A Bayesian Hierarchical Regional Flood Model

Tom Micevski[1], George Kuczera[1], Stewart W. Franks[1]

1 Civil, Surveying, and Environmental Engineering, University of Newcastle, NSW, 2308.

Abstract: Recent studies have shown that flood data from eastern Australian catchments may demonstrate variability in flood risk over multidecadal time scales, characterised by crossings of the Interdecadal Pacific Oscillation (IPO) climate index. This nonhomogeneity of flood risk may lead to a significant prospect of biased long-run flood risk from at-site flood data with insufficient coverage of both IPO epochs. This paper develops a Bayesian hierarchical regional model, implemented using the Gibbs sampler, to overcome this possible bias in flood risk. The hierarchical model proposes that the parameters of the flood frequency distribution at any site are random samples from a regional probability model, allowing for intersite variability, while also permitting spatial correlation between concurrent floods. An outcome is that the predictive uncertainty at an ungauged or gauged site may be quantified.

Keywords: Regional model, Bayesian analysis, Gibbs sampler, ungauged site, correlation.

1. INTRODUCTION

A basic assumption of flood frequency analysis is that flood peaks are independent and identically distributed. However, recent studies have questioned the validity of this assumption with evidence showing the existence of more than one distribution within eastern Australian flood data, especially for New South Wales (NSW) [Erskine and Warner 1988; Franks 2002; Franks and Kuczera 2002].

The association between El Nino/Southern Oscillation (ENSO) and Australian climate was found [Power et al. 1999] to be modulated by the Interdecadal Pacific Oscillation (IPO), a climate index of multidecadal Pacific Ocean sea surface temperature anomalies (see Section 2). Kiem et al. [2003] analysed IPO-stratified flood data from NSW and found that the IPO modulated both the magnitude and frequency of ENSO events (El Nino and La Nina) resulting in multidecadal periods of elevated and reduced flood risk. La Nina events were found to be the primary drivers of flood risk and this was further enhanced under negative IPO phases. These results have obvious consequences for Australian flood risk. Micevski et al. [2006] performed flood frequency analyses on IPO-stratified flood data from eastern Australia. The IPO was found to modulate the flood risk in NSW and southern Qld, with flood quantiles being increased by a factor of approximately 1.7 during negative IPO epochs. Thus, there is a large prospect of significant bias in long-run flood risk when using at-site data with poor coverage of one of the IPO epochs.

This study describes a Bayesian regional hierarchical flood model to overcome the possible bias in long-run flood risk associated with a nonhomogeneous flood record. The hierarchical model proposed in Micevski et al. [2003] is refined and implemented using the Gibbs sampler. Results are presented for eastern Australian flood data.

2. DATA

2.1 Interdecadal Pacific Oscillation (IPO)

The IPO is the coherent pattern of SST variability occurring on interdecadal time scales over the (entire) Pacific Ocean [Power et al. 1999; Folland et al. 2002]. It is characterised by the third empirical orthogonal function of 13-year low-pass filtered global SSTs, projected onto annual data [Folland et al. 2002]. Note that the IPO has a similar time series to that of the Pacific Decadal Oscillation (PDO) [Mantua et al. 1997], which is defined as the leading principal component of North Pacific Ocean SST anomalies, poleward of 20°N.

The annual IPO time series is presented in Figure 1 and represents an average of four seasonal values — the IPO data were obtained directly from the United Kingdom Met Office. Note that the time series reveals extended epochs above and below the long-term average.

2.2 Flood data

Annual maximum flood data from NSW and Qld were used in this study. The NSW data was obtained from the NSW Department of Land and Water Conservation Pinneena database, while the Qld data was provided directly by the Qld Department of Natural Resources and

Mines. Flood data was obtained for a total of 127 sites, with Qld having 85 sites and NSW 42 sites. The locations of the sites are shown in Figure 2. The yearly peak flows were extracted using a water year from April to March, which corresponds to the typical ENSO cycle.
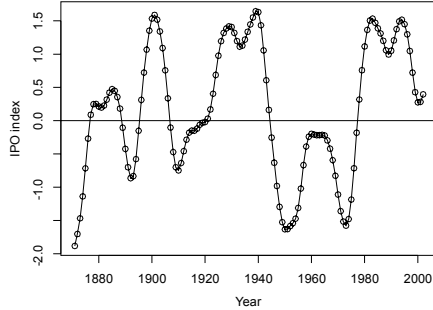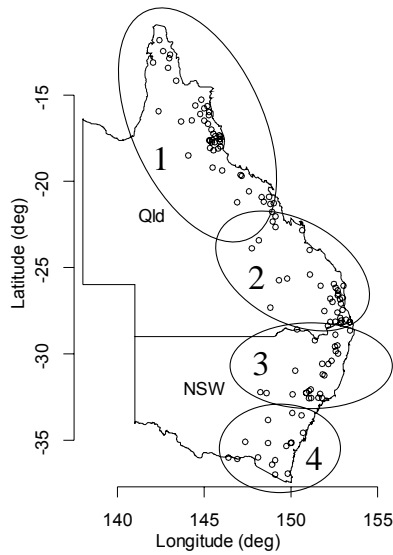

Figure 1. Annual IPO time series.

Figure 2. Location of sites and regions.

3. HIERARCHICAL MODEL

A hierarchical model has observable outcomes that are modelled conditionally on certain parameters, which themselves are given a probabilistic specification in terms of further parameters [Gelman et al. 1995, p. 119]. That is, the observable outcomes (annual floods) are sampled from a site's own flood model whose parameters (eg. mean and standard deviation) are considered to be sampled from a regional flood model. Thus, the hierarchical model consists of a two-level structure (see Figure 3), the two levels being the regional and site models, and can be thus parameterised $\Theta = \{\theta_R, \theta_S\}$, where Θ is the vector of all regional model parameters, and θ_R and θ_S are the vectors of regional and site model parameters, respectively.
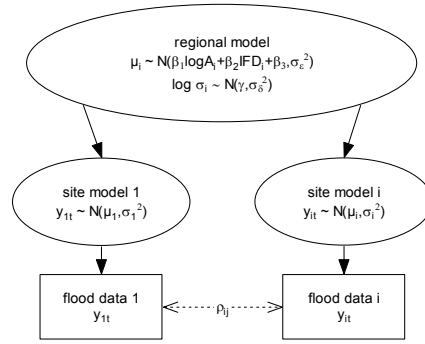

Figure 3. Hierarchical model structure.

Assume that the annual maximum flood is lognormally distributed, $y_{it} \sim N(\mu_i, \sigma_i^2)$, for $i=1,\ldots,n$ and $t=1,\ldots,T_i$; where y_{it} is the (natural) logarithm of the flood for site i at time t, n and T_i are the number of gauged sites and years of data at each site respectively, and μ_i and σ_i are respectively the mean and standard deviation of y_i. Also, assume that the floods at each site are independent in time, but are spatially correlated. This is important because the information content of the flood data is reduced when floods are spatially correlated. Often spatial correlation is ignored, which may then overestimate the predictive power of the regional model — the generalised least squares procedure [Stedinger and Tasker 1985] is a notable exception. The intersite correlation may be described using an exponential decay function:

$$\rho_{ij} = \exp(-[d_{ij}/A]^B) \qquad [1]$$

where ρ_{ij} is the correlation coefficient between sites i and j, d_{ij} is the distance between the sites, and A and B are correlation parameters to be estimated.

In this study, the regional model took the form of a normal linear model, with the site mean being assumed to be a function of the logarithm of catchment area (log A) and the 2-year, 12-hour rainfall intensity (IFD), while the site standard deviation is assumed to have no dependent variables. Note that the approach is quite general allowing other catchment descriptors to be used. Thus, the hierarchical model consists of the regional model

$$\mu_i \sim N(\beta_1 \log A_i + \beta_2 IFD_i + \beta_3, \sigma_\varepsilon^2) \qquad [2]$$
$$\log \sigma_i \sim N(\gamma, \sigma_\delta^2) \qquad [3]$$

and the site model, for independent sites

$$y_{it} \sim N(\mu_i, \sigma_i^2) \qquad [4]$$

or, for correlated sites

$$\mathbf{y}_t \sim N(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t) \qquad [5]$$

where $\mathbf{y}_t$ is the vector of flood data, $\boldsymbol{\mu}_t$ is the vector of site means, and $\boldsymbol{\Sigma}_t$ is the covariance matrix (for the current year t), where the elements of the covariance matrix are formed using $\Sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$, where ρ_{ij} is defined in [1]. Note that the regional flood model `sits above' the gauged sites and furnishes different means μ_i and standard deviations σ_i to each site, allowing for variation between sites (see Figure 3). Also, note that $\boldsymbol{\theta}_R = \{\sigma^2_\varepsilon, \boldsymbol{\beta}, \sigma^2_\delta, \gamma\}$ and $\boldsymbol{\theta}_S = \{\boldsymbol{\mu}, \log \boldsymbol{\sigma}\}$.

4.   MODEL CALIBRATION

A Bayesian approach is used to infer the parameters of the hierarchical model (model calibration) — Bayesian methods explicitly account for parameter uncertainty, allowing for a rigorous treatment of the flood regionalisation problem. Thus, the use of Bayes Theorem is essential in this analysis. Bayes Theorem is expressed:

$$p(\boldsymbol{\Theta}|Y) = f(Y|\boldsymbol{\Theta})p(\boldsymbol{\Theta}) / p(Y) \propto f(Y|\boldsymbol{\Theta})p(\boldsymbol{\Theta}) \qquad [6]$$

where $p(\boldsymbol{\Theta}|Y)$ is the posterior density describing the current knowledge about the model parameter vector $\boldsymbol{\Theta}$, given the observed data Y, $f(Y|\boldsymbol{\Theta})$ is the likelihood function that defines the model fit for a particular set of model parameters, $p(\boldsymbol{\Theta})$ is the prior density that contains our subjective belief about the true value of $\boldsymbol{\Theta}$, and $p(Y)$ is the marginal likelihood (a normalising constant).

The Gibbs sampler [eg. Gelman et al. 1995] is particularly well suited to hierarchical models and is used for parameter inference.   The hierarchical model is partitioned into two subvectors $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_R, \boldsymbol{\theta}_S\}$, and each iteration of the Gibbs sampler cycles through the two subvectors of $\boldsymbol{\Theta}$, randomly sampling each subvector from its conditional posterior distribution. The conditional posteriors can be simplified:

$$\begin{aligned} p(\boldsymbol{\Theta}_R|\boldsymbol{\Theta}_S, Y) &= p(Y|\boldsymbol{\Theta}_R, \boldsymbol{\Theta}_S)p(\boldsymbol{\Theta}_R|\boldsymbol{\Theta}_S) / p(Y|\boldsymbol{\Theta}_S) \\ &= p(Y|\boldsymbol{\Theta}_S)p(\boldsymbol{\Theta}_R|\boldsymbol{\Theta}_S) / p(Y|\boldsymbol{\Theta}_S) \\ &= p(\boldsymbol{\Theta}_R|\boldsymbol{\Theta}_S) \qquad [7] \end{aligned}$$

using Bayes Theorem and exploiting the hierarchical structure (the regional parameters $\boldsymbol{\theta}_R$ only affect the data Y indirectly through the site parameters $\boldsymbol{\theta}_S$). Similarly,

$$p(\boldsymbol{\Theta}_S|\boldsymbol{\Theta}_R, Y) = p(Y|\boldsymbol{\Theta}_S, \boldsymbol{\Theta}_R)p(\boldsymbol{\Theta}_S|\boldsymbol{\Theta}_R) / p(Y|\boldsymbol{\Theta}_R)$$

$$= p(Y|\boldsymbol{\Theta}_S)p(\boldsymbol{\Theta}_S|\boldsymbol{\Theta}_R) / p(Y|\boldsymbol{\Theta}_R) \qquad [8]$$

The algorithm for the Gibbs sampler is:

(0) Assign starting value $\boldsymbol{\theta}_S^{(0)}$
(1) Randomly sample regional subvector:
$\boldsymbol{\theta}_R^{(j)} \leftarrow p(\boldsymbol{\theta}_R|\boldsymbol{\theta}_S^{(j-1)}, Y) = p(\boldsymbol{\theta}_R|\boldsymbol{\theta}_S^{(j-1)})$
(2) Randomly sample site subvector:
$\boldsymbol{\theta}_S^{(j)} \leftarrow p(\boldsymbol{\theta}_S|\boldsymbol{\theta}_R^{(j)}, Y) = p(Y|\boldsymbol{\theta}_S)p(\boldsymbol{\theta}_S|\boldsymbol{\theta}_R^{(j)})/p(Y|\boldsymbol{\theta}_R^{(j)})$
(3) Iterate steps (1) and (2), j=1,…,N times.

Some important Gibbs sampler implementation issues are now outlined.

4.1   Regional parameters

Since the components of the regional model consist of normal linear models, standard regression results may be used to sample these parameters [eg. Gelman et al. 1995, pp. 235-7].

4.2   Site parameters

No standard results are available for the site model parameters, so a `Metropolis-within-Gibbs' step [eg. Gelman et al. 1995] must be used to sample the site parameters.   The conditional posterior is used for the Metropolis-within-Gibbs algorithm:

$$p(\boldsymbol{\theta}_S|\boldsymbol{\theta}_R, Y) \propto p(Y|\boldsymbol{\theta}_S)\, p(\boldsymbol{\theta}_S|\boldsymbol{\theta}_R) \qquad [9]$$

where these terms are evaluated using:

$$p(Y|\boldsymbol{\Theta}_S) = p(Y|\boldsymbol{\mu}, \log \boldsymbol{\sigma}) = \Pi^T_{t=1} N(\mathbf{y}_t|\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$$
$$p(\boldsymbol{\Theta}_S|\boldsymbol{\Theta}_R) = p(\boldsymbol{\mu}|\boldsymbol{\Theta}_R)p(\log \boldsymbol{\sigma}|\boldsymbol{\Theta}_R)$$
$$p(\boldsymbol{\mu}|\boldsymbol{\Theta}_R) = \Pi^n_{t=1} N(\mu_i|\beta_1^{(j)}\log A_i + \beta_2^{(j)}IFD_i + \beta_3^{(j)}, \sigma^2_\varepsilon{}^{(j)}I)$$
$$p(\log \boldsymbol{\sigma}|\boldsymbol{\Theta}_R) = \Pi^n_{t=1} N(\log \sigma_i|\gamma^{(j)}, \sigma^2_\delta{}^{(j)}I) \qquad [10]$$

where $N(x|a,b)$ denotes a normal (Gaussian) density evaluated at x with mean a and variance b, and I is the identity matrix. Note that using a `single-block' Metropolis-within-Gibbs step for the $\boldsymbol{\theta}_S$ vector may lead to poor acceptance rates because some components of $\boldsymbol{\theta}_S$ may be poorly sampled in each Gibbs sampler iteration, leading to the entire $\boldsymbol{\theta}_S$ proposal being rejected. This problem may be overcome by modifying the Metropolis procedure so that the parameters of each site are sampled individually, rather than all in a single block, using n Metropolis-within-Gibbs steps instead of one only.   This modification increases the computational effort required.

4.3   Starting value

The starting value $\boldsymbol{\theta}_S^{(0)}$ is obtained using a mode searching algorithm (eg. quasi-Newton

method) on the likelihood component of the $p(Y|\boldsymbol{\theta}_S)$ unnormalised conditional posterior [9]. The two correlation parameters (A and B) are also determined at this time, if required.

5. RESULTS

The IPO-stratified flood data were analysed using the hierarchical model, using four regions (see Figure 2). The regional posterior parameters for the independent-site analyses are given in Table 1.

Table 1. Regional posterior parameters (independent-site analysis).

Parameter	IPO-		IPO+	
	Mean	SD	Mean	SD
Region 1				
β_1	0.67	0.06	0.63	0.06
β_2	0.13	0.03	0.13	0.03
β_3	4.72	0.65	4.85	0.59
σ^2_ε	0.29	0.08	0.25	0.07
γ	-0.21	0.07	0.03	0.05
σ^2_δ	0.19	0.05	0.11	0.03
Region 2				
β_1	0.52	0.06	0.52	0.06
β_2	0.27	0.06	0.29	0.06
β_3	4.24	0.74	3.74	0.79
σ^2_ε	0.19	0.09	0.33	0.10
γ	0.31	0.04	0.25	0.04
σ^2_δ	0.02	0.02	0.05	0.02
Region 3				
β_1	0.65	0.06	0.67	0.06
β_2	0.46	0.08	0.43	0.07
β_3	2.56	0.84	2.12	0.75
σ^2_ε	0.27	0.10	0.20	0.08
γ	0.10	0.05	0.06	0.05
σ^2_δ	0.04	0.02	0.04	0.02
Region 4				
β_1	0.56	0.10	0.58	0.10
β_2	0.20	0.10	0.22	0.10
β_3	4.62	1.10	4.01	1.15
σ^2_ε	0.34	0.19	0.37	0.21
γ	0.11	0.08	0.09	0.09
σ^2_δ	0.08	0.04	0.11	0.05

Table 1 displays no apparent significant difference for the regional parameters between the IPO- and IPO+ epochs, after consideration of the associated standard deviations. Similar results were also achieved for the correlated-site model (not shown). This appears to contradict previous studies showing that at-site flood risk in eastern Australian is affected by IPO modulation (see Section 1). However, the regional model appears to be affected by considerable noise, due to the large areal extent of the regions and the small number of sites (see Figure 2) — the noise in the regional model swamps any differences due to IPO. This can be demonstrated by considering the equivalent gauged record length of the regional model for the mean, which is summarised in Table 2. The equivalent record lengths range from 2-12 years, so it is not surprising that the

regional model has difficulty detecting any IPO-related differences in regional parameters.

Table 2. Equivalent record lengths (years).

Reg.	Independent-site anal.		Correlated-site anal.	
	IPO-	IPO+	IPO-	IPO+
1	2	4	4	4
2	10	5	12	5
3	5	6	6	6
4	4	3	4	5

The regional parameters also show little discernible difference between the independent- and correlated-site analyses. The most likely explanation for the apparent lack of difference is that the considerable noise in the regional model masks any possible correlation effects. This lack of difference can be confirmed through plots of the independent- and correlated-site residuals. Figure 4 presents a scatter plot of the independent- and correlated-model mean residuals for region 1 (for the IPO- epoch), which is representative of the scatter plots for the other regions (and the IPO+ epoch and also for the standard-deviation residual). The plots show some scatter, but the majority of residuals plot near the `45° line'. The differences in the residuals between the independent- and correlated-site analyses are judged sufficiently small to warrant the use of the independent-site analysis for the remainder of this study.

Finally, note that this Gibbs sampler-based approach did not suffer from the sampling difficulties that were experienced with the Metropolis algorithm-based approach of Micevski et al. [2003].

6. PREDICTION AT A NEW SITE

The regional posterior distribution (the Gibbs samples) can be used to predict the flood frequency distribution at a new site, which may be ungauged or gauged, where this site was not used in the development of the regional model. The procedure is derived using Monte Carlo importance (particle) sampling, with the final algorithm summarised (for a gauged site with data G):

(1) Randomly sample the IPO epoch: $IPO^{(j)} \leftarrow P(IPO)$, where $P(IPO)$ is the probability of sampling an IPO epoch (IPO- or IPO+).
(2) Randomly sample $\boldsymbol{\theta}_S^{(j)} = \{\mu^{(j)}, \log \sigma^{(j)}\}$ from the regional model in two steps (sample $\boldsymbol{\theta}_R^{(j)}$ using the Gibbs samples from the appropriate IPO epoch and then sample $\boldsymbol{\theta}_S^{(j)}$ (see [2] and [3])).
(3) Compute the T-year flood: $q_T^{(j)} = \mu^{(j)} + z_T\sigma^{(j)}$, where z_T is the standard normal deviate.

(4) Compute the (unnormalised) particle weight: $p(G| \theta_S^{(j)})$.

(5) Repeat steps (1)-(4) for all j=1,…,N posterior samples.

(6) Compute the normalised particle weights for all j=1,…,N: $w^{(j)} = p(G| \theta_S^{(j)}) / \Sigma_{k=1}^{N} p(G| \theta_S^{(k)})$.

(7) Sort q_T and compute the (sorted) cumulative particle weights $W^{(j)}$ for all j=1,…,N.

(8) Extract probability limits using the cumulative particle weights.

Note, if the ungauged distribution is required, then the steps involving particle weights can be ignored.
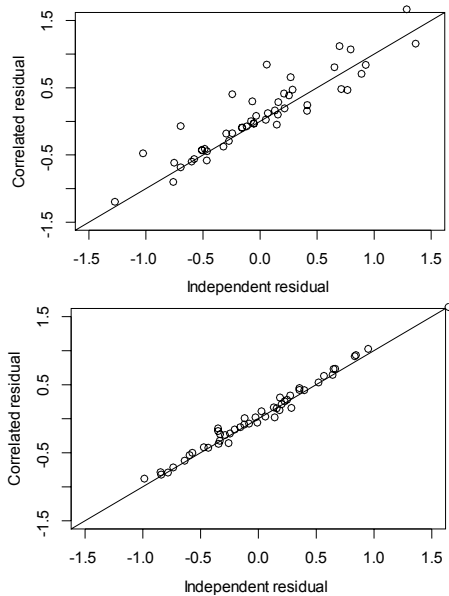
Figure 4. Region 1 residual plots for the site mean comparing (top) independent- and (bottom) correlated-site analyses (IPO- epoch).
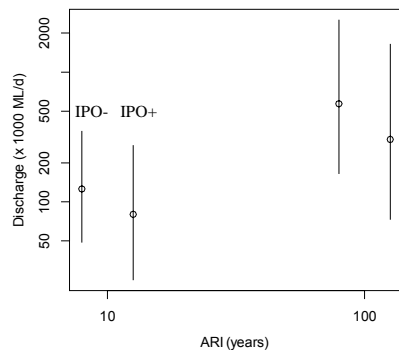

Figure 5. Posterior distribution of selected flood quantiles for an ungauged site in region 2 (90% limits are shown).

The algorithm is illustrated for an ungauged site in Figure 5, which gives the at-site flood frequency curves for the IPO- and IPO+ epochs. The figure corresponds to a site in region 2 with an observed IPO modulation of flood risk of about 1.7, with the mean discharge being commensurately larger during the IPO-epoch. The probability limits for the two IPO epochs largely overlap, which indicates that the uncertainty and noise of the regional model have masked any possible IPO-related difference.

This procedure is now used to examine the improvements, if any, from augmenting the regional flood distribution with some at-site gauged data. Gauged data (5 or 10 years) is randomly sampled from an existing site's full gauged record and is used to augment the regional distribution. Figure 6 compares 2 sites, from regions 1 and 2, for the following flood frequency distributions: G(full) — full gauged record (of length x years); G(10) — shortened 10-year gauged record; UG — `ungauged' record (regional distribution only); UG(5) — ungauged record + 5 years of gauged data; UG(10) — ungauged record + 10 years of gauged data. Note that the same gauged data is used for both G(10) and UG(10), with UG(5) using the first 5 of 10 years.

The top panel has a shortened (10-year) gauged record which is consistent with the long-run (full) record. The augmented distribution had reduced variability (smaller probability limits), even with a noisy regional model. The bottom panel has a shortened record which is inconsistent with the long-run record (most years are sampled from the IPO+ epoch). The augmented distribution had slightly increased variability; however, it remains consistent with the full record and the resultant bias is less than that associated with the 10-year record alone. These results suggest that combining regional and gauged data may have significant benefits when used to predict at-site flood frequency distributions with limited gauged data.

7.   CONCLUSIONS

A Bayesian hierarchical flood regionalisation procedure of Micevski et al [2003] was further developed in this paper. The use of regional flood methods was motivated by an attempt to help overcome the possible bias in long-run flood risk associated with a nonhomogeneous flood record, such as that associated with the IPO modulation of flood risk. The regional model was used to analyse IPO-stratified flood data from eastern Australia. The regional model could not detect any IPO-related difference, most likely due to considerable noise in the regional model. An algorithm to

combine regional and gauged information was presented. Of importance is the rigorous handling of uncertainty in the regional and site models. The results suggest that the use of a regional model may help protect against bias in long-run flood risk at sites with short records which largely sample from one IPO epoch.
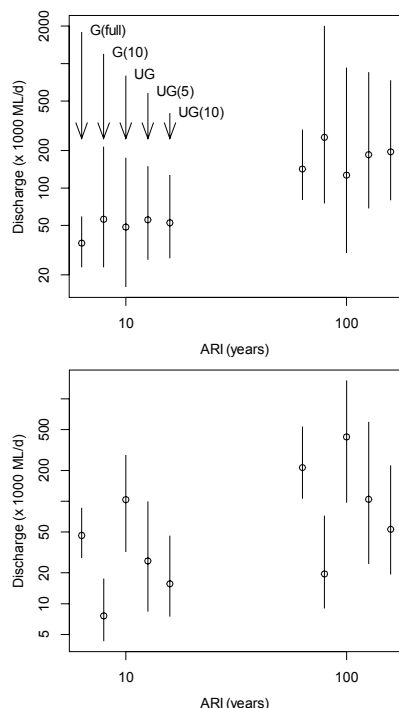

Figure 6. Posterior distribution of selected flood quantiles for gauged sites in regions (top) 1 and (bottom) 2 (90% limits are shown).

8. ACKNOWLEDGEMENTS

9. REFERENCES

Erskine, W.D. and R.F. Warner, Geomorphic effects of alternating flood- and drought-dominated regimes on a NSW coastal river, in Warner, R.F. (editor), *Fluvial Geomorphology of Australia*, Academic Press, Sydney, 223–244, 1988.

Folland, C.K., J.A. Renwick, M.J. Salinger and A.B. Mullan. Relative influences of the Interdecadal Pacific Oscillation and ENSO on the South Pacific Convergence Zone. *Geophysical Research Letters*, 29(13), 1643, 2002.

Franks, S.W, Identification of a change in climate state using regional flood data, *Hydrology and Earth System Sciences*, 6(1), 11–16, 2002.

Franks, S.W. and G. Kuczera, Flood frequency analysis: Evidence and implications of secular climate variability, New South Wales, *Water Resources Research*, *38*(5), 1062, 2002.

Gelman, A., J.B. Carlin, H.S. Stern and D.B. Rubin, *Bayesian Data Analysis*, Chapman & Hall, London, 1995.

Kiem, A.S., S.W. Franks and G. Kuczera, Multi-decadal variability of flood risk, *Geophysical Research Letters*, *30*(2), 1035, 2003.

Mantua, N.J., S.R. Hare, Y. Zhang, J.M. Wallace and R.C. Francis. A Pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the American Meteorological Society*, 78(6), 1069-1079, 1997.

Micevski, T., S.W. Franks and G. Kuczera, Multidecadal variability in coastal eastern Australian flood data, *Journal of Hydrology*, 327(1-2), 219-225, 2006.

Micevski, T., G. Kuczera and S.W. Franks, Bayesian flood regionalisation for New South Wales, paper presented at 28th International Hydrology and Water Resources Symposium, Institution of Engineers Australia, Wollongong, NSW, November 10–14, 2003.

Power, S., T. Casey, C. Folland, A. Colman and V. Mehta, Inter-decadal modulation of the impact of ENSO on Australia, *Climate Dynamics*, *15*(5), 319–324, 1999.

Stedinger, J.R. and G.D. Tasker, Regional hydrologic analysis 1. Ordinary, weighted and generalized least squares compared. *Water Resources Research*, *21*(9), 1421–1432, 1985.

A Bayesian hierarchical regional flood model

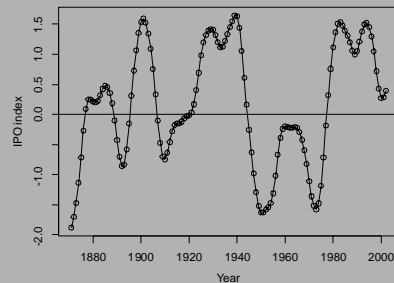**Tom Micevski**, George Kuczera, Stewart W. Franks

The University of Newcastle

5 December 2006

Motivation

- Regionalisation
  - Estimate flood at a new site using records from 'nearby' gauged sites
  - Used in 3 circumstances
    - Ungauged site
    - Poorly-gauged site (short record)
    - Gauged sites affected by nonhomogeneity of flood risk
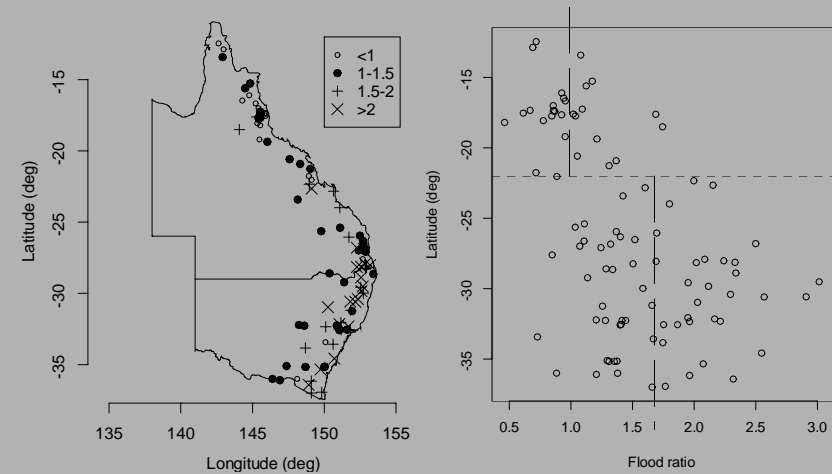- We will concentrate on 3rd point

Aims

- Develop a regional flood model
  - To allow for dependence on IPO
  - To allow for missing data
  - To incorporate spatial correlation
    - Loss of information?
  - To allow for (and quantify) uncertainty
- At both ungauged and gauged sites

Regional model

- Mechanism used to transfer info from gauged sites to new (ungauged/gauged) site
- Transfer is noisy
  - Data errors
  - Sampling variability
  - Use a probability model
- Bayesian hierarchical model
  - Conditional modelling
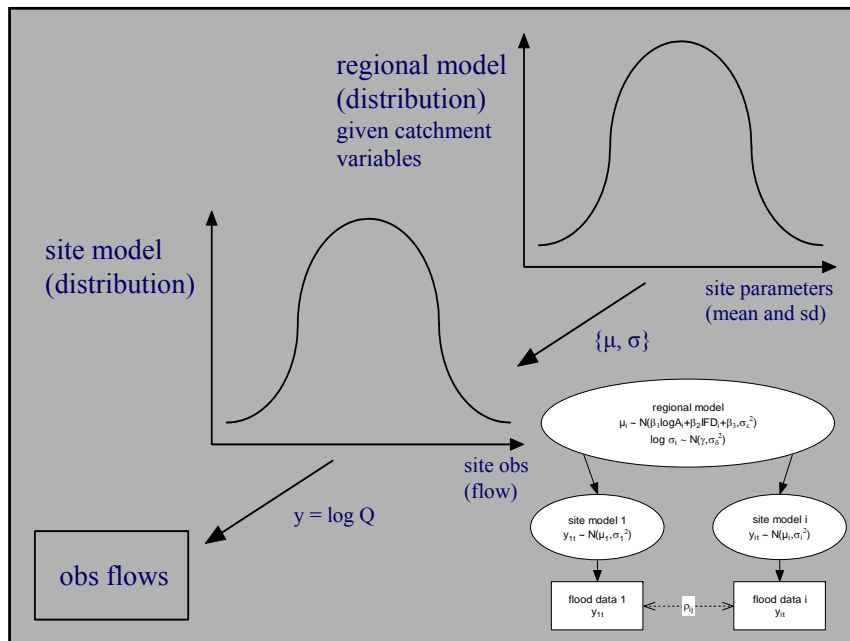
Hierarchical model for each IPO epoch

- Observed data ← site distribution
  $$\log Q = y_{it} \sim N(\mu_i, \sigma_i^2)$$
- Site mean and sd ← regional distribution
  $$\mu_i \sim N(\beta_1 \log A_i + \beta_2 IFD_i + \beta_3, \sigma_\varepsilon^2)$$
  $$\log \sigma_i \sim N(\gamma, \sigma_\delta^2)$$
- Correlation
  $$\rho_{ij} = \exp(-[d_{ij} / A]^B)$$
  $$\mathbf{y}_t \sim N(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$$
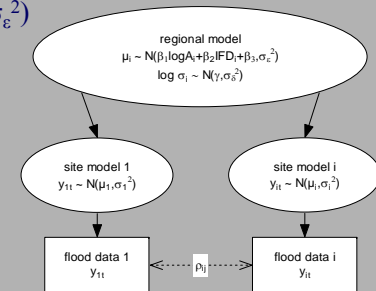
Model calibration using Gibbs sampler

- Bayesian approach

$$p(\Theta \mid Y) = \frac{f(Y \mid \Theta)\, p(\Theta)}{p(Y)} \propto f(Y \mid \Theta)\, p(\Theta)$$

- Conditional sampling (updating) using Gibbs sampler
  - Partition entire parameter vector: $\Theta = \{\theta_R, \theta_S\}$
  - $\theta_R = \{\sigma_\varepsilon^2, \beta, \sigma_\delta^2, \gamma\}$
  - $\theta_S = \{\mu, \log \sigma\}$
- Gibbs sampler algorithm:
  - (0) Assign starting value $\theta_S^{(0)}$
  - (1) Randomly sample regional subvector:
    $\theta_R^{(j)} \leftarrow p(\theta_R | \theta_S^{(j-1)}, Y) = p(\theta_R | \theta_S^{(j-1)})$
  - (2) Randomly sample site subvector:
    $\theta_S^{(j)} \leftarrow p(\theta_S | \theta_R^{(j)}, Y) = p(Y | \theta_S)\, p(\theta_S | \theta_R^{(j)}) / p(Y | \theta_R^{(j)})$
  - Repeat steps (1) and (2) many times (say 10000)

Study regions
(geographical)

5

5
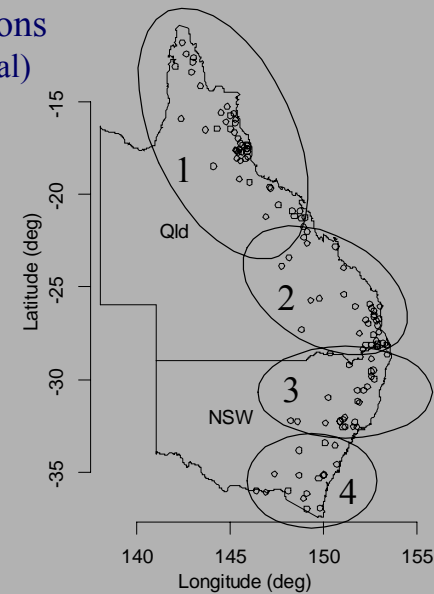
Results

- Independent-site analysis
- Differences between individual IPO+/- pars not 'statistically significant'
- Noise in regional model
  - Large regions
  - Few sites
  - Varying size of sites
- Swamps differences
 - Due to IPO dependence
- Equivalent record lengths
  - 2–12 years

Par	IPO-		IPO+	
Region 1				
β_1	0.67	(0.06)	0.63	(0.06)
β_2	0.13	(0.03)	0.13	(0.03)
β_3	4.72	(0.65)	4.85	(0.59)
σ^2_ε	0.29	(0.08)	0.25	(0.07)
γ	-0.21	(0.07)	0.03	(0.05)
σ^2_δ	0.19	(0.05)	0.11	(0.03)
Region 2				
β_1	0.52	(0.06)	0.52	(0.06)
β_2	0.27	(0.06)	0.29	(0.06)
β_3	4.24	(0.74)	3.74	(0.79)
σ^2_ε	0.19	(0.09)	0.33	(0.10)
γ	0.31	(0.04)	0.25	(0.04)
σ^2_δ	0.02	(0.02)	0.05	(0.02)

Prediction at a new (gauged) site

- Site flood frequency curve and its 90% probability limits
  - Derived from regional model (Gibbs samples)
  - Allows for (incorporates) variability of both the regional model and its parameters
- Rigorous method to combine regional and gauged information
  - Monte Carlo importance (particle) sampling

Prediction at a new (gauged) site
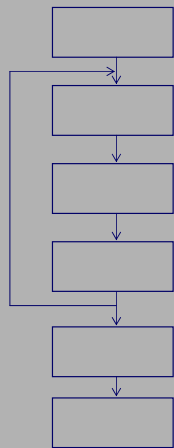
NEW SITE: $\mathbf{g}$ = gauged record, $\mathbf{x}$ = {log A, IFD, …}
* regional posterior (previous calibration: 2×10000)
* IPO probs: Pr(being in either IPO epoch) ≈ 0.50

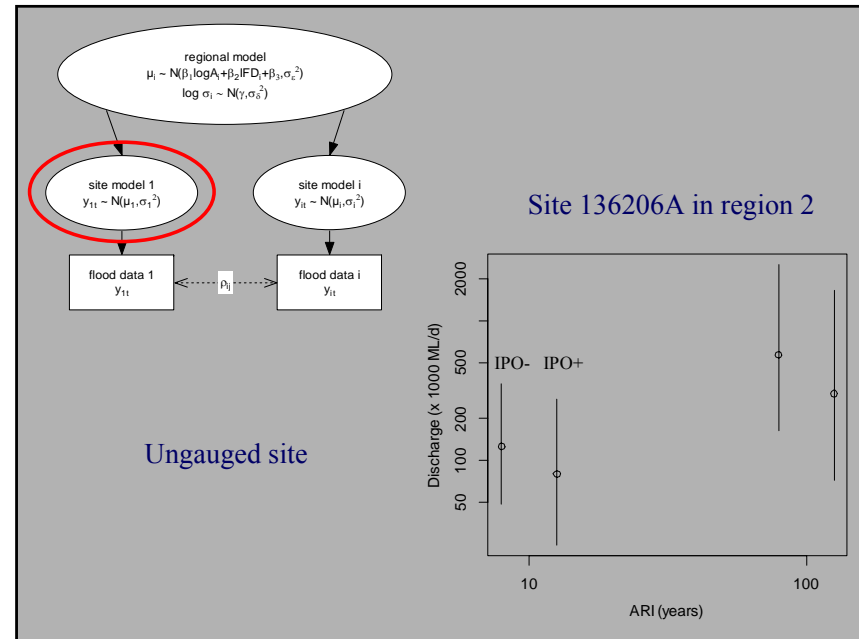Sample IPO epoch: $\text{IPO}^{(j)} \leftarrow p(\text{IPO})$

Sample $\boldsymbol{\theta}_S^{(j)}$ from regional model for current IPO epoch
   2 steps: $\boldsymbol{\theta}_R^{(j)} \leftarrow p(\boldsymbol{\theta}_R|Y, \text{IPO}^{(j)})$; $\boldsymbol{\theta}_S^{(j)} \leftarrow p(\boldsymbol{\theta}_S|\boldsymbol{\theta}_R^{(j)}, \mathbf{x})$
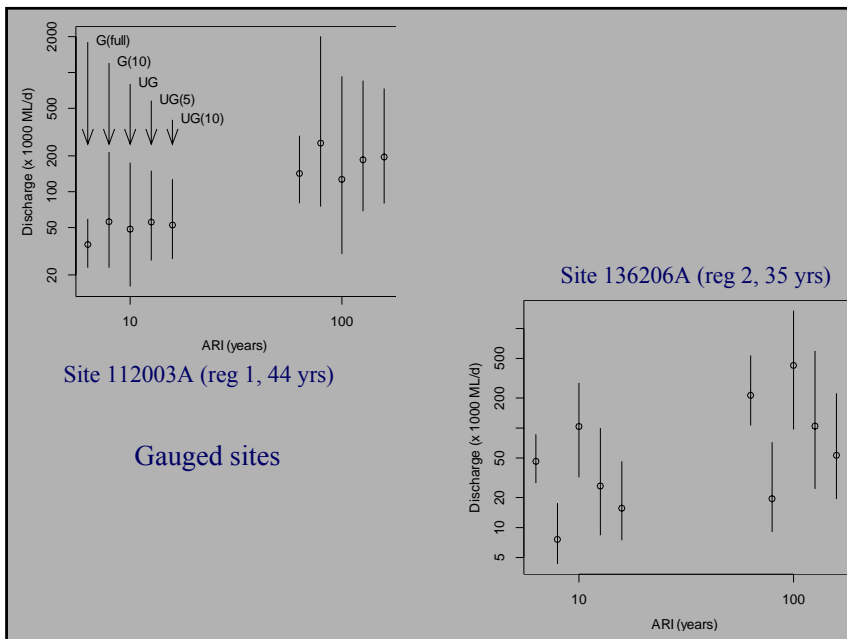
Compute T-year flood: $\log Q_T^{(j)} = \mu^{(j)} + z_T \sigma^{(j)}$
Compute particle weight: $p(\mathbf{g}|\boldsymbol{\theta}_S^{(j)}) = \Pi_k N(g_k| \mu^{(j)}, \sigma^{(j)2})$

Normalise particle weights: $w^{(j)} = p(\mathbf{g}|\boldsymbol{\theta}_S^{(j)}) / [\Sigma\, p(\mathbf{g}|\boldsymbol{\theta}_S)]$

Extract quantiles for Q_T (using cumulative particle weights)

regional model
$\mu_i \sim N(\beta_1 \log A_i + \beta_2 IFD_i + \beta_3, \sigma_\varepsilon^2)$
$\log \sigma_i \sim N(\gamma, \sigma_\delta^2)$

site model 1
$y_{1t} \sim N(\mu_1, \sigma_1^2)$

site model i
$y_{it} \sim N(\mu_i, \sigma_i^2)$

flood data 1
y_{1t}

ρ_{1i}

flood data i
y_{it}

Site 136206A in region 2

IPO-   IPO+

Ungauged site

Discharge (x 1000 ML/d)

ARI (years)

7

Conclusions

- Nonhomogeneity (IPO dependence) of flood record in eastern Australia
  - Possible bias in long-run flood risk
  - Overcome using regional flood distribution
    - Use many sites with adequate samples from both IPO epochs

Gauged sites

Site 112003A (reg 1, 44 yrs)

Site 136206A (reg 2, 35 yrs)

8

Conclusions (cont.)

- Bayesian hierarchical regional flood methodology can handle
  - IPO dependence of flood risk
  - Intersite correlation
  - Missing data
  - *Rigorous* estimate of predictive uncertainty at both ungauged and gauged sites!

Thank you

- Questions?