

# **Bayesian Artificial Neural Networks in Water Resources Engineering**

by Greer B. Kingston

Submitted in fulfilment of the requirements for the degree of  
**DOCTOR OF PHILOSOPHY**

May 2006

**FACULTY OF ENGINEERING, COMPUTER AND MATHEMATICAL SCIENCES**

---

School of Civil and Environmental Engineering



# Bayesian Artificial Neural Networks in Water Resources Engineering

By: Greer B. Kingston, *B.E. Civil and Environ. (Hons)*

May 2006

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

School of Civil and Environmental Engineering  
Faculty of Engineering, Computer and Mathematical Sciences  
The University of Adelaide SA 5005 Australia

Telephone: +61 8 8303 5451

Facsimile: +61 8 8303 4359

Web: [www.civeng.adelaide.edu.au](http://www.civeng.adelaide.edu.au)

Email: [enquiries@civeng.adelaide.edu.au](mailto:enquiries@civeng.adelaide.edu.au)

# Table of Contents

---

<b>Table of Contents</b>	<b>iii</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>xv</b>
<b>Abstract</b>	<b>xix</b>
<b>Statement of Originality</b>	<b>xxi</b>
<b>Acknowledgements</b>	<b>xxiii</b>
<b>List of Publications</b>	<b>xxv</b>
<b>Nomenclature &amp; Abbreviations</b>	<b>xxvii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Research Objectives . . . . .	3
1.3 Layout and Contents of Thesis . . . . .	5
<b>2 Research Background</b>	<b>7</b>
2.1 Water Resources Modelling . . . . .	7
2.2 Artificial Neural Networks (ANNs) . . . . .	9
2.2.1 Background and Description . . . . .	9
2.2.2 Multi-Layer Perceptrons (MLPs) . . . . .	10
2.2.3 Training . . . . .	12
2.2.4 Advantages . . . . .	12
2.2.5 Issues and Limitations . . . . .	13
2.3 Bayesian Methods . . . . .	24
2.3.1 Background to Bayesian Methodology . . . . .	24

2.3.2	Use of Bayesian Methods in Water Resources Modelling . . . . .	25
2.3.3	Bayesian Neural Networks . . . . .	26
2.3.4	Training and Prediction . . . . .	27
2.3.5	Model Selection . . . . .	33
2.3.6	Limitations of Current Bayesian Neural Network Practices . . . . .	35
2.4	Conclusion . . . . .	36
<b>3</b>	<b>State-of-the-Art Deterministic ANN Methodology</b>	<b>39</b>
3.1	Introduction . . . . .	39
3.2	Review of The Current State-of-the-Art . . . . .	40
3.2.1	Choice of Performance Criteria . . . . .	41
3.2.2	Choice of Data Sets . . . . .	46
3.2.3	Data Pre-Processing . . . . .	50
3.2.4	Determination of ANN Inputs . . . . .	55
3.2.5	Determination of ANN Architecture . . . . .	60
3.2.6	ANN Training . . . . .	68
3.2.7	ANN Validation . . . . .	74
3.3	Approach Adopted and Investigations Required . . . . .	75
3.4	Investigation of Deterministic ANN Development Methods . . . . .	77
3.4.1	Synthetic Data Sets . . . . .	77
3.4.2	Comparison of Training Algorithms . . . . .	80
3.4.3	Assessment of Model Selection Criteria . . . . .	93
3.4.4	Assessment of Input Importance Measures . . . . .	94
3.4.5	Results . . . . .	102
3.4.6	Evaluation of Best Models . . . . .	126
3.4.7	Conclusions . . . . .	133
<b>4</b>	<b>A New Bayesian Framework for ANNs</b>	<b>137</b>
4.1	Introduction and Motivation . . . . .	137
4.2	Bayesian Training and Prediction . . . . .	139
4.2.1	Markov Chain Monte Carlo Methods . . . . .	140
4.2.2	MCMC Methods Previously Applied to ANNs . . . . .	146
4.2.3	Proposed Bayesian Training Approach . . . . .	149
4.3	Bayesian Model Selection . . . . .	159
4.3.1	Computation of Evidence via Posterior Simulation . . . . .	159
4.3.2	Bayes Factors . . . . .	161

---

4.3.3	BMS Previously Applied to ANNs . . . . .	163
4.3.4	Proposed BMS Framework . . . . .	164
4.4	Assessment of Bayesian Techniques with Synthetic Data . . . . .	167
4.4.1	Bayesian Training and Prediction . . . . .	167
4.4.2	Bayesian Model Selection . . . . .	173
4.4.3	Results . . . . .	174
4.4.4	Evaluation of Best Models . . . . .	199
4.4.5	Conclusions . . . . .	212
<b>5</b>	<b>Case Study 1 - Salinity Forecasting in the River Murray</b>	<b>215</b>
5.1	Introduction . . . . .	215
5.2	Background . . . . .	215
5.2.1	Salinity in the River Murray . . . . .	216
5.2.2	Forecasting Salinity in the River Murray with ANNs . . . . .	220
5.3	Available Data and Model Inputs . . . . .	223
5.4	Deterministic ANN Development . . . . .	231
5.4.1	Methods . . . . .	231
5.4.2	Results . . . . .	233
5.5	Bayesian ANN Development . . . . .	241
5.5.1	Methods . . . . .	241
5.5.2	Results . . . . .	243
5.6	Conclusions . . . . .	254
<b>6</b>	<b>Case Study 2 - Cyanobacteria Forecasting</b>	<b>255</b>
6.1	Introduction . . . . .	255
6.2	Background . . . . .	256
6.2.1	Cyanobacteria in the River Murray . . . . .	256
6.2.2	Modelling Cyanobacteria Concentrations with ANNs . . . . .	258
6.3	Available Data and Model Inputs . . . . .	260
6.4	Deterministic ANN Development . . . . .	267
6.4.1	Methods . . . . .	267
6.4.2	Results . . . . .	269
6.5	Bayesian ANN Development . . . . .	275
6.5.1	Methods . . . . .	275
6.5.2	Results . . . . .	277
6.6	Conclusions . . . . .	287

<b>7</b>	<b>Conclusions and Recommendations</b>	<b>289</b>
7.1	Contributions of the Research . . . . .	289
7.1.1	Deterministic ANN Development . . . . .	289
7.1.2	Bayesian ANN Development . . . . .	291
7.2	Conclusions . . . . .	293
7.2.1	General . . . . .	293
7.2.2	Deterministic ANN Development . . . . .	295
7.2.3	Bayesian ANN Development . . . . .	296
7.3	Recommendations for Future Work . . . . .	297
	<b>References</b>	<b>299</b>
	<b>Appendices</b>	<b>325</b>
	<b>Appendix A Training Investigation Results</b>	<b>327</b>
	<b>Appendix B Results of Assessment of Input Importance Measures</b>	<b>331</b>

# List of Figures

---

2.1	Layer structure of a multi-layer perceptron . . . . .	10
2.2	Operation of a single node . . . . .	11
2.3	Evidence incorporating Occam's razor . . . . .	35
3.1	Main steps in the development of an ANN . . . . .	40
3.2	SOM data division . . . . .	48
3.3	Typical activation functions for ANNs . . . . .	61
3.4	Generalising to, overfitting and underfitting the data . . . . .	63
3.5	Examples of different local minima on the error surface . . . . .	68
3.6	The effect of the learning rate used for backpropagation . . . . .	70
3.7	The impact of momentum in backpropagation . . . . .	71
3.8	Indirect route taken towards the minimum of the error surface . . . . .	72
3.9	Probability density of data set I . . . . .	78
3.10	Probability density of data set II . . . . .	79
3.11	Probability density of data set III . . . . .	80
3.12	Schematic of backpropagation . . . . .	82
3.13	Schematic of a genetic algorithm . . . . .	85
3.14	Two child average crossover operator . . . . .	87
3.15	Schematic of the SCE-UA algorithm . . . . .	89
3.16	Example ANN . . . . .	95
3.17	The squashing effect of nonlinear activation functions . . . . .	96
3.18	Error surfaces of 2 and 10 hidden node ANNs . . . . .	107
3.19	In-sample BIC results when training was stopped early and run to convergence . . . . .	112
3.20	Average $RI$ values estimated for data set I . . . . .	120
3.21	Average $RI$ values estimated for data set II . . . . .	122
3.22	Average $RI$ values estimated for data set III . . . . .	124
3.23	Scatter plots of model predictions for data set I . . . . .	127

3.24	Time series plot of model predictions for data set I . . . . .	128
3.25	Scatter plots of model predictions for data set II . . . . .	129
3.26	Time series plot of model predictions for data set II . . . . .	130
3.27	Scatter plots of the 5 hidden node ANN predictions for data set III . . . . .	132
3.28	Scatter plots of the 6 hidden node ANN predictions for data set III . . . . .	132
3.29	5 hidden node ANN predictions for data set III . . . . .	134
4.1	The probability of a jump using the Metropolis Algorithm . . . . .	142
4.2	The effect of the proposal distribution scale used for Metropolis . . . . .	143
4.3	Scale and degrees of freedom parameters of the scaled inverse chi-squared distribution . . . . .	153
4.4	The effect of simulated annealing on the target distribution . . . . .	158
4.5	Posterior hidden-output weight distributions that include zero . . . . .	166
4.6	SSE traces obtained when local and global minimum solutions were found for data set II . . . . .	169
4.7	SSE traces obtained when local and global minimum solutions were found for data set III . . . . .	170
4.8	SSE traces obtained with 1 and 10 hidden node ANNs used to model data set I . . . . .	172
4.9	SSE traces obtained with 3 and 10 hidden node ANNs used to model data set II . . . . .	173
4.10	Mean MCMC traces for the 2, 6, . . . , 10 hidden node ANNs applied to data set II . . . . .	174
4.11	Individual $\log p^*(\mathbf{w} \mathbf{y})$ traces obtained for the 2, 6, . . . , 10 hidden node ANNs applied to data set II . . . . .	175
4.12	Mean MCMC traces resulting from each form of prior distribution when applied to poorly initialised data set II model - unfixed hyperparameters . . . . .	176
4.13	Mean MCMC traces resulting from each form of prior distribution when applied to poorly initialised data set III model - unfixed hyperparameters . . . . .	177
4.14	Individual $\log p^*(\mathbf{w} \mathbf{y})$ traces resulting from each form of prior distribution when applied to poorly initialised data set II model - unfixed hyperparameters . . . . .	178
4.15	Individual $\log p^*(\mathbf{w} \mathbf{y})$ traces resulting from each form of prior distribution when applied to poorly initialised data set III model - unfixed hyperparameters . . . . .	178



---

4.16	Mean MCMC traces resulting from each form of prior distribution when applied to poorly initialised data set II model - fixed hyperparameters . . .	179
4.17	Individual MCMC traces resulting from each form of prior distribution when applied to poorly initialised data set II model - fixed hyperparameters	180
4.18	Mean MCMC traces resulting from each form of prior distribution when applied to poorly initialised data set III model - fixed hyperparameters . . .	181
4.19	Individual MCMC traces resulting from each form of prior distribution when applied to poorly initialised data set III model - fixed hyperparameters	181
4.20	Mean MCMC traces obtained for the data set III model with a longer simulation . . . . .	182
4.21	Mean MCMC traces obtained for the data set II model with poor and good weight initialisations . . . . .	183
4.22	Mean MCMC traces obtained for the data set III model with poor and good weight initialisations . . . . .	183
4.23	Mean MCMC traces obtained using simulated annealing . . . . .	184
4.24	Mean MCMC traces resulting from different prior distributions when the MCMC algorithm was used to train the 10 hidden node data set I model - overfitting prevented . . . . .	185
4.25	Mean MCMC traces resulting from different prior distributions when the MCMC algorithm was used to train the 10 hidden node data set II model - overfitting prevented . . . . .	185
4.26	Mean MCMC traces resulting from each form of prior when applied to the overtrained data set I model . . . . .	186
4.27	Mean MCMC traces resulting from each form of prior when applied to the overtrained data set II model . . . . .	187
4.28	Average magnitudes of different weight groups for the data set I model, given each form of prior . . . . .	188
4.29	Average magnitudes of different weight groups for the data set II model, given each form of prior . . . . .	188
4.30	Evidence estimates for the 1, . . . , 10 hidden node ANNs applied to data set I . . . . .	191
4.31	Marginal posterior hidden-output weight distribution for the 1 hidden node ANN . . . . .	192
4.32	Marginal posterior hidden-output weight distributions for the 2 hidden node ANN . . . . .	193

4.33	Evidence estimates for the 1, . . . , 10 hidden node ANNs applied to data set II . . . . .	194
4.34	Marginal posterior hidden-output weight distributions for the 3 hidden node ANN applied to data set II . . . . .	195
4.35	Marginal posterior hidden-output weight distributions for the 4 hidden node ANN applied to data set II . . . . .	195
4.36	Evidence estimates for the 1, . . . , 10 hidden node ANNs applied to data set III . . . . .	196
4.37	Marginal posterior hidden-output weight distributions for the 5 hidden node ANN applied to data set III . . . . .	197
4.38	Marginal posterior hidden-output weight distributions for the 6 hidden node ANN applied to data set III . . . . .	198
4.39	Marginal posterior hidden-output weight distributions for the 7 hidden node ANN applied to data set III . . . . .	199
4.40	Scatter plots of mean predictions and 95% prediction limits for data set I .	200
4.41	Mean predictions and 95% prediction limits obtained for data set I . . . .	201
4.42	First 100 mean predictions and 95% prediction limits obtained for data set I	202
4.43	Estimated <i>RI</i> distributions for the inputs of data set I . . . . .	202
4.44	Scatter plots of mean predictions and 95% prediction limits for data set II	203
4.45	First 100 mean predictions and 95% prediction limits obtained for data set II	203
4.46	Mean predictions and 95% prediction limits obtained for data set II . . . .	204
4.47	Estimated <i>RI</i> distributions for the inputs of data set II . . . . .	205
4.48	Scatter plots of the 5 hidden node ANN mean predictions and 95% prediction limits for data set III . . . . .	206
4.49	Scatter plots of the 6 hidden node ANN mean predictions and 95% prediction limits for data set III . . . . .	206
4.50	Mean predictions and 95% prediction limits obtained using the 5 hidden node ANN applied to data set III . . . . .	207
4.51	First 100 mean predictions and 95% prediction limits obtained using the 5 hidden node ANN applied to data set III . . . . .	208
4.52	<i>RI</i> distributions for the inputs of data set III estimated using the 5 hidden node ANN weights . . . . .	209
4.53	<i>RI</i> distributions for the inputs of data set III estimated using the 6 hidden node ANN weights . . . . .	210
5.1	Murray-Darling Basin . . . . .	216

5.2	Pipelines delivering River Murray water to South Australia . . . . .	217
5.3	Increasing salinity in the River Murray with distance downstream . . . . .	218
5.4	Economic impacts for a 1 EC unit increase in Morgan salinity . . . . .	219
5.5	Time series of salinity in the River Murray at Murray Bridge . . . . .	221
5.6	Fourier series seasonal mean for salinity and flow . . . . .	222
5.7	Salinity at Mannum with a lag of 1 day ( $MAS_{t-1}$ ) . . . . .	226
5.8	Salinity at Morgan with a lag of 60 days ( $MOS_{t-60}$ ) . . . . .	226
5.9	Salinity at Waikerie with a lag of 1 day ( $WAS_{t-1}$ ) . . . . .	227
5.10	Salinity at Loxton with a lag of 25 days ( $MAS_{t-1}$ ) . . . . .	227
5.11	Flow at lock 7 with a lag of 1 day ( $L7F_{t-1}$ ) . . . . .	228
5.12	River level at Murray Bridge with a lag of 1 day ( $MBL_{t-1}$ ) . . . . .	228
5.13	River level at Mannum with a lag of 57 days ( $MAL_{t-57}$ ) . . . . .	229
5.14	River level upstream of Lock 1 with a lag of 1 day ( $L1UL_{t-1}$ ) . . . . .	229
5.15	Histogram of Murray Bridge salinity data . . . . .	230
5.16	Standardised model residuals . . . . .	234
5.17	BIC and AIC values for 5, 10, . . . , 30 hidden node ANN models . . . . .	235
5.18	BIC and AIC values for 2, 4, 6 and 8 hidden node ANN models . . . . .	236
5.19	BIC and AIC values for 3, 4 and 5 hidden node ANN models . . . . .	237
5.20	Scatter plots of the 4 hidden node ANN model predictions . . . . .	237
5.21	Salinity forecasts obtained for the “model development” data . . . . .	238
5.22	Salinity forecasts obtained for the “real-time forecasting” data . . . . .	240
5.23	Mean MCMC traces for the 2, 4, . . . , 10 hidden node ANNs . . . . .	243
5.24	Individual $\log p^*(\mathbf{w} \mathbf{y})$ traces obtained for the 2, 4, . . . , 10 hidden node ANNs . . . . .	244
5.25	Evidence estimates for the 2, 4, . . . , 10 hidden node ANNs . . . . .	245
5.26	Marginal posterior hidden-output weight distributions of the 4 hidden node ANN . . . . .	246
5.27	Marginal posterior hidden-output weight distributions of the 6 hidden node ANN . . . . .	246
5.28	Scatter plot of hidden-output weights for the 6 hidden node ANN . . . . .	247
5.29	Estimated evidence values including those for the 3 and 5 hidden node ANNs . . . . .	247
5.30	Marginal posterior hidden-output weight distributions for the 3 hidden node ANN . . . . .	248

5.31	Marginal posterior hidden-output weight distributions for the 5 hidden node ANN . . . . .	248
5.32	Scatter plots of mean predictions and 95% prediction limits . . . . .	249
5.33	Mean salinity forecasts and 95% prediction limits obtained for the “model development” data . . . . .	249
5.34	<i>RI</i> distributions for salinity inputs . . . . .	251
5.35	Mean salinity forecasts and 95% prediction limits obtained for the “real-time forecasting” data . . . . .	252
6.1	Regulatory structures in the lower River Murray, South Australia. . . . .	257
6.2	Algal dynamics . . . . .	259
6.3	<i>Anabaena</i> spp. time series . . . . .	262
6.4	Silica versus <i>Anabaena</i> spp. at Morgan. . . . .	264
6.5	Flow into South Australia versus <i>Anabaena</i> spp. at Morgan. . . . .	264
6.6	Temperature versus <i>Anabaena</i> spp. at Morgan. . . . .	265
6.7	pH versus <i>Anabaena</i> spp. at Morgan. . . . .	265
6.8	Probability density of the available <i>Anabaena</i> spp. data . . . . .	267
6.9	Residuals resulting from linearly scaled data . . . . .	269
6.10	Residuals resulting from log transformed data . . . . .	270
6.11	BIC and AIC values for 5, 10, 15 and 20 hidden node ANN models . . . . .	271
6.12	BIC and AIC values for 2, 4, 6 and 8 hidden node ANN models . . . . .	271
6.13	BIC and AIC values for 1, 2, 3, 4 and 5 hidden node ANN models . . . . .	272
6.14	<i>Anabaena</i> spp. forecasts obtained for the entire data . . . . .	273
6.15	<i>Anabaena</i> spp. forecasts obtained for the validation data . . . . .	274
6.16	Mean MCMC traces for the 2, 4, . . . , 10 hidden node ANNs . . . . .	277
6.17	Individual $\log p^*(\mathbf{w} \mathbf{y})$ traces obtained for the 2, 4, . . . , 10 hidden node ANNs . . . . .	278
6.18	Evidence estimates for the 2, 4, . . . , 10 hidden node ANNs . . . . .	279
6.19	Marginal posterior hidden-output weight distributions for the 2 hidden node ANN . . . . .	280
6.20	Scatter plot of hidden-output weights for the 2 hidden node ANN . . . . .	280
6.21	Marginal posterior hidden-output weight distributions for the 4 hidden node ANN . . . . .	281
6.22	Scatter plots of the hidden-output weights for the 4 hidden node ANN . . . . .	282
6.23	Estimated evidence values including that for the 1 hidden node ANN . . . . .	283
6.24	Mean <i>Anabaena</i> forecasts and 95% prediction limits obtained for all data . . . . .	283

6.25 Mean <i>Anabaena</i> forecasts and 95% prediction limits obtained for validation data . . . . .	284
6.26 <i>RI</i> distributions for <i>Anabaena</i> spp. inputs . . . . .	286



# List of Tables

---

3.1	Default parameter values for SCE-UA algorithm . . . . .	91
3.2	User-defined parameters adopted for BP, GA and SCE-UA . . . . .	92
3.3	ANN weights specified in the comparison of input importance measures carried out by <i>Sarle</i> (2002) . . . . .	98
3.4	Limitations of previous studies conducted to compare input importance measures . . . . .	99
3.5	Summary of training set MSE results for data set I . . . . .	103
3.6	Summary of training set MSE results for data set II . . . . .	104
3.7	Summary of training set MSE results for data set III . . . . .	106
3.8	Average training times for the BP, GA and SCE-UA algorithms . . . . .	108
3.9	Optimal number of hidden nodes for data set I . . . . .	108
3.10	Optimal number of hidden nodes for data set II . . . . .	109
3.11	Optimal number of hidden nodes for data set III . . . . .	110
3.12	In-sample AIC and BIC results for data set I . . . . .	110
3.13	In-sample AIC and BIC results for data set II . . . . .	111
3.14	In-sample AIC and BIC results for data set III . . . . .	111
3.15	Out-of-sample results for data set I . . . . .	113
3.16	Out-of-sample results for data set II . . . . .	114
3.17	Out-of-sample results for data set III . . . . .	115
3.18	Validation set results for data set I . . . . .	116
3.19	Validation set results for data set II . . . . .	117
3.20	Validation set results for data set III . . . . .	117
3.21	PMI results for data sets I, II, and III . . . . .	119
3.22	Evaluation of input importance measures for data set I . . . . .	121
3.23	Evaluation of input importance measures for data set II . . . . .	123
3.24	Evaluation of input importance measures for data set III . . . . .	125
3.25	Overall evaluation of input importance measures . . . . .	126
3.26	Performance of best model developed for data set I . . . . .	127

3.27	<i>RI</i> values for the 1 hidden node ANN inputs for data set I . . . . .	129
3.28	Performance of best model developed for data set II . . . . .	131
3.29	<i>RI</i> values for the 3 hidden node ANN inputs for data set II . . . . .	131
3.30	Performance of the 5 hidden node ANN developed for data set III . . . . .	132
3.31	Performance of the 6 hidden node ANN developed for data set III . . . . .	133
3.32	<i>RI</i> values for the 5 and 6 hidden node ANN inputs for data set III . . . . .	133
4.1	Interpretive scale for Bayes factors . . . . .	162
4.2	SSE values obtained for each model developed in the overfitting investigation . . . . .	189
4.3	Evidence estimates for data set I ANN models . . . . .	191
4.4	Log Bayes Factors in favour of the highest ranked model for data set I . . . . .	192
4.5	Evidence estimates for data set II ANN models . . . . .	193
4.6	Log Bayes Factors in favour of the highest ranked model for data set II . . . . .	194
4.7	Evidence estimates for data set III ANN models . . . . .	196
4.8	Log Bayes Factors in favour of the highest ranked model for data set III . . . . .	197
4.9	Mean <i>RI</i> values for the 1 hidden node ANN inputs for data set I . . . . .	202
4.10	Mean <i>RI</i> values for the 3 hidden node ANN inputs for data set II . . . . .	205
4.11	Mean <i>RI</i> values for the 5 and 6 hidden node ANN inputs for data set III . . . . .	209
4.12	Mean performance of the ANNs developed using Bayesian methods in comparison to the performance of the corresponding deterministic ANNs . . . . .	211
5.1	Available data for salinity forecasting model . . . . .	224
5.2	Inputs used in salinity forecasting ANN model . . . . .	225
5.3	Modified parameter values for SCE-UA algorithm . . . . .	232
5.4	Training times using the SCE-UA algorithm . . . . .	234
5.5	Model-based <i>RI</i> estimates and order of input importance . . . . .	239
5.6	Log Bayes Factors in favour of the highest ranked model . . . . .	245
5.8	Performance of the 4 hidden node ANN developed using deterministic and Bayesian methods . . . . .	253
6.1	National cyanobacterial alert levels . . . . .	258
6.2	Available data for <i>Anabaena</i> forecasting model . . . . .	261
6.3	Inputs used in <i>Anabaena</i> forecasting ANN model . . . . .	263
6.4	RMSEs (log scale) for the 1 hidden node ANN developed using deterministic methods . . . . .	273



---

6.5	Model-based <i>RI</i> estimates and order of input importance in comparison to the PMI-based <i>RI</i> estimates . . . . .	274
6.6	Log Bayes Factors in favour of the highest ranked model . . . . .	279
6.7	RMSEs (log scale) for the 1 hidden node ANN developed using deterministic methods . . . . .	285
A.1	MSE results for data set I trained with BP . . . . .	328
A.2	MSE results for data set I trained with the GA . . . . .	328
A.3	MSE results for data set I trained with the SCE-UA algorithm . . . . .	328
A.4	MSE results for data set II trained with BP . . . . .	329
A.5	MSE results for data set II trained with the GA . . . . .	329
A.6	MSE results for data set II trained with the SCE-UA algorithm . . . . .	329
A.7	MSE results for data set III trained with BP . . . . .	330
A.8	MSE results for data set III trained with the GA . . . . .	330
A.9	MSE results for data set III trained with the SCE-UA algorithm . . . . .	330
B.1	<i>RI</i> results for data set I trained by BP . . . . .	332
B.2	<i>RI</i> results for data set I trained by GA . . . . .	333
B.3	<i>RI</i> results for data set I trained by SCE-UA . . . . .	334
B.4	<i>RI</i> results for data set II trained by BP . . . . .	335
B.5	<i>RI</i> results for data set II trained by GA . . . . .	336
B.6	<i>RI</i> results for data set II trained by SCE-UA . . . . .	337
B.7	<i>RI</i> results for data set III trained by BP . . . . .	338
B.8	<i>RI</i> results for data set III trained by GA . . . . .	339
B.9	<i>RI</i> results for data set III trained by SCE-UA . . . . .	340



# Abstract

---

A new Bayesian framework for training and selecting the complexity of artificial neural networks (ANNs) is developed in this thesis, based on Markov chain Monte Carlo (MCMC) techniques. The primary motivation of the research presented is the incorporation of uncertainty into ANNs used for water resources modelling, with emphasis placed on obtaining accurate results, while maintaining simplicity of implementation, which is considered to be of utmost importance for adoption of the framework by practitioners in this field. By applying the Bayesian framework to a number of synthetic and real-world case studies and by comparison with a state-of-the-art ANN development approach, it is shown throughout this thesis how the Bayesian approach can be used to address the three most significant issues facing the wider acceptance of ANNs in this field; namely generalisability, interpretability and uncertainty. The state-of-the-art approach is devised through reviewing and, where necessary, improving current best practice deterministic ANN development methods, leading to the recommended use of the global SCE-UA optimisation algorithm, which has not been used before for ANN training, and the development of a modified connection weight approach for extracting knowledge from trained ANNs. The real-world case studies used in this research, which involve salinity forecasting in the River Murray at Murray Bridge, South Australia, and the forecasting of cyanobacteria (*Anabaena* spp.) in the River Murray at Morgan, South Australia, are used to demonstrate the practical value of the Bayesian framework, particularly when extrapolation is required and when the available data are of poor quality. These issues lead to poor model performance when deterministic ANN development methods are applied, yet as the generated predictions are deterministic, there is no direct way of assessing their quality. Application of the proposed Bayesian framework leads to better average performance of the ANN models developed, since a minimal ANN structure is selected and a more generalised input-output mapping is obtained. More importantly, prediction limits are provided which quantify the uncertainty in the predictions and enable management and design decisions to be made based on a known level of confidence.



# Statement of Originality

---

*I Greer B. Kingston hereby declare that this work contains no material that has been accepted for the award of any other degree or diploma in any university or other tertiary institution. To the best of my knowledge and belief, it contains no material previously published or written by any other person, except where due reference is made in the text.*

*I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.*

*SIGNED: ..... DATE: .....*



# Acknowledgements

---

*Firstly, I would like to express my sincere gratitude to my principal supervisor Associate Professor Holger Maier, whose continual encouragement, direction and constructive and timely comments helped me to complete this thesis and were always greatly appreciated throughout the duration of my PhD. Secondly, I would like to thank my co-supervisor Associate Professor Martin Lambert who helped to shape this project by introducing me to various algorithms and Bayesian procedures and whose assistance in my understanding of these methods is kindly acknowledged. I would also like to thank Dr Gavin Bowden for providing the code used to implement a number of the deterministic ANN methods used to complete this work.*

*Thanks must also go to my fellow postgraduates in the School of Civil and Environmental Engineering for providing such a friendly and supportive working environment and for the de-stressing ritual that is “Friday drinks”. Thankyou also to the support staff in the school for their assistance throughout my time here.*

*A special thanks to my friends and family for their encouragement and belief in me, particularly my mother Kathy and my sister Lauren, whose various forms of support helped me to complete this PhD.*

*Finally, the financial support for this project, provided by an Australian Research Council (ARC) Discovery grant together with an Australian Postgraduate Award (APA), is gratefully acknowledged.*





# List of Publications

---

The following publications are related to the research presented in this thesis:

## Journal Papers:

Kingston, G. B., M. F. Lambert, and H. R. Maier (2005), Bayesian training of artificial neural networks used for water resources modeling, *Water Resources Research*, 41(12), W12409, doi:10.1029/2005WR004152.

Kingston, G. B., H. R. Maier, and M. F. Lambert (2005), Calibration and validation of neural networks to ensure physically plausible hydrological modeling, *Journal of Hydrology*, 314(1–4), 158–176, doi:10.1016/j.jhydrol.2005.03.013.

Kingston, G. B., H. R. Maier, and M. F. Lambert (2005), A probabilistic method to assist knowledge extraction from artificial neural networks used for hydrological prediction, *Mathematical and Computer Modelling*, *In press*, doi:10.1016/j.mcm.2006.01.008.

## Conference Papers:

Kingston, G. B., H. R. Maier, and M. F. Lambert (2006), Forecasting cyanobacteria with Bayesian and deterministic artificial neural networks, accepted for publication in *Proceedings of the IEEE World Congress on Computational Intelligence (WCCI 2006)*, Vancouver, Canada, July 2006.

Kingston, G. B., H. R. Maier, and M. F. Lambert (2005), A Bayesian approach to artificial neural network model selection, in Zerger, A. and Argent, R. M. (eds), *Proceedings of the MODSIM 2005 International Congress on Modelling and Simulation*, pp. 1853–1859. Melbourne, Australia, December 2005.  
<http://www.mssanz.org.au/modsim05/papers/kingston.pdf>

Kingston, G. B., H. R. Maier, and M. F. Lambert (2005), A Bayesian method to improve the extrapolation ability of ANNs, in Hamza, M. H. (ed), *Proceedings of the IASTED International Conference on Applied Simulation and Modelling*, no. 469–126. Benalmádena, Spain, June 2005.

Kingston, G. B., H. R. Maier, and M. F. Lambert (2004), A statistical input pruning method for artificial neural networks used in environmental modelling, in Pahl-Wostl, C., Schmidt, S., Rizzoli, A. E. and Jakeman, A. J. (eds), *Complexity and Integrated Resources Management, Transactions of the 2nd Biennial Meeting of the International*

*Environmental Modelling and Software Society*, Volume 1, pp. 87–92. iEMSs, June 2004.

Kingston, G. B., T. M. Heneker, M. F. Lambert, and H. R. Maier (2003), A comparison of conceptual and empirical modelling methods for simulation of catchment runoff, in *Proceedings of the 28th Hydrology and Water Resources Symposium*, Volume 2, pp. 299–306. Wollongong, Australia, November 2003.

Kingston, G. B., H. R. Maier, and M. F. Lambert (2003), Understanding the mechanisms modelled by artificial neural networks for hydrological prediction, in Post, D. A. (ed), *Proceedings of the MODSIM 2003 International Congress on Modelling and Simulation*, Volume 2, pp. 825–830. Townsville, Australia, July 2003.

Kingston, G. B., M. F. Lambert, and H. R. Maier, (2003), Development of stochastic artificial neural networks for hydrological prediction, in Post, D. A. (ed), *Proceedings of the MODSIM 2003 International Congress on Modelling and Simulation*, Volume 2, pp. 837–842. Townsville, Australia, July 2003.

# Nomenclature & Abbreviations

---

Symbol	Description
<i>General</i>	
$\mathfrak{R}^d$	$d$ -dimensional set of all real numbers
$\Theta$	search space; $\Theta \in \mathfrak{R}^d$
$\mathbf{I}$	identity matrix
$\mathbf{H}$	Hessian matrix
$\mu$	scalar mean
$\sigma^2$	variance ( $\sigma$ is the standard deviation)
$\Sigma$	covariance matrix
$\lambda$	signal-to-noise ratio
$r^2$	coefficient of determination
<i>Probability Distribution Notation</i>	
$N(\mu, \sigma^2)$	Normal distribution with mean $\mu$ and variance $\sigma^2$
$U(a, b)$	Uniform distribution with boundaries $a, b$ , where $b > a$
$\text{Inv-}\chi^2(\nu, S)$	Inverse chi-square distribution with $\nu$ degrees of freedom and scale $S$
<i>ANN and Modelling Notation</i>	
$f(\cdot)$	function modelled by an ANN
$K$	number of inputs
$J$	number of hidden layer nodes
$M$	number of outputs
$I_k$	$k$ th input node
$H_j$	$j$ th hidden node
$O_m$	$m$ th output node
$w_i$	$i$ th “true” weight
$\hat{w}_i$	$i$ th estimated weight
$\mathbf{w}$	“true” vector of connection and bias weights $\equiv (w_1, \dots, w_d)$
$\hat{\mathbf{w}}$	estimated vector of connection and bias weights $\equiv (\hat{w}_1, \dots, \hat{w}_d)$

*continued on next page*

<b>Symbol</b>	<b>Description</b>
$d$	dimension of weight vector
$g(\cdot)$	activation function
$zin_j$	summed input into hidden node $j$
$z_j$	output from hidden node $j \equiv g(zin_j)$
$\hat{y}in$	summed input into output node
$\hat{y}_m$	output from output node $m \equiv g(\hat{y}in_m)$
$E_y$	error, or objective, function
$E_w$	penalty term used to regularise weights
$\nabla E$	gradient of error function
$\epsilon$	model residuals
$\sigma_y^2$	scale (variance) of model residuals
$\sigma_w^2$	scale (variance) of weights
$\hat{\sigma}_y^2$	scale of model residuals at optimum of $E_y$
$\hat{\sigma}_w^2$	scale of weights at optimum of $E_y$
 <i>Data Notation</i>	
$N$	number of samples in the data set
$y$	general term for observed target data
$y_i$	$i$ th observed target data scalar
$\mathbf{y}$	general term for set of scalar target data $\equiv (y_1, \dots, y_N)$
$\mathbf{y}^M$	general term for observed target data vector $\equiv (y_1, \dots, y_M)$
$\mathbf{y}_i^M$ or $\mathbf{y}_i$	$i$ th target vector $\equiv (y_{1,i}, \dots, y_{M,i})$
$\mathbf{y}_m$	$m$ th target variable $\equiv (y_{m,1}, \dots, y_{m,N})$
$\mathbf{Y}$	general term for set of target vectors $\equiv (\mathbf{y}_1^M, \dots, \mathbf{y}_N^M)$
$\hat{y}_i$	$i$ th predicted data scalar
$\hat{\mathbf{y}}$	general term for set of scalar predicted data $\equiv (\hat{y}_1, \dots, \hat{y}_N)$
$\hat{\mathbf{y}}^M$	general term for predicted data vector $\equiv (\hat{y}_1, \dots, \hat{y}_M)$
$\hat{\mathbf{y}}_i^M$ or $\hat{\mathbf{y}}_i$	$i$ th predicted output vector $\equiv (\hat{y}_{1,i}, \dots, \hat{y}_{M,i})$
$\hat{\mathbf{y}}_m$	$m$ th predicted output variable $\equiv (\hat{y}_{m,1}, \dots, \hat{y}_{m,N})$
$\mathbf{x}$	general term for input variable $\equiv (x_1, \dots, x_N)$
$\mathbf{x}^K$	general term for input vector $\equiv (x_1, \dots, x_K)$
$\mathbf{x}_i^K$ or $\mathbf{x}_i$	$i$ th input vector $\equiv (x_{1,i}, \dots, x_{K,i})$
$\mathbf{x}_k$	$k$ th input variable $\equiv (x_{k,1}, \dots, x_{k,N})$
$\mathbf{X}$	general term for set of input vectors $\equiv (\mathbf{x}_1^K, \dots, \mathbf{x}_N^K)$
$\mathcal{D}$	set of data pairs $\equiv [(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)]$

*continued on next page*

<b>Symbol</b>	<b>Description</b>
<i>Probability Notation</i>	
$\theta$	general term for model parameters
$\mathcal{H}$	general term for model structure, including implicit assumptions
$p(\theta)$	probability density of model parameters $\theta$ , also known as the prior probability density
$p(\theta, \mathbf{y})$	joint probability of the model parameters $\theta$ and the data $\mathbf{y}$
$p(\theta \mathbf{y})$	conditional probability of $\theta$ , given the data $\mathbf{y}$ , also known as the posterior probability density of $\theta$
$p^*(\theta \mathbf{y})$	unnormalised posterior density
$p(\mathbf{y} \theta)$	conditional probability of the data $\mathbf{y}$ , given the model parameters $\theta$ , also known as the likelihood function
$L(\theta)$	the likelihood function, as above
$p(\mathbf{y} \mathcal{H})$	conditional probability of the data $\mathbf{y}$ , given the model $\mathcal{H}$ , also known as the marginal likelihood or evidence
$\hat{p}(\mathbf{y} \mathcal{H})$	approximate evidence
$p(\mathcal{H} \mathbf{y})$	conditional probability of the model $\mathcal{H}$ , given the model $\mathbf{y}$ , also known as the posterior probability density of $\mathcal{H}$
<i>Training Notation</i>	
<b>Backpropagation (BP)</b>	
$\gamma$	stepsize of gradient descent
$\mathbf{d}$	direction of gradient descent
$\eta$	learning rate
$\phi$	momentum rate
$\Delta\mathbf{w}$	weight increment
$\delta$	delta function
$\kappa$	epoch size
<b>Genetic Algorithm (GA)</b>	
$G$	population of chromosomes
$s$	population size
$\rho_{cross}$	crossover rate
$\rho_{mut}$	mutation rate
$\tau$	mutation stepsize
$gene'$	mutated gene value
<b>Shuffled Complex Evolution (SCE-UA)</b>	
$p$	number of complexes
$m$	number of points in a complex

*continued on next page*

<b>Symbol</b>	<b>Description</b>
$s$	population size = $m \times p$
$q$	number of points in a subcomplex
$\alpha$	number of offspring generated by a subcomplex
$\beta$	number of evolution steps taken by each complex
<b>Markov chain Monte Carlo (MCMC)</b>	
$T(\cdot)$	transition distribution
$Q(\cdot)$	proposal density
$\alpha(\cdot)$	acceptance probability distribution
$\theta^*$	candidate parameter state
$c$	adaptive scaling parameter
$T$	temperature used for simulated annealing
$\varphi$	simulated annealing schedule parameter
$t_0$	number of iterations for which $\Sigma$ is held constant
$t_{\sigma_0^2}$	number of iterations for which hyperparameters are held constant
$t_b$	number of burn-in iterations

<b>Abbreviation</b>	<b>Description</b>
AM	Adaptive Metropolis
AIC	Akaike's information criterion
ANN	Artificial neural network
ARD	Automatic relevance determination
AWQC	Australian Water Quality Centre
BF	Bayes factor
BIC	Bayesian information criterion
BMS	Bayesian model selection
BP	Backpropagation
CCE	Competitive complex evolution
CDF	Cumulative distribution function
CE	Coefficient of efficiency
C-J	Chib-Jeliazkov
DWR	South Australian Department for Water Resources
EBMLP	Evolutionary backpropagation multi-layer perceptron
EP	Evolutionary programming
GA	Genetic algorithm
G-D	Gelfand-Dey

*continued on next page*

<b>Abbreviation</b>	<b>Description</b>
GLUE	Generalized likelihood uncertainty estimation
GRNN	General regression neural network
HMC	Hybrid Monte Carlo
L1LF	Lock 1 lower flow
L1LL	Lock 1 lower river level
L1UL	Lock 1 upper river level
L7F	Lock 7 flow
LOL	Loxton river level
LOS	Loxton salinity
MAE	Mean absolute error
MAL	Mannum river level
MAS	Mannum salinity
MBL	Murray Bridge river level
MBS	Murray Bridge salinity
MCMC	Markov chain Monte Carlo
MDB	Murray-Darling Basin
MDBC	Murray-Darling Basin Commission
MDBMC	Murray-Darling Basin Ministerial Council
MI	Mutual information
MLP	Multi-layer perceptron
MOL	Morgan river level
MOS	Morgan salinity
MSE	Mean squared error
MSRE	Mean squared relative error
MT	Model tree
NGO	NeuroGenetic Optimizer
OCF	Overland Corner flow
OCL	Overland Corner river level
<i>OCW</i>	Overall connection weight
PC	Principle component
PCA	Principle component analysis
PDF	Probability density function
PMI	Partial mutual information
<i>RI</i>	Relative importance
RMSE	Root mean squared error
SCE-UA	Shuffled complex evolution - University of Arizona
SOM	Self-organising map

*continued on next page*

<b>Abbreviation</b>	<b>Description</b>
SSE	Sum squared error
SVM	Support vector machine
WAL	Waikerie river level
WAS	Waikerie salinity