

***Modelling Long-Term Persistence  
in Hydrological Time Series***

Mark Andrew Thyer B.E. (Civil) (Hons. I)

*A thesis submitted for the degree of Doctor of Philosophy  
at The University of Newcastle*

December 2000

*I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.*

---

Mark A. Thyer

*“If we knew what we were doing  
it would not be called research”*

*Albert Einstein*

---

# ***Acknowledgements***

The completion of a PhD thesis represents a major undertaking and it would not have been possible without the support and advice from my family, friends and colleagues.

It was been an absolute pleasure to have Associate Professor George Kuczera as my supervisor. His enthusiasm and wisdom are a continual source of inspiration. Countless times I have walked into George's office with a seemingly insurmountable problem and then emerged (usually hours later) fully motivated to tackle this thesis.

I would like to gratefully acknowledge the financial support provided by the industry sponsors of this project, Sydney Catchment Authority and Hunter Water Corporation. Personal thanks go to Mark Powell from SCA for answering my many questions and to Janice Lough from AIMS for providing the Burdekin data.

Special thanks go to my friends, Craig Gardiner - my SIB brewing partner, Kerry McIntosh, Greg Hancock - the Dad of the postgrads, Jeff (and Wendy) Walker for finding some desk space at NASA and Emily Slatter for providing me with a home during the throes of writing up this thesis. The times I have shared with these good people have made my stay in Newcastle very memorable.

Thank you to the postgrads and academics, Andrew Frost, Lyndon Bell, Scott Wooldridge, Heber Sugo and Dr Stewart Franks for helpful advice and fun distractions. Thank you also to the support staff of my Department, they do an excellent job.

Jodie Berkefeld deserves a special mention. She was my friend, my housemate, my girlfriend and now my friend again. I think she deserves a PhD for putting up with me over the last 3 years and 9 months. Thank you for everything, Jodie.

I cannot express my appreciation for the continual love and support provided by my family. To my parents, Alan and Tricia, and my sister, Jo-Anne, thank you for always believing in me, I am forever grateful.

And finally, thank you to all my friends. You make my life what it is, an enjoyable journey.

---

---

# ***Table of Contents***

<b>Abstract</b>	.....	<b>ix</b>
<b>Notation</b>	.....	<b>x</b>
<b>Chapter 1 - Introduction</b>		
1.1	General Introduction.....	1-1
1.2	Stochastic Modelling in Hydrology .....	1-2
1.3	Objectives and Scope .....	1-2
1.4	Overview .....	1-3
<b>Chapter 2 - Background</b>		
2.1	Introduction.....	2-1
2.2	Stochastic Modelling Approaches for Long-term Hydrological Time Series Simulation.....	2-2
2.3	Physical Processes that Influence Long-term Hydrological Time Series .....	2-9
2.3.1	Ocean-atmosphere circulations .....	2-9
2.3.2	Land-atmosphere interactions .....	2-14
2.3.3	Preliminary evidence of two-state long-term persistence in the Australian climatic regime .....	2-15
2.3.4	Palaeoclimatic evidence of two-state long-term persistence.....	2-18
2.4	Critical Evaluation of the Conceptual Basis of the Common Stochastic Modelling Approaches.....	2-21
2.5	Conclusion.....	2-22
<b>Chapter 3 - A New Conceptual Framework</b>		
3.1	Introduction.....	3-1
3.2	The Hidden State Markov Model.....	3-2
3.2.1	The single site framework .....	3-4
3.2.2	The multi-site framework .....	3-4
3.3	Previous Applications of Hidden State Markov Models.....	3-6
3.4	Conclusion.....	3-9

---

**Chapter 4 - Hydrological Data**

4.1	Introduction.....	4-1
4.2	Australia's Climatology .....	4-2
4.3	Selected Australian Capital Cities Rainfall Data .....	4-4
4.3.1	Sydney metropolitan data (066062) .....	4-5
4.3.2	Brisbane metropolitan data (040214) .....	4-6
4.3.3	Melbourne metropolitan data (086071).....	4-6
4.3.4	Adelaide metropolitan data (023000).....	4-7
4.3.5	Perth metropolitan data (009034).....	4-7
4.4	Burdekin River Reconstructed Runoff Data .....	4-9
4.5	Warragamba Catchment Rainfall Data .....	4-10
4.5.1	Selection of sites.....	4-10
4.5.2	Filling-in of missing data.....	4-15
4.5.3	Identifying and correcting inconsistent data.....	4-15
4.6	Williams River Catchment Rainfall Data.....	4-16
4.7	Conclusion.....	4-21

**Chapter 5 - Calibration Procedure for Single Site HSM Model**

5.1	Introduction.....	5-1
5.2	Development of Calibration Procedure.....	5-2
5.2.1	Bayesian inference background.....	5-2
5.2.2	Preliminary attempts to calibrate the HSM Model.....	5-4
5.2.3	Markov chain Monte Carlo simulation and the Gibbs sampler.....	5-6
5.2.4	Applying the Gibbs sampler to the HSM model.....	5-7
5.2.4.a	Specification of priors.....	5-9
5.2.4.b	Initialising the parameter vector .....	5-11
5.2.4.c	Assessing convergence .....	5-12
5.2.4.d	Other implementation issues.....	5-13
5.2.5	Verification of calibration procedure .....	5-14
5.2.6	Applying the Metropolis algorithm to the HSM model.....	5-14
5.3	Conclusion.....	5-15

---

## Chapter 6 - Developing an Interpretative Framework for HSM Model using Synthetic Data

6.1	Introduction.....	6-1
6.2	Motivation.....	6-2
6.3	Different Aspects of Two-State Persistence.....	6-2
6.3.1	Strength of the wet and dry state persistence structure .....	6-2
6.3.2	Length of the time series .....	6-3
6.3.3	Separation of the wet and dry state distributions.....	6-5
6.3.4	Two–state persistence structure notation.....	6-6
6.4	Methods Developed for Analysing the Calibration Results.....	6-7
6.4.1	Verifying the HSM modeling assumptions .....	6-7
6.4.2	Selecting the appropriate water year .....	6-10
6.5	Synthetic Data Scenarios.....	6-11
6.5.1	Independent data scenario .....	6-13
6.5.2	$TSP = [M, M, 1.0]$ scenario.....	6-16
6.5.2.a	140 year series.....	6-16
6.5.2.b	100 and 70 year series.....	6-21
6.5.3	$TSP = [W, S, 1.0]$ scenario.....	6-23
6.5.4	Discussion of other scenarios .....	6-27
6.6	An Interpretative Framework.....	6-28
6.7	Conclusion.....	6-30

## Chapter 7 - Single Site HSM Model Calibration Results

7.1	Introduction.....	7-1
7.2	Analysis of Calibration Results.....	7-2
7.2.1	Australian capital cities rainfall data .....	7-2
7.2.1.a	Sydney.....	7-2
7.2.1.b	Brisbane .....	7-3
7.2.1.c	Melbourne .....	7-4
7.2.1.d	Adelaide .....	7-5
7.2.1.e	Perth.....	7-6
7.2.2	Burdekin River runoff data.....	7-9
7.2.3	Warragamba catchment rainfall data.....	7-12

7.2.3.a	Mt. Victoria composite, Yarra composite and Cataract Dam .....	7-13
7.2.3.b	Moss Vale .....	7-13
7.2.3.c	Taralga .....	7-13
7.2.4	Williams River catchment rainfall data.....	7-17
7.2.4.a	Clarence Town, Dungog and Raymond Terrace.....	7-17
7.2.4.b	Stroud and Branxton.....	7-18
7.3	Discussion of Results .....	7-22
7.3.1	Selection of appropriate water year.....	7-26
7.4	Conclusion.....	7-28

## **Chapter 8 - AR(1) Model Analysis**

8.1	Introduction.....	8-1
8.2	The AR(1) Model.....	8-2
8.3	Development of Calibration Procedure.....	8-3
8.3.1	The Metropolis algorithm.....	8-3
8.3.1.a	Implementation.....	8-4
8.3.2	Verification of calibration procedure .....	8-6
8.4	Analysis of Calibration Results.....	8-9
8.5	Comparison to HSM Model Calibration Results .....	8-12
8.6	Conclusion.....	8-13

## **Chapter 9 - Simulations and Implications: Discussion of Single Site**

### **Calibration Results**

9.1	Introduction.....	9-1
9.2	Comparison of AR(1) and HSM Model.....	9-2
9.2.1	Simulation of long-term rainfall distributions .....	9-2
9.2.2	Comparison of autocorrelation function.....	9-9
9.2.3	Evaluation of hydrological insight .....	9-13
9.3	Further Implications from HSM Model.....	9-14
9.4	Other HSM Modelling Issues.....	9-17
9.5	Future Research Directions for Single Site HSM Model.....	9-17
9.5.1	Developing a linkage to the climatic mechanisms .....	9-18
9.5.2	Modification of HSM modelling approach.....	9-19
9.5.3	Further calibration to sites around Australia .....	9-21



9.5.4	Adaptation to a seasonal HSM model.....	9-22
9.6	Motivation for Multi-site HSM Modelling Framework.....	9-23
9.7	Conclusion.....	9-25

## **Chapter 10 - Calibration Procedure for Multi-site HSM Model**

10.1	Introduction.....	10-1
10.2	Development of Calibration Procedure.....	10-2
10.2.1	Applying the Gibbs sampler to the multi-site HSM model.....	10-2
10.2.1.a	Specification of priors for the state rainfall parameters.....	10-3
10.2.1.b	Sampling distribution for the missing data values.....	10-7
10.2.1.c	Initialising the parameter vector .....	10-8
10.2.1.d	Assessing convergence .....	10-8
10.2.1.e	Other implementation issues.....	10-9
10.2.2	Verification of calibration procedure .....	10-9
10.2.3	Development of P3 prior specification.....	10-10
10.2.4	Further investigation of influence of spatial correlation between sites .....	10-14
10.2.5	Preliminary guidelines for applying Gibbs sampler to the multi-site HSM model.....	10-17
10.3	Conclusion.....	10-19

## **Chapter 11 - Multi-site HSM Model Calibration Results**

11.1	Introduction.....	11-1
11.2	Analysis of Calibration Results.....	11-2
11.2.1	Warragamba catchment rainfall data.....	11-2
11.2.1.a	Four-site analysis .....	11-2
11.2.1.b	Five-site analysis.....	11-3
11.2.2	Williams River catchment rainfall data.....	11-6
11.3	Discussion of Multi-site Results .....	11-9
11.3.1	Comparison to single site results - the influence of multi-site data .....	11-9
11.3.2	Selection of appropriate water year.....	11-13
11.3.3	Identification of a homogenous persistence region.....	11-14
11.3.4	Influence of sampling for missing data .....	11-15
11.3.5	Implications of calibration results.....	11-17

11.4	Future Research Directions .....	11-20
11.4.1	A methodology for identifying a homogenous persistence region.....	11-20
11.4.2	Alternative multi-HSM model frameworks.....	11-20
11.4.3	Further synthetic calibration runs .....	11-21
11.5	Conclusion.....	11-22
 <b>Chapter 12 - Conclusions</b>		
12.1	Introduction.....	12-1
12.2	Summary and Conclusions.....	12-2
12.2.1	Motivation.....	12-2
12.2.2	A new conceptual framework.....	12-2
12.2.3	Development of single site calibration procedure and an interpretative framework for the calibration results .....	12-2
12.2.4	Insight from calibrating the single site HSM model to real data.....	12-3
12.2.5	Comparison to the AR(1) model.....	12-5
12.2.6	Application to water resources management.....	12-6
12.2.7	Development of a multi-site calibration procedure and calibration results.....	12-7
12.3	Future Research Directions .....	12-8
12.3.1	Single site HSM model.....	12-8
12.3.2	Multi-site HSM model.....	12-9
12.4	Concluding Remarks.....	12-9
 <b>References .....</b>		<b>R-1</b>
 <b>Appendix A - Estimation of Hurst Coefficient</b>		
A.1	Introduction.....	A-1
A.1.1	Rescaled range statistic.....	A-2
A.1.2	Aggregated variance method.....	A-2
A.1.3	Differenced variance method .....	A-3
A.2	Summary.....	A-4
 <b>Appendix B - Filling-In and Correction of Hydrological Data</b>		
B.1	Filling-in of Missing Data.....	B-1
B.2	Correction of Inconsistencies in Monthly Data.....	B-2

---

## Appendix C - Markov Chain Monte Carlo Methods: An Overview

C.1	Introduction.....	C-1
C.2	Markov Chain Monte Carlo Simulation.....	C-2
C.3	The Metropolis-Hastings Algorithm.....	C-5
C.4	The Metropolis Algorithm.....	C-8
C.5	The Gibbs Sampler.....	C-9
C.6	Summary.....	C-13

## Appendix D - Sampling Distributions for the Gibbs Sampler

D.1	Introduction.....	D-1
D.2	Sampling the Hidden State Time Series.....	D-2
D.3	Derivation of the Likelihood Function for the HSM model.....	D-4
D.4	Sampling the State Transition Probabilities.....	D-6
D.5	Sampling the State Rainfall Parameters.....	D-8
D.5.1	Sampling the state mean conditioned on the state variance.....	D-9
D.5.2	Sampling the state variance conditioned on the state mean.....	D-11
D.5.3	Sampling the state multivariate mean and covariance matrix jointly.....	D-12
D.5.4	Sampling the state mean and variance jointly.....	D-17

## Appendix E - Original Single Site Prior Specification

E.1	Introduction.....	E-1
E.2	Original Prior Specification.....	E-2
E.3	Comparison of Results.....	E-3
E.4	Implications.....	E-6
E.5	Conclusions.....	E-6

## Appendix F - Verification of Single Site HSM Model Calibration Procedure

F.1	Introduction.....	F-1
F.2	Results.....	F-2
F.3	Conclusion.....	F-3

## Appendix G - Comparison of Priors to Posteriors for Single Site HSM Model

G.1	Introduction.....	G-1
-----	-------------------	-----

---

G.2	Results .....	G-2
G.3	Conclusion.....	G-7
<b>Appendix H - Simulated Rainfall Results for Single Site HSM Model</b>		
H.1	Introduction.....	H-1
<b>Appendix I - Derivation of Likelihood Function for AR(1) Model</b>		
I.1	Introduction.....	I-1
I.2	Derivation of the Likelihood.....	I-2
I.3	Relationship Between Parameters in Transformed and Untransformed Space.....	I-5
<b>Appendix J - Verification of AR(1) Model Calibration Procedure</b>		
J.1	Introduction.....	J-1
J.2	Results .....	J-2
J.2.1	Bivariate Gaussian synthetic data.....	J-2
J.2.2	AR(1) synthetic data.....	J-3
J.3	Conclusion.....	J-6
<b>Appendix K - Simulated Rainfall Results for AR(1) Model</b>		
K.1	Introduction.....	K-1
<b>Appendix L - Verification of Multi-site HSM Model Calibration Procedure</b>		
L.1	Introduction.....	L-1
L.2	Results .....	L-2
L.3	Conclusion.....	L-5
<b>Appendix M - Comparison of Priors to Posteriors for Multi-site HSM Model</b>		
M.1	Introduction.....	M-1
M.2	Results .....	M-2
M.3	Conclusion.....	M-7
<b>Appendix N - Williams River Catchment Multi-site Results</b>		
N.1	Introduction.....	N-1
N.2	Results .....	N-2
N.3	Conclusion.....	N-2

---

---

# ***Abstract***

The hidden state Markov (HSM) model is introduced as a new conceptual framework for modelling long-term persistence in hydrological time series. Unlike the stochastic models currently used, the conceptual basis of the HSM model can be related to the physical processes that influence long-term hydrological time series in the Australian climatic regime. A Bayesian approach was used for model calibration. This enabled rigorous evaluation of parameter uncertainty, which proved crucial for the interpretation of the results. Applying the single site HSM model to rainfall data from selected Australian capital cities provided some revealing insights. In eastern Australia, where there is a significant influence from the tropical Pacific weather systems, the results showed a weak wet and medium dry state persistence was likely to exist. In southern Australia the results were inconclusive. However, they suggested a weak wet and strong dry persistence structure may exist, possibly due to the infrequent incursion of tropical weather systems in southern Australia. This led to the postulate that the tropical weather systems are the primary cause of two-state long-term persistence. The single and multi-site HSM model results for the Warragamba catchment rainfall data supported this hypothesis. A strong two-state persistence structure was likely to exist in the rainfall regime of this important water supply catchment. In contrast, the single and multi-site results for the Williams River catchment rainfall data were inconsistent. This illustrates further work is required to understand the application of the HSM model. Comparisons with the lag-one autoregressive [AR(1)] model showed that it was not able to reproduce the same long-term persistence as the HSM model. However, with record lengths typical of real data the difference between the two approaches was not statistically significant. Nevertheless, it was concluded that the HSM model provides a conceptually richer framework than the AR(1) model.

---

# Notation

Following will be a description of the notation commonly used in this thesis. Unless otherwise stated the convention is to refer to scalars using lowercase italic nonbold type, vectors using lowercase nonitalic bold type and matrices using uppercase nonitalic bold type.

## *Probability Notation*

$\theta$	general term for model parameters
$y$	general term for the observed data
$y_t$	observed data scalar at time $t$
$\mathbf{y}_t$	observed data vector at time $t$
$Y_N$	general term for the set of observed scalar data, $Y_N = \{y_1, \dots, y_n\}$
$\mathbf{Y}_N$	general term for the set of observed vector data, $\mathbf{Y}_N = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$
$p(\theta, y)$	joint probability density of model parameters $\theta$ and observed data $y$
$p(\theta   y)$	conditional probability density of model parameters $\theta$ conditioned on the observed data $y$ , also referred to as the posterior probability density
$p(\theta)$	probability density of the model parameters $\theta$ , also known as the prior probability density
$p(y   \theta)$	conditional probability density of observed data $y$ conditioned on the model parameters $\theta$ , also known as the likelihood function
$p(y^{rep}   Y_N)$	posterior predictive distribution of the replicated data $y^{rep}$ simulated given the model parameters conditioned on the observed data

## *Probability Distribution Notation*

$U(0,1)$	uniform probability distribution with limits 0 and 1
$N(\mu, s^2)$	univariate Gaussian distribution with scalar mean $\mu$ and variance $s^2$ ( $\sigma$ is standard deviation)

---

$N_r(\boldsymbol{\mu}, \mathbf{S})$	multivariate Gaussian distribution with $r$ dimensions, mean vector $\boldsymbol{\mu}$ of length $r$ and $r \times r$ symmetric positive definite covariance matrix $\mathbf{S}$ , the inverse of the covariance matrix, $\mathbf{S}^{-1}$ is referred to as a precision matrix
$Inv-\chi^2(\nu, s^2)$	scaled inverse-chi-square distribution with degrees of freedom $\nu$ and scale $s^2$
$Beta(\alpha, \beta)$	Beta distributions with parameters $\alpha$ and $\beta$
$Gamma(\alpha, \beta)$	Gamma distribution with shape $\alpha$ and inverse scale $\beta$
$W_r(\nu, \mathbf{W})$	Wishart distribution with dimension $r$ , degrees of freedom $\nu$ and $r \times r$ symmetric positive definite scale matrix $\mathbf{W}$

### *Prior and Posterior Parameter Notation*

In general, the prior parameters are denoted by the subscript  $o$  and the posterior parameters are by the subscript  $n$ .

$\boldsymbol{\mu}_0, \tau_0^2$	prior mean and variance for state mean
$\boldsymbol{\mu}_n, \tau_n^2$	posterior mean and variance for state mean
$\nu_0, \sigma_0^2$	prior degrees of freedom and scale for state variance
$\nu_n, \sigma_n^2$	posterior degrees of freedom and scale for state variance
$\boldsymbol{\mu}_0, \kappa_0$	prior mean and number of prior measurements on $\mathbf{S}$ scale for state mean vector
$\boldsymbol{\mu}_n, \kappa_n$	posterior mean and number of posterior measurements on $\mathbf{S}$ scale for state mean vector
$\nu_0, \mathbf{W}_0$	prior degrees of freedom and prior scale matrix for the state precision matrix, $\mathbf{W}_0$ can also be thought of as the prior precision matrix
$\nu_n, \mathbf{W}_n$	posterior degrees of freedom and posterior precision matrix for the state precision matrix

*Hidden State Markov Model Notation*

<b>P</b>	Markovian state transition probability matrix
$p_{ij}$	state transition probability that represents the probability of moving from state $i$ to state $j$ where $i, j \in W, D$
$W$	wet state
$D$	dry state
$s_t$	hidden state (wet or dry) at time $t$
$S_N$	time series of hidden states, $S_N = \{s_1, \dots, s_n\}$
$Y^W$	observed data that has been classified in the wet state
$Y^D$	observed data that has been classified in the dry state

*Two State Persistence Structure Notation*

$E(SRT_k)$	expected state residence time in state $k$ , calculated from $1/p_{TRANS}$ , where $p_{TRANS}$ is probability of transition out of state $k$
$SPS_k$	strength of the persistence structure in state $k$ Defined as either weak ( $W$ ), medium ( $M$ ), strong ( $S$ ), or very strong ( $VS$ ) using the expected state residence time, refer to Table 6.1
$WADSI$	wet and dry separation index, calculated from $(\mu_w - \mu_D) / \sqrt{\sigma_w^2 + \sigma_D^2}$
$TSP$	two state persistence structure notation, for the single site it is defined as $[SPS_w, SPS_D, WADSI]$ whereas for the multi-site there is a vector of $WADSI$ values, one for each site

*ARMA model notation*

$\mu$	scalar mean of time series
$\phi_j$	autoregressive parameter at lag $j$
$\alpha_j$	moving average parameter at lag $j$
$\varepsilon_t$	error term, an uncorrelated Gaussian random variable, $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$



---

*Markov Chain Monte Carlo Derivation Notation*

This notation is primarily used in Appendix C. The usual convention of denoting vectors as nonitalic bold type is not used in Appendix C.

$x^i$	a sample $x$ , possibly scalar, generally a vector, generated by a MCMC method at iteration $i$
$\pi(\cdot)$	“target” probability density
$\pi^*(\cdot)$	target probability distribution
$P(A x)$	Markov chain transition probability kernel, that represents the probability of making a transition from a point $x$ to a point in the region defined by $A$
$q(y x)$	MCMC candidate generating density, represents the probability of generating a sample $y$ given the current state of the process $x$
$p(y x)$	function that gives the probability of a transition from $x$ to $y$
$r(x)$	probability that chain remains at $x$ , $r(x) = 1 - \int p(y x) dy$
$\alpha(y x)$	probability of move from $x$ to $y$
$\delta(x)$	Dirac delta function, with property that $\int_{-\infty}^{\infty} \delta(x) dx = 1$ , with $\delta(x) \neq 0$ when $x \in A$ , for some region $A$ and $\delta(x) = 0$ when $x \notin A$
$b$	number of iterations required to “warm-up” the MCMC chain before convergence is achieved