

NONHOMOGENEITY IN EASTERN AUSTRALIAN FLOOD FREQUENCY DATA: IDENTIFICATION AND REGIONALISATION

Tom Micevski
BE (Hons), BSurv (Hons)

A thesis submitted for the degree of

Doctor of Philosophy

SCHOOL OF ENGINEERING



THE UNIVERSITY OF NEWCASTLE
NEW SOUTH WALES, 2308
AUSTRALIA

September 2007

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.

Tom Micevski

Acknowledgements

Many thanks must first go to my supervisors, Professor George Kuczera and Associate Professor Stewart Franks, whose wisdom, guidance, and encouragement have helped me considerably in this long journey towards the completion of this thesis¹.

Thanks also to the other postgrads in the Civil Engineering Department who have helped make my time at university much more enjoyable: (twin brother) Dmitri Kavetski, Andrew Frost, Adam Wyatt, Danielle Verdon, Micah Bell, Brendan Drake, Bobby Paquet, Chad Jefferson, and Matthew McCabe (go JT!).

My family have provided much support over the duration of my thesis. Many thanks to all, especially to my dad.

¹This thesis was initially submitted for examination in August 2006.

Associated publications

The research work embodied in this thesis led to the following series of research papers:

1. Micevski, T., Kiem, A. S., Franks, S. W., Kuczera, G., 2003. Multidecadal variability in New South Wales flood data. In: Proc. 28th International Hydrology Water Resources Symposium. Vol. 1. Institution of Engineers Australia, Wollongong, NSW, pp. 173–180, 10–13 Nov.
2. Micevski, T., Kuczera, G., Franks, S. W., 2003. Bayesian flood regionalisation for New South Wales. In: Proc. 28th International Hydrology Water Resources Symposium. Vol. 3. Institution of Engineers Australia, Wollongong, NSW, pp. 19–28, 10–13 Nov.
3. Micevski, T., Kuczera, G., Franks, S. W., 2004. A Bayesian hierarchical nonhomogeneous regional flood model for New South Wales, Australia. In: Proc. World Water and Environmental Resources Congress 2004 (CD-ROM). Environmental and Water Resources Institute (EWRI), Salt Lake City, UT, USA, 27 Jun–1 Jul.
4. Micevski, T., Franks, S. W., Kuczera, G., 2005. Multidecadal variability in Queensland flood data. In: Proc. 29th International Hydrology Water Resources Symposium (CD-ROM). Engineers Australia, Canberra, ACT, 21–23 Feb.
5. Micevski, T., Kuczera, G., Franks, S. W., 2005. Flood frequency censoring errors associated with daily-read flood observations. *Water Resour. Res.* 41 (7), W07002, doi:10.1029/2004WR003881.
6. Micevski, T., Franks, S. W., Kuczera, G., 2006. Multidecadal variability in coastal eastern Australian flood data. *J. Hydrol.* 327 (1-2), 219–225.

Contents

Abstract	x
1 Introduction	1
1.1 Overview	1
1.2 Objectives	2
1.3 Outline	4
2 Flood data affected by censoring errors	5
2.1 Introduction	5
2.2 Flood frequency data from eastern Australia	5
2.3 Relationship between true peak and daily-read flows	7
2.4 Evaluation of methods	12
2.4.1 Monte Carlo Experiment	12
2.4.2 Method 0: Gauged Only (GO)	13
2.4.3 Method 1: Censored As Gauged (CAG)	14
2.4.4 Method 2: Binomial Censoring (BC)	14
2.4.5 Method 3: Random Dependent Censoring (RDC)	14
2.4.6 Results and discussion	16
2.5 Conclusion	21
3 Multidecadal variability in flood risk	24
3.1 Introduction	24
3.2 Climatic variability	24
3.3 Interdecadal Pacific Oscillation (IPO)	26
3.4 Data and methodology	27
3.5 Results	29
3.6 Discussion	33
3.7 Conclusion	36
4 A Bayesian hierarchical regional flood model	38
4.1 Overview	38
4.2 Background	39
4.2.1 Index flood regionalisation methods	39
4.2.2 Map-based regionalisation methods	43

4.2.3	Regression-based regionalisation methods	44
4.2.4	Methods for the delineation of regions	46
4.2.5	Discussion	47
4.3	Flood regionalisation: A hierarchical model perspective	48
4.4	Bayesian inference for the regional flood model	51
4.4.1	Bayesian inference framework	51
4.4.2	Overview of MCMC sampling techniques	52
4.4.3	Gibbs sampler implementation for the hierarchical regional flood model	59
4.4.4	Assessment of hierarchical regional flood model inference using syn- thetic data	65
4.4.5	Predicting the flood frequency distribution at a new site	72
4.5	Discussion	76
4.6	Conclusion	77
5	Case study investigations	78
5.1	Overview	78
5.2	Case studies	79
5.2.1	Selection of catchment variables for the site-mean regression equation	79
5.2.2	Delineation of regions	79
5.2.3	Results of regional analysis for IPO-stratified flood data	80
5.2.4	Analysis of residuals for IPO-stratified flood data	84
5.2.5	Joint procedure for regional model with correlated errors	87
5.2.6	Results of regional analysis using joint procedure	91
5.2.7	Analysis of residuals for joint procedure	94
5.2.8	Prediction at an ungauged site	100
5.2.9	Prediction at a gauged site	103
5.3	Discussion	105
5.3.1	Regional-parameter variability	105
5.3.2	Intersite correlation	106
5.4	Conclusion	106
6	Conclusions	109
6.1	Introduction	109
6.2	Summary	109
6.2.1	Development of a new procedure for censored data in flood frequency analysis	109
6.2.2	Investigation of the nonhomogeneity of eastern Australian flood data	110
6.2.3	Development of a hierarchical regional flood model using a Bayesian framework	111
6.2.4	Techniques for prediction at a new (ungauged or gauged) site using the regional posterior distribution	112

6.2.5	Application of the regional flood model to flood data from eastern Australia	112
6.3	Future work	113
6.3.1	Further investigations relating to censored data	113
6.3.2	Areas affected by the nonhomogeneity of flood risk	114
6.3.3	Reducing the variability in the regional analysis	114
6.3.4	Testing the correlated-site infilling hypothesis	114
6.3.5	Comparison between joint and two-stage models	114
6.4	Final remarks	115
A	Summary of site data	116
	Bibliography	119

List of Figures

1.1	Synthetic flood series showing multidecadal drought flood epochs.	2
2.1	Location of the 127 sites within New South Wales and Queensland.	6
2.2	Comparison of the annual maximum true peak and daily-read discharges. . .	8
2.3	Stage and hydrograph for site 210006.	10
2.4	Quantile-quantile plots of the censoring ratio.	11
2.5	Quantile rmse and bias for the CAG and GO methods [$\log Q \sim N(8.0, 1.2^2)$, $E(r) = 1.0$, $CV(r) = 1.0$].	16
2.6	Quantile bias for the CAG and GO methods [$\log Q \sim N(8.0, 1.2^2)$, $E(r) = 0.2$, $CV(r) = 1.0$].	18
2.7	Quantile rmse and bias for the BC and GO methods [$\log Q \sim N(8.0, 1.2^2)$, $E(r) = 1.0$, $CV(r) = 1.0$].	18
2.8	Quantile rmse and bias for the RDC and GO methods [$\log Q \sim N(8.0, 1.2^2)$, $E(r) = 1.0$, $CV(r) = 1.0$].	19
2.9	Quantile rmse and bias for the RDC and GO methods [$\log Q \sim N(8.0, 0.2^2)$, $E(r) = 1.0$, $CV(r) = 1.0$].	20
2.10	Quantile rmse and bias for the CAG and GO methods [$\log Q \sim N(8.0, 0.2^2)$, $E(r) = 1.0$, $CV(r) = 1.0$].	21
2.11	Quantile rmse and bias for the RDC and GO methods with misspecification errors [$\log Q \sim N(8.0, 1.2^2)$, true $E(r) = 1.0$, erroneous $E(r) = 0.2$, $CV(r) = 1.0$].	22
2.12	Quantile rmse and bias for the RDC and GO methods with misspecification errors [$\log Q \sim N(8.0, 1.2^2)$, true $E(r) = 0.2$, erroneous $E(r) = 1.0$, $CV(r) = 1.0$].	22
3.1	Annual IPO index time series.	27
3.2	Histograms of record length for both IPO epochs.	28
3.3	A typical good lognormal fit to the IPO-stratified data.	29
3.4	A poor lognormal fit to the IPO-stratified data.	30
3.5	Histogram of p values for the Kolmogorov–Smirnov test.	30
3.6	Histograms of flood ratios for several ARIs.	31
3.7	Relationship between 10-year flood ratio and catchment area.	32
3.8	Relationship between 10-year flood ratio and latitude.	33
3.9	Spatial distribution of the 10-year flood ratio.	34

3.10	Flood series showing IPO stratifications.	36
4.1	Hierarchical model structure.	49
4.2	Schematic representation of the Markov-chain transition kernel $K(\mathbf{b} \mid \mathbf{a})$. . .	53
4.3	Histograms of σ_{ε}^2	67
4.4	Histograms of σ_{δ}^2	68
4.5	Schematic representation of distributions.	69
4.6	Histograms of regional parameters derived using the uniform sampling procedure.	71
5.1	Location of sites within the four regions.	80
5.2	Histograms of record lengths for both IPO epochs for all four regions. . . .	82
5.3	Residual plots for the site mean comparing IPO- and IPO+ analyses.	85
5.4	Residual plots for the (logarithm of the) site standard deviation comparing IPO- and IPO+ analyses.	85
5.5	Residual plots for the site mean comparing IPO- and IPO+ analyses.	86
5.6	Residual plots for the (logarithm of the) site standard deviation comparing IPO- and IPO+ analyses.	86
5.7	Residual plots for the site mean comparing IPO- and IPO+ analyses.	95
5.8	Residual plots for the (logarithm of the) site standard deviation comparing IPO- and IPO+ analyses.	95
5.9	Residual plots comparing the site mean and (log) standard deviation.	96
5.10	Spatial patterns of the residuals.	98
5.11	Site mean normal probability plots.	99
5.12	Site standard deviation normal probability plots.	99
5.13	Posterior distribution of selected flood quantiles for IPO- and IPO+ epochs for site 210001.	100
5.14	Posterior distribution of regional mean for region 3.	102
5.15	Posterior distribution of selected flood quantiles for the “true”, “perfect parameter”, and “perfect regression” cases for site 210001.	102
5.16	At-site and regionalised flood frequency distributions for sites located in all four regions.	104
5.17	Years of available data in region 3 (to illustrate possible infilling of missing years).	107

Abstract

Flood frequency data from the eastern Australian states of New South Wales (NSW) and Queensland (Qld) were investigated to determine the magnitude and extent of multi-decadal variability (nonhomogeneity) in flood risk. Some flood data from NSW were found to be systematically in error because daily-read discharges were used instead of instantaneous peak discharges. A new approach, based on the method of maximum likelihood, was developed to overcome the potential artefacts introduced by the use of daily-read data in flood frequency analysis. However, it was shown for flood data typical of NSW, the treatment of daily-read data as instantaneous peaks did not introduce sufficiently large quantile bias and loss of mean-squared-error performance to warrant use of the new estimation method.

The flood data were stratified by Interdecadal Pacific Oscillation (IPO) value and flood frequency analyses performed on the IPO-stratified flood data — the IPO is a climate index of Pacific Ocean sea surface temperature anomalies, which displays variability on a long-term (multidecadal) time scale. The IPO was found to modulate the flood risk in NSW and southern Qld, with flood quantiles being increased, on average, by approximately 1.7 times during IPO-negative epochs, whereas little effect was detected for sites in northeast Qld located approximately north of the Tropic of Capricorn. The IPO modulation (nonhomogeneity) of flood risk has great practical significance — the use of at-site flood data with inadequate coverage of both IPO epochs may result in biased estimates of long-run flood risk.

A Bayesian regional flood model framework, based on hierarchical modelling concepts, was developed to overcome the possible bias in long-run flood risk associated with a non-homogeneous flood record. Importantly, the model allows for the consideration of intersite correlation. Bayesian methods were used to enable a rigorous treatment of uncertainty in the flood regionalisation problem. The Gibbs sampler was used to infer uncertainty in the regional model parameters, while importance-sampling-based procedures were developed to compute the Bayesian predictive distribution and the posterior distribution of quantiles at a new site, which may be ungauged or gauged. This represents the first truly-general Bayesian solution for combining regional and gauged information in flood frequency analysis.

The flood data from eastern Australia were partitioned into four regions and analysed using the Bayesian hierarchical regional model. The (correlated-site) regional model found significant differences (at the 10% level) in the regional means for the two regions in NSW. The regional standard deviations showed significant differences for the two regions in Qld,

but these differences were opposite in sign, with IPO-positive standard deviations being greater than IPO-negative values. The equivalent gauged length provided by the regional model had a maximum of 4 to 8 years.

There is a large overlap in the probability limits between the IPO-positive and IPO-negative regional distributions for a flood quantile. However, because the regional model errors for the IPO phases are highly correlated, the difference in IPO-positive and IPO-negative quantiles is likely to be significant. For larger return periods, the opposite-in-sign differences in the regional mean and standard deviation may reduce the resultant IPO-related differences (in discharge).

The value of the regional model was demonstrated by pooling the regional information with the information in short gauged records at selected sites in each region. The pooled at-site flood frequency distribution provided substantial improvements over the gauged record alone (in terms of prediction limits and bias). Indeed, this was especially evident in a situation where a shortened (10-year) gauged record was found inconsistent with the true (long-run) gauged record — the shortened gauged record consisted mainly of years from the IPO-positive epoch. These results suggest that the use of the regional model may protect against bias in long-run flood risk at sites with short records sampled largely in one IPO epoch.