IDENTIFICATION AND MODELLING OF HYDROLOGICAL PERSISTENCE WITH HIDDEN MARKOV MODELS

Julian Peter Whiting
B.E. (Civil & Env.) (Hons.), B.Sc.

A thesis submitted for Doctor of Philosophy
at the University of Adelaide

October 2006

Table of contents

List of figures

List of tables

Abstract

Hydrological observations are characterised by wet and dry cycles, a characteristic that is termed hydrological persistence. Interactions between global climate phenomena and the hydrological cycle result in rainfall and streamflow data clustering into wetter and drier states. These states have implications for the management and planning of water resources. Statistical tests constructed from the theory of wet and dry spells indicate that evidence for persistence in monthly observations is more compelling than at an annual scale. This thesis demonstrates that examination of monthly data yields spatially-consistent patterns of persistence across a range of hydrological variables. It is imperative that time series models for rainfall and streamflow replicate the observed fluctuations between the climate regimes. Monthly time series are generally represented with linear models such as ARMA variants; however simulations from such models may underestimate the magnitude and frequency of persistence. A different approach to modelling these data is to incorporate shifting levels in the broader climate with a tendency to persist within these regimes. Hidden Markov models (HMMs) provide a strong conceptual basis for describing hydrological persistence, and are shown to provide accurate descriptions of fluctuating climate states. These models are calibrated here with a full Bayesian approach to quantify parameter uncertainty.

A range of novel variations to standard HMMs are introduced, in particular Autoregressive HMMs and hidden semi-Markov models which have rarely been used to model monthly rainfall totals. The former model combines temporal persistence within observations with fluctuations between persistent climate states, and is particularly appropriate for modelling streamflow time series. The latter model extends the modelling capability of HMMs by fitting explicit probability distributions for state durations. These models have received little attention for modelling persistence at monthly scale. A non-parametric (NP) HMM, which overcomes the major shortcomings of standard parametric HMMs, is also described. Through removing the requirement to assume parametric forms of conditional distributions prior to model calibration, the innovative NP HMM framework provides an improved estimation of persistence in discrete and continuous data that remains unaffected by incorrect parametric assumptions about the state distributions. Spatially-consistent persistence is identified across Australia with the NP HMM, showing a tendency toward stronger persistence in low-rainfall regions. Coherent signatures of persistence are also identified across time series of total monthly rainfall, numbers of rain-days each month, and the intensities of the most extreme rain events recorded each month over various short durations, illustrating that persistent climate states modulate both the numbers of rain events and the amount of moisture contained within these events. These results provide a new interpretation of the climatic interactions that underlie hydrological persistence.

The value of HMMs to water resource management is illustrated with the accurate simulation of a range of hydrologic data, which in each case preserves statistics and spell properties over a range of aggregations. Catchment-scale rainfall for the Warragamba Reservoir is simulated accurately with HMMs, and rainfall-runoff transformations from these simulations provide reservoir inflows of lower drought risk than provided from ARMA models.

Statement of Originality

I hereby certify that this work has not been submitted for the award of any other degree or diploma in any university of 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. I give my consent to this thesis, when deposited in the University Library, being available for photocopying and loan

Julian Peter Whiting   Date

Acknowledgements

I would firstly like to acknowledge the support and guidance offered by my two supervisors Martin Lambert and Andrew Metcalfe throughout the period of my study. I am proud of what we have achieved through this time, and of the confidence that I have found to pursue my own research directions.

I would like to thank my friends and colleagues in the School of Civil and Environmental Engineering for making the past four years such an enjoyable time. Thank you to all who assisted in various ways, or generally contributed to the fun of it all. I am also grateful for the financial support provided to me through the School.

To the large ensemble originating from the University of Newcastle: George Kuczera, Mark Thyer, Stewart Franks and Andrew Frost, my thanks for your many ideas, criticisms and suggestions along the way. Thanks also to Tom Micevski for making your computer code and insights available to me.

To my parents Michael and Janine, and to Matt and Linden, an enormous thanks for your continual encouragement.

And to Lauren, thank you so much, it's done – this is half yours too.

List of Symbols

Climate terminology

ACW	Antarctic Circumpolar Wave
DMI	(Indian Ocean) Dipole Mode Index
ENSO	El-Niño Southern Oscillation
IOD	Indian Ocean Dipole
IPO	Interdecadal Pacific Oscillation
NAO	North Atlantic Oscillation
PDO	Pacific Decadal Oscillation
PC_1	First Principal Component calculated from NINO3, IPO and DMI
SLP	Sea-level atmospheric pressures
SST	Sea-surface temperature
SO	Southern Oscillation
SOI	Southern Oscillation Index

Time series notation

y_t	Observation at time *t*
$\{y_t\}$	Time series of observations
Y_t	Series of observations up to time *t*, $\{y_1, y_2, ..., y_t\}$
$\{u_t\}$	Time series of observations transformed into $(0,1)$ interval
$\{x_t\}$	Time series of HMM states
$\{s_1, s_2, ..., s_k\}$	Series of HMM states
$s_{i,t}$	HMM state series being in state i at time t
$\{z_t\}$	Time series of white noise/ random variation
T	Time series length
ϕ_p	Serial correlation coefficient at lag *p* for a $AR(p)$ model
φ_q	Moving-average coefficient at lag *q* for a $MA(q)$ model
5-mrm	Five-month running-means
H	Hurst exponent

Probability notation

$BF_{i,j}$	Bayes factor comparing model i to model j
$N(\mu_j, \sigma_j^2)$	Gaussian distribution with mean μ_j and standard deviation σ_j^2, with j corresponding to model state where applicable
$Gamma(\alpha, \beta)$	Gamma distribution with shape α and scale β
H	Hurst exponent
P_{ij}	HMM transition probability from state *i* to state *j*
ΣTP	Sum of HMM transition probabilities
$P(a \mid b)$	Conditional probability of a given b
$P(a, b)$	Joint probability of a and b
θ	Conditional vector of model parameters
W	HMM wet state
D	HMM dry state

Stochastic modelling abbreviations

AD	Anderson-Darling goodness-of-fit statistic (Stephens, 1974)
AIC	Akaike's Information Criterion (Akaike, 1974)
AM	Adaptive Metropolis algorithm (Haario *et al.*, 2001)
LORT	Gold's length-of-runs test (Gold, 1929)
MCMC	Monte Carlo Markov Chain method
SCE	Shuffled Complex Evolution algorithm
SL	Shifting Level model