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
THE UNIVERSITY OF ADELAIDE
AUSTRALIA

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
Email:	enquiries@civeng.adelaide.edu.au

Electronic File Created: May 18, 2006, 9:08am

Table of Contents

List of Figures

List of Tables

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**
General	
$\Re^d$	d-dimensional set of all real numbers
Θ	search space; $\Theta \in \Re^d$
$\mathbf{I}$	identity matrix
$\mathbf{H}$	Hessian matrix
μ	scalar mean
σ^2	variance (σ is the standard deviation)
Σ	covariance matrix
λ	signal-to-noise ratio
r^2	coefficient of determination
Probability Distribution Notation	
$N(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
$U(a, b)$	Uniform distribution with boundaries a, b, where $b > a$
Inv-$\chi^2(\nu, S)$	Inverse chi-square distribution with ν degrees of freedom and scale S
ANN and Modelling Notation	
$f(\cdot)$	function modelled by an ANN
K	number of inputs
J	number of hidden layer nodes
M	number of outputs
I_k	kth input node
H_j	jth hidden node
O_m	mth output node
w_i	ith "true" weight
$\hat{w}_i$	ith estimated weight
$\mathbf{w}$	"true" vector of connection and bias weights $\equiv (w_1, \ldots, w_d)$
$\hat{\mathbf{w}}$	estimated vector of connection and bias weights $\equiv (\hat{w}_1, \ldots, \hat{w}_d)$

continued on next page

Symbol	**Description**
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 $m \equiv g(\hat{y}in_m)$
$E_{\mathbf{y}}$	error, or objective, function
$E_{\mathbf{w}}$	penalty term used to regularise weights
∇E	gradient of error function
ϵ	model residuals
$\sigma^2_{\mathbf{y}}$	scale (variance) of model residuals
$\sigma^2_{\mathbf{w}}$	scale (variance) of weights
$\hat{\sigma}^2_{\mathbf{y}}$	scale of model residuals at optimum of $E_{\mathbf{y}}$
$\hat{\sigma}^2_{\mathbf{w}}$	scale of weights at optimum of $E_{\mathbf{y}}$
Data Notation	
N	number of samples in the data set
y	general term for observed target data
y_i	ith observed target data scalar
$\mathbf{y}$	general term for set of scalar target data $\equiv (y_1, \ldots, y_N)$
$\mathbf{y}^M$	general term for observed target data vector $\equiv (y_1, \ldots, y_M)$
$\mathbf{y}^M_i$ or $\mathbf{y}_i$	ith target vector $\equiv (y_{1,i}, \ldots, y_{M,i})$
$\mathbf{y}_m$	mth target variable $\equiv (y_{m,1}, \ldots, y_{m,N})$
$\mathbf{Y}$	general term for set of target vectors $\equiv (\mathbf{y}^M_1, \ldots, \mathbf{y}^M_N)$
$\hat{y}_i$	ith predicted data scalar
$\hat{\mathbf{y}}$	general term for set of scalar predicted data $\equiv (\hat{y}_1, \ldots, \hat{y}_N)$
$\hat{\mathbf{y}}^M$	general term for predicted data vector $\equiv (\hat{y}_1, \ldots, \hat{y}_M)$
$\hat{\mathbf{y}}^M_i$ or $\hat{\mathbf{y}}_i$	ith predicted output vector $\equiv (\hat{y}_{1,i}, \ldots, \hat{y}_{M,i})$
$\hat{\mathbf{y}}_m$	mth predicted output variable $\equiv (\hat{y}_{m,1}, \ldots, \hat{y}_{m,N})$
$\mathbf{x}$	general term for input variable $\equiv (x_1, \ldots, x_N)$
$\mathbf{x}^K$	general term for input vector $\equiv (x_1, \ldots, x_K)$
$\mathbf{x}^K_i$ or $\mathbf{x}_i$	ith input vector $\equiv (x_{1,i}, \ldots, x_{K,i})$
$\mathbf{x}_k$	kth input variable $\equiv (x_{k,1}, \ldots, x_{k,N})$
$\mathbf{X}$	general term for set of input vectors $\equiv (\mathbf{x}^K_1, \ldots, \mathbf{x}^K_N)$
$\mathcal{D}$	set of data pairs $\equiv [(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)]$

continued on next page

Symbol	**Description**
Probability Notation	
θ	general term for model parameters
$\mathcal{H}$	general term for model structure, including implicit assumptions
$p(\theta)$	probability density of model parameters θ, also known as the prior probability density
$p(\theta, \mathbf{y})$	joint probability of the model parameters θ and the data $\mathbf{y}$
$p(\theta\|\mathbf{y})$	conditional probability of θ, given the data $\mathbf{y}$, also known as the posterior probability density of θ
$p^*(\theta\|\mathbf{y})$	unnormalised posterior density
$p(\mathbf{y}\|\theta)$	conditional probability of the data $\mathbf{y}$, given the model parameters θ, 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 $\mathcal{H}$
Training Notation	
Backpropagation (BP)	
γ	stepsize of gradient descent
$\mathbf{d}$	direction of gradient descent
η	learning rate
ϕ	momentum rate
$\Delta\mathbf{w}$	weight increment
δ	delta function
κ	epoch size
Genetic Algorithm (GA)	
G	population of chromosomes
s	population size
ρ_{cross}	crossover rate
ρ_{mut}	mutation rate
τ	mutation stepsize
$gene'$	mutated gene value
Shuffled Complex Evolution (SCE-UA)	
p	number of complexes
m	number of points in a complex

continued on next page

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

Abbreviation	Description
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

Abbreviation	**Description**
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
OCW	Overall connection weight
PC	Principle component
PCA	Principle component analysis
PDF	Probability density function
PMI	Partial mutual information
RI	Relative importance
RMSE	Root mean squared error
SCE-UA	Shuffled complex evolution - University of Arizona
SOM	Self-organising map

continued on next page

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