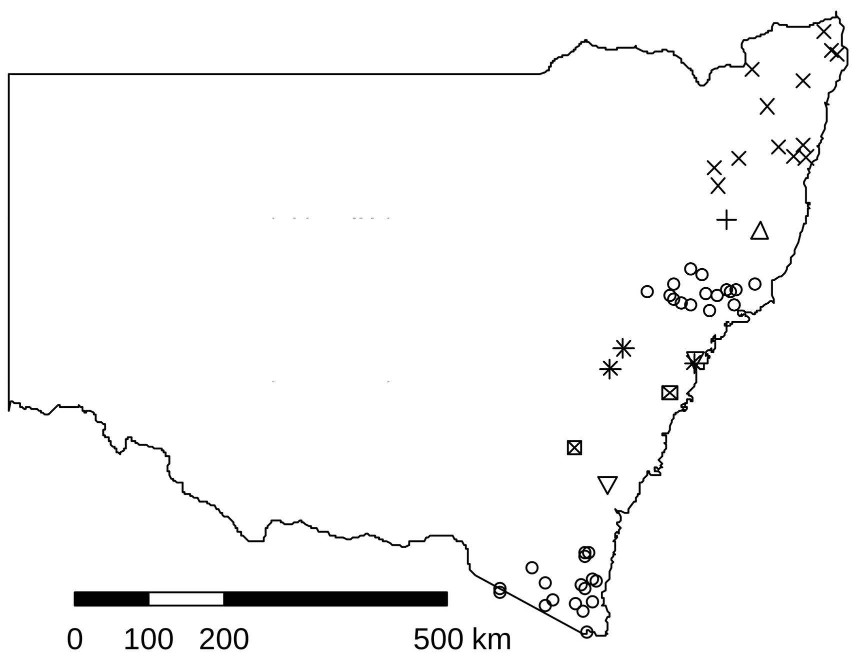
Regional Flood Frequency Analysis using Bayesian Generalized Least Squares in a Region-of-Influence Context

Tom Micevski[1], André Hackelbusch[1,2], Khaled Haddad[3], George Kuczera[1] and Ataur Rahman[3]

1. Introduction

As part of the revision of Australian Rainfall and Runoff, procedures for estimation of flood peaks discharge at ungauged catchments were reviewed. This study assesses the performance of the fusion of the region of influence (RoI) approach with Bayesian generalised least squares (GLS) regional flood frequency regression. The GLS procedure regionalises the mean, standard deviation and skewness of the log-Pearson III (LP3) distribution with simultaneous consideration of model and sampling error. The RoI concept is used to reduce unaccounted-for heterogeneity in the GLS regression. A case study was undertaken for 55 high-quality data catchments located in eastern New South Wales, Australia.



2. Why Linear Regression?

The index flood method assumes homogeneity in the growth curve across all sites. However, the case study region strongly failed the Hosking and Wallis (1993) homogeneity test invalidating application of the index flood method.

No of sites	H(1)	H(2)	H(3)	No of discordant sites
55	10.9	8.3	5.8	2
53	10.6	8.0	5.9	0

To better account for regional heterogeneity a linear regression approach was adopted.

$$\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta} + \boldsymbol{\varepsilon}$$

where $\hat{\mathbf{y}}$ is a vector of an estimated hydrologic variable at N sites, $\mathbf{X}$ is a NxP matrix of catchment attributes, $\boldsymbol{\beta}$ is a P-vector of parameters, $\boldsymbol{\varepsilon}$ is a N-vector of normally distributed sampling errors and $\boldsymbol{\delta}$ is a N-vector of normally distributed model errors.

3. Why LP3 Parameter Bayesian GLS?

In regional frequency analysis the estimated hydrologic variable $\hat{\mathbf{y}}$ can be either a quantile or a parameter from the underlying flood probability model.

We chose to regress catchment attributes against flood probability model parameters in conjunction with Bayesian GLS for the following reasons:

1. Flood quantiles increase smoothly with increasing average recurrence interval (ARI);
2. It is straightforward to combine any at-site flood information with regional estimates using the approach described by Micevski and Kuczera (2009) to produce more accurate quantile estimates; and
3. Quantiles can be estimated for any ARI in the range of interest.

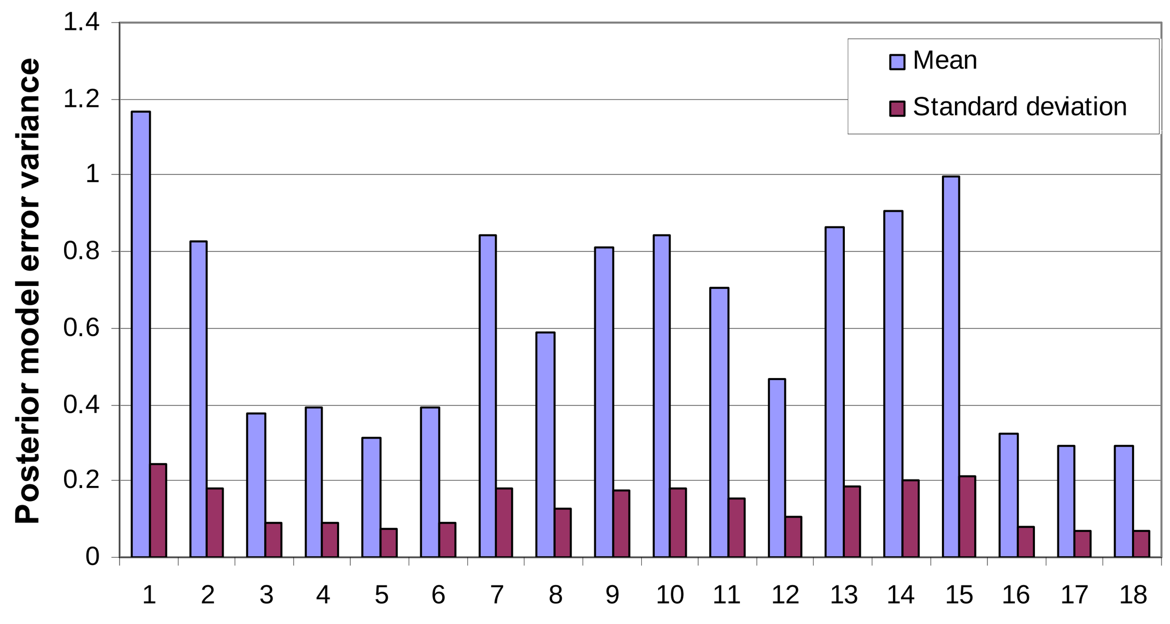
4. Selection of Catchment Attributes

The following catchment attributes were evaluated:

- A catchment area [km^2]
- $I_{h,y}$ rainfall intensity [mm/h] for y-year ARI and duration of h hours.
- R: mean annual rainfall [mm/yr]
- E mean annual areal evaporation [mm/yr]
- S_D stream density [km/km^2]
- S_L stream length [km]
- S_{1085} slope of the central 75% of mainstream [m/km]
- F fraction of catchment area under forest [-]

A search, similar in concept to stepwise regression, was conducted to identify the subset of catchment attributes that minimizes model error variance

The Figure shows the posterior mean and standard deviation of the model variance for the LP3 mean parameter. Combination 1 used no catchment attributes (only a constant term), combination 2 used only area, while combinations 3 to 15, 16 to 17, and 18 used two, three, and four attributes respectively. The combinations with the smallest model error variance were 5, 17, and 18. Of these combinations, 5 was selected being the simplest with only two attributes, area and 50-year, 12-hour rainfall intensity.



The optimal linear models for the LP3 parameters were:

LP3 mean: $\mu = \beta_0 + \beta_1 \log(A) + \beta_2 \log(I_{12,50})$

LP3 standard deviation: $\sigma = \beta_0$

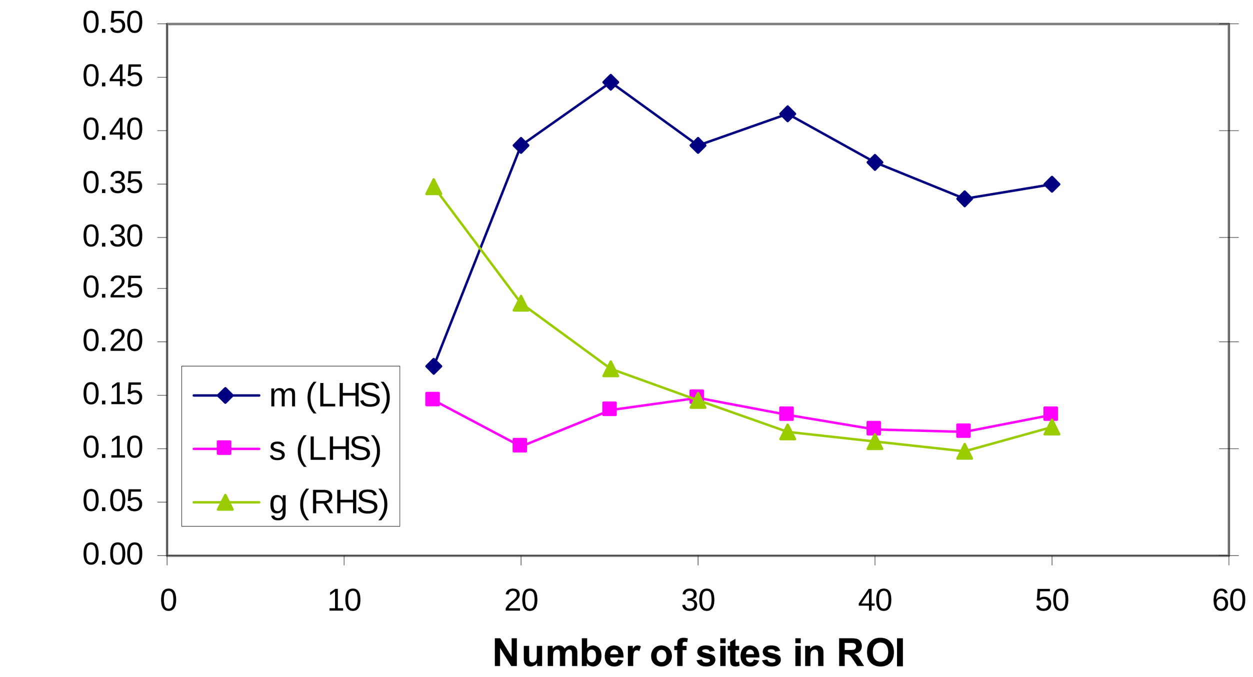
LP3 skew: $\gamma = \beta_0$

On first glance, the absence of catchment attributes for LP3 standard deviation and skew seems to suggest the index flood assumptions are supported by the data, in contradiction with the Hosking-Wallis test. In fact, index flood assumes the same standard deviation and skew at every site, whereas the linear regression model allows for random variation about β_0.

5. Selection of Region of Influence to Minimize Model Error

The ROI approach uses the physical distance between sites as the distance metric. It starts with the nearest 15 sites and then adds the next 5 closest sites until the minimum predictive variance is found.

The Figure illustrates how predictive variance changes with the size of the RoI. For the LP3 mean, the smallest RoI was the preferred, highlighting that unaccounted-for heterogeneity increases with increasing RoI. For the LP3 skew, the opposite occurs. Because sampling error dominates model error, the nearest 45 (out of 54) sites defined the best RoI.

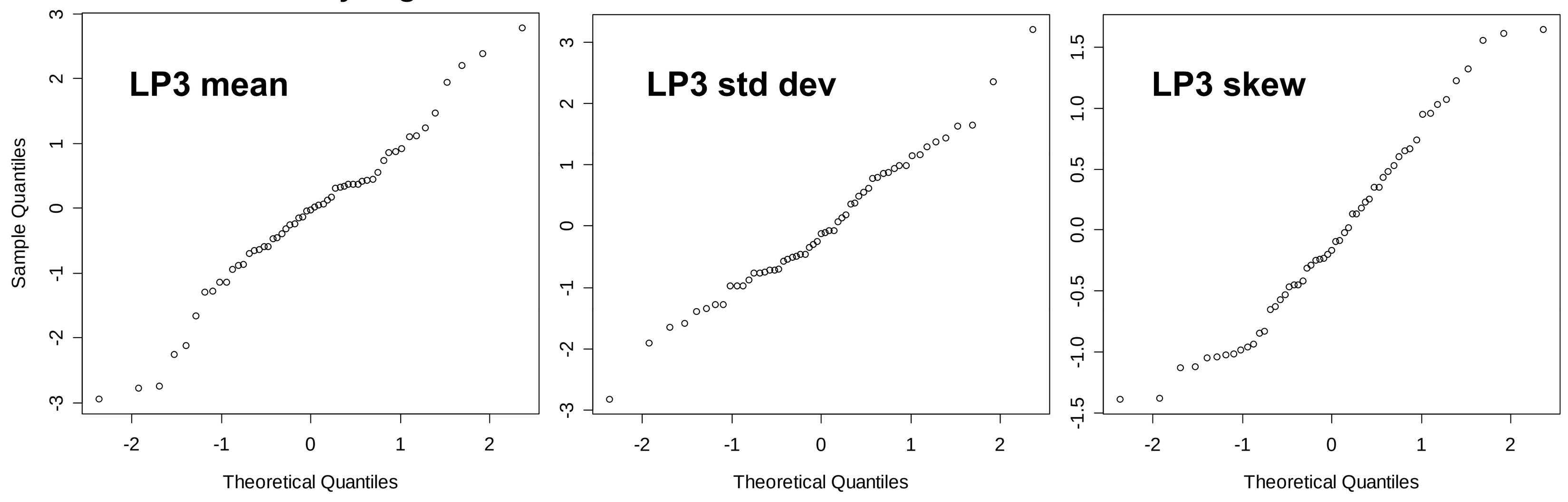


On average, RoIs for the LP3 mean had 23 sites, 33 sites for the standard deviation, and 46 sites for the skewness.

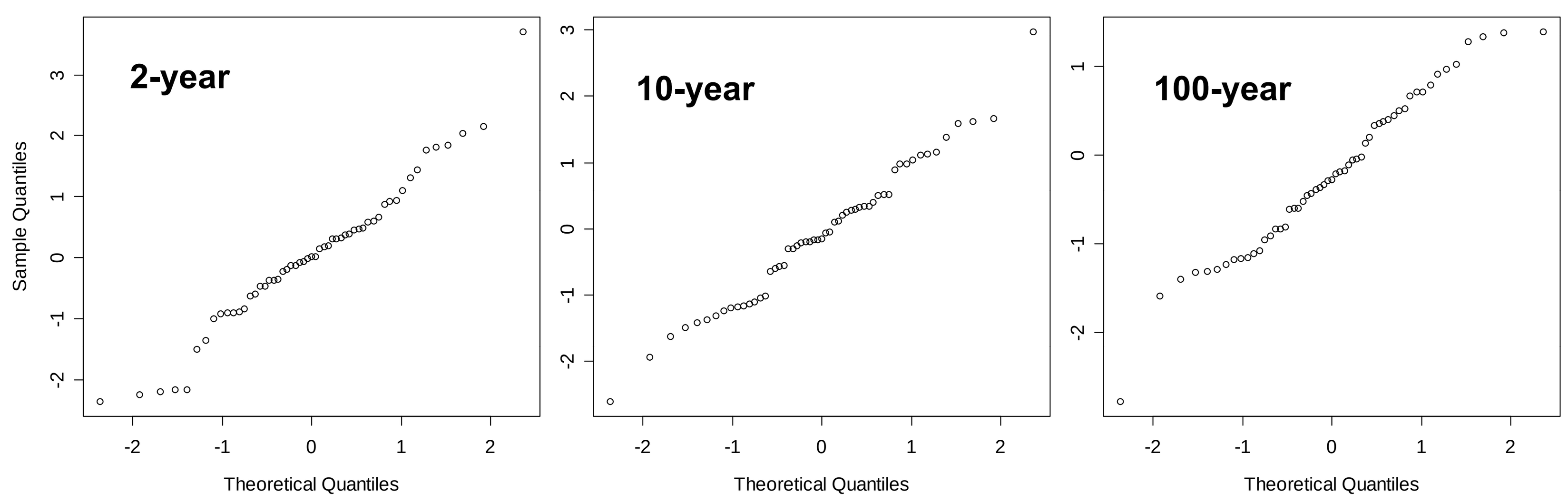
6. Can GLS Regression Be Trusted?

Regional frequency analysis must be robust. One-at-a-time cross validation was used to establish if there were any genuine outliers.

Normal QQ plots of standardized residuals for LP3 mean, standard deviation and skew reveal no statistically significant outliers.

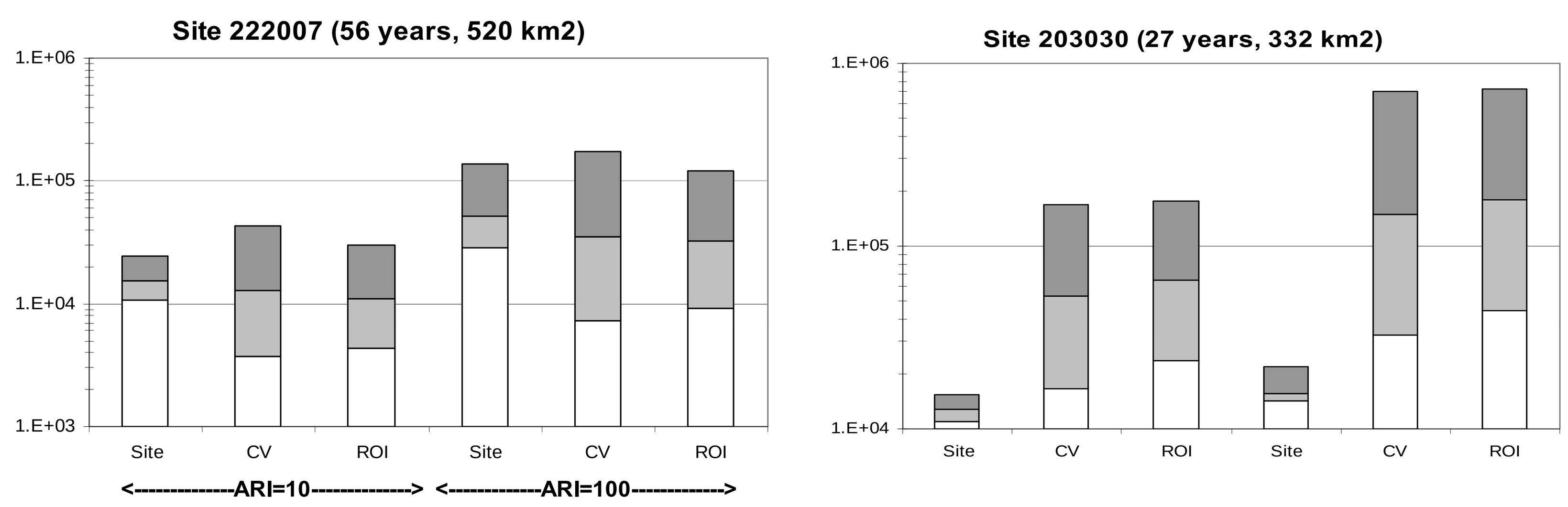


Likewise, normal QQ plots of standardized 2-, 10- and 100-year quantiles reveal no outliers.



7. Prediction at Ungauged Sites

The Figures illustrate two cases of 10 and 100-year flood quantile prediction at ungauged sites. The first shows a typical result where the at-site quantile distribution overlaps with the regional quantile distributions. The use of RoI to reduce model error results in tighter prediction limits compared with use of all sites in the regional analysis. The second presents the worst case encountered in the study where the 90% prediction limits do not overlap, reminding us of the intrinsic error in regional quantile prediction.



8. Conclusions

The fusion of the regional of influence approach with Bayesian GLS linear regression using LP3 parameters exploits the strengths of each individual approach. The case study for eastern New South Wales suggests the following findings:

- The method is robust. As no outliers were detected, the GLS predictive uncertainty is expected to provide an honest, reliable assessment of uncertainty.
- Coupling RoI with GLS regression reduces unaccounted-for heterogeneity when compared with GLS regression on all sites in the region. As a result, this produces more accurate quantile estimates at ungauged sites.
- Because GLS distinguishes between sampling and model error, the search for optimal catchment attributes can be guided by explicit model error minimization.
- The use of Bayesian GLS with LP3 parameters provides a rigorous method for combining regional estimates with any gauged or historic site flood data. Limited site data can significantly reduce uncertainty because the greatest regional model error is associated with the LP3 mean.

[1]School of Engineering, University of Newcastle, Callaghan NSW 2308, Australia
[2]Faculty of Civil- and Environmental Engineering, Ruhr-University Bochum, 44870 Bochum, Germany
[3]School of Engineering, University of Western Sydney, Penrith South DC NSW 1797, Australia