is accounted for by removing the influence of already selected inputs. This is achieved by developing models between the selected input and the output and assessing the strength of the relationship (in terms of MI) between the remaining potential inputs and the residuals of these models during the next iteration, which is referred to as PMI.

Although PMI IVS has already been applied successfully to a number of studies in hydrological and water resource modelling, present implementations predominantly depend on the assumption that the data used to develop the model follow a Gaussian distribution. This assumption has the

## ABSTRACT

---

potential to affect two steps in the PMI process, including the estimation of MI/PMI and the estimation of the residuals. In terms of MI/PMI estimation, this requires kernel density estimates of the modelling data to be obtained for the estimation of marginal and joint probability density functions (PDFs), which rely on estimates of kernel bandwidths (or smoothing parameters) and in most studies, the Gaussian reference rule is used for this purpose, which only results in optimal bandwidth estimates if the modelling data follow a Gaussian distribution. However, this is unlikely to be the case when dealing with water resources and other environmental data. In terms of residual estimation (RE), this has generally been done using general regression neural networks (GRNNs), which also require estimates of kernel bandwidths to be obtained and therefore suffers from the same issues as MI/PMI estimation.

The purpose of this thesis is to assess the impact the assumption that the data follow a Gaussian distribution has on the performance of PMI IVS and the efficacy of potential methods for overcoming any problems associated with this assumption. In order to achieve this, a large number of numerical tests are conducted on synthetic data with different degrees of normality and non-linearity, investigating the effectiveness of a range of options for (i) bandwidth estimation (caused by making Gaussian assumptions for non-Gaussian circumstances when adopting kernel based estimations in both MI/PMI and RE) and (ii) for dealing with boundary issues (caused by using a symmetrical kernel for bounded/unsymmetrical data when implementing kernel based estimations in both MI/PMI and RE), as well as methods for RE that do not require kernel density estimates. The results from these tests are used to develop preliminary guidelines for the selection of the most appropriate bandwidth and the most effective treatment of the boundary issue, which are then validated for two water resources case studies with different data properties and problem linearity, including forecasting of river salinity in the River Murray, Australia, and rainfall-runoff modelling in the Kentucky River, USA.

The major research contributions are presented in three journal publications. The motivations underlying these publications include: 1) the development and testing of rigorous and novel analytical procedures for assessing if, and to

## ABSTRACT

---

what degree, the performances of residual and MI estimates are affected by bandwidth selection and boundary issues; 2) clear explanation of the inaccurate performance of conventional PMI IVS under the influence of bandwidth selection and boundary issues; 3) the development of effective preliminary guidelines based upon synthetic studies dealing with both bandwidth selection and boundary issues under different scenarios categorised by data normality and problem linearity; 4) the development of more robust and reliable PMI IVS software for realistic environmental and water resource problems. Overall, the research outcomes suggest that the performance of PMI IVS is significantly influenced by bandwidth selection and boundary issues and can be effectively improved by following the proposed empirical guidelines, although the findings of this work could be tested more broadly, including for data sets with a wider range of attributes, such as different degrees of noise, collinearity and interdependency, as well as incomplete information.





## STATEMENT OF ORIGINALITY

I **Xuyuan Li** hereby certify that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

A list of works contained within this thesis is given in Section 5.3.

Signature \_\_\_\_\_ Date \_\_\_\_\_

## STATEMENT OF ORIGINALITY

---

### ACKNOWLEDGEMENTS

I would take this opportunity to gratefully acknowledge my principal supervisor Professor Holger R. Maier and co-supervisor Dr. Aaron C. Zecchin, who have supervised me with invaluable support, encouragement, and patience. I really appreciate the critical and rigorous attitude, timely and effective feedback, great foresight, and comfortable research environment provided by both of my supervisors. I have enjoyed studying under their supervision and learnt a lot from them during my candidature. Without their excellent supervision, the present thesis would hardly be possible.

I would like to sincerely thank Prof. A. Sharma, Dr. G.J. Bowden, Dr. R.J. May and Dr. G.B. Humphrey who brought me into the present research topic and kindly provided their suggestions and research materials.

I would express my appreciation to my fellow postgraduates within the School for sharing their wisdom, experiences, successes, and lessons with me. It was my pleasure to meet Dr. Xun Sun, Dr. Wenyan Wu, Dr. Feifei Zheng, Dr. Liang Huang, Dr. Jeffrey P. Newman, Dr. Christopher Stokes, and Dr. Tao Zhang and have had a number of in-depth academic discussions with them. Special thanks are also given to School Editor Barbara Brougham, Computer Technician Dr. Stephen Carr, all School Administrators, and the research scholarship (AGRS) provided by the University of Adelaide.

Last but not least, I would like to give an immense gratitude to my parents, Mr Xuelong Li and Ms Yongjing Liu, for their altruistic care and support from beginning to end (*ab ovo usque ad mala*).

## ACKNOWLEDGEMENTS

---