HYDROLOGICAL PROCESSES
Hydrol. Process. **26**, 1008–1021 (2012)
Published online 29 June 2011 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/hyp.8189

Regional flood frequency analysis using Bayesian generalized least squares: a comparison between quantile and parameter regression techniques

Khaled Haddad,[1] Ataur Rahman[1]* and Jery R. Stedinger[2]

[1] *School of Engineering, University of Western Sydney, Penrith, Sydney, Australia*
[2] *School of Civil and Environmental Engineering, Cornell University, Ithaca, NY 14850, USA*

Abstract:

Regression-based regional flood frequency analysis (RFFA) methods are widely adopted in hydrology. This paper compares two regression-based RFFA methods using a Bayesian generalized least squares (GLS) modelling framework; the two are quantile regression technique (QRT) and parameter regression technique (PRT). In this study, the QRT focuses on the development of prediction equations for a flood quantile in the range of 2 to 100 years average recurrence intervals (ARI), while the PRT develops prediction equations for the first three moments of the log Pearson Type 3 (LP3) distribution, which are the mean, standard deviation and skew of the logarithms of the annual maximum flows; these regional parameters are then used to fit the LP3 distribution to estimate the desired flood quantiles at a given site. It has been shown that using a method similar to stepwise regression and by employing a number of statistics such as the model error variance, average variance of prediction, Bayesian information criterion and Akaike information criterion, the best set of explanatory variables in the GLS regression can be identified. In this study, a range of statistics and diagnostic plots have been adopted to evaluate the regression models. The method has been applied to 53 catchments in Tasmania, Australia. It has been found that catchment area and design rainfall intensity are the most important explanatory variables in predicting flood quantiles using the QRT. For the PRT, a total of four explanatory variables were adopted for predicting the mean, standard deviation and skew. The developed regression models satisfy the underlying model assumptions quite well; of importance, no outlier sites are detected in the plots of the regression diagnostics of the adopted regression equations. Based on 'one-at-a-time cross validation' and a number of evaluation statistics, it has been found that for Tasmania the QRT provides more accurate flood quantile estimates for the higher ARIs while the PRT provides relatively better estimates for the smaller ARIs. The RFFA techniques presented here can easily be adapted to other Australian states and countries to derive more accurate regional flood predictions. Copyright © 2011 John Wiley & Sons, Ltd.

KEY WORDS regional flood frequency; Bayesian method; quantile regression; parameter regression; design floods

Received 11 November 2010; Accepted 23 May 2011

INTRODUCTION

Flood quantile estimation in ungauged catchments is a common problem in hydrology. Regional flood frequency analysis (RFFA) is often used for this purpose, which is to 'trade space for time' (Hosking and Wallis, 1997). Regression-based methods are widely used in RFFA which is based on the concept that spatial variations in flood flow statistics are closely related with variations in regional catchment and climatic characteristics (Gupta *et al*., 2006; Pandey and Nguyen, 1999; Nezhad *et al*., 2010). The most common form of the regression approach is to develop a regression equation for a flood quantile of interest, known as the quantile regression technique (QRT) (Benson, 1962; Thomas and Benson, 1970). The United States Geological Survey has adopted the QRT as the standard RFFA method since the 1960s (Gupta *et al*., 1994).

Hydrologists commonly use ordinary least squares (OLS) estimators that are appropriate and statistically efficient if the flow records are of equal length and if concurrent flows between any pair of stations are uncorrelated (Tasker *et al*., 1986). These are often violated with regional annual maximum flood series data. To overcome the problems with the OLS regression, Stedinger and Tasker (1985, 1986) developed a GLS model that accounts for the differences in at-site record lengths and inter-site correlation among at-site estimators. Stedinger and Tasker (1985, 1986) showed in a Monte Carlo simulation that the GLS estimators provide model regression parameters with smaller mean-squared errors than the competing OLS estimators, provide relatively unbiased estimates of the variance of the regression parameters and results in a more accurate estimate of the regression model error. GLS regression has been widely adopted in hydrology (Tasker and Stedinger, 1989; Madsen *et al*., 1995; Madsen and Rosbjerg, 1997; Kroll and Stedinger, 1999; Reis *et al*., 2005; Eng *et al*., 2005; Griffis and Stedinger, 2007; Gruber and Stedinger, 2008; Hackelbusch *et al*., 2009; Micevski and Kuczera, 2009).

* Correspondence to: Ataur Rahman, School of Engineering, University of Western Sydney, Penrith, Sydney, Australia.
E-mail: a.rahman@uws.edu.au

Reis *et al*. (2005) and Gruber *et al*. (2007) introduced a Bayesian analysis of the GLS model which provides more accurate measure of the model error variance and a more realistic description of the possible values of the model error variance in cases where the method of moments estimator of the model error variance as described by Stedinger and Tasker (1985) may be zero or close to it; this occurs when sampling errors dominate the regional analysis.

As an alternative to the QRT, the parameters of a probability distribution can be regressed against the explanatory variables (Tasker and Stedinger, 1989; Madsen *et al*., 2002). In the case of the log Pearson Type 3 (LP3) distribution, regression equations can be developed for the first three moments, i.e. the mean, standard deviation and skewness of the logarithms of annual maximum flood series. For an ungauged catchment, these equations can then be used to predict the mean, standard deviation and skewness to fit an LP3 distribution. This method here is referred to as 'parameter regression technique' (PRT). The advantages of the PRT are that the estimated flood quantiles from the fitted distribution increase consistently with average recurrence intervals (ARI) and the flood quantiles for any ARI (within the range of validity of the method) can be estimated. However, there has been little research on comparison of PRT with QRT in RFFA.

The objective of this paper is to compare the QRT and PRT approaches under a Bayesian GLS modelling framework. A method is presented for selecting the most appropriate set of explanatory variables in the GLS regression by employing a number of statistics such as the model error variance, average variance of prediction, Bayesian information criterion, Akaike information criterion and pseudo coefficient of determination values. In previous applications of the GLS regression, explanatory variables selected by the OLS regression using stepwise regression have generally been adopted. The method presented here provides an improvement in selecting the right set of predictor variables in the GLS regression. A number of statistics, diagnostic plots and one-at-a-time cross validation approach have been adopted in evaluating the regression models. The methods have been applied to a data set from the state of Tasmania in Australia.

METHODS

Generalized least squares model description

The GLS regression assumes that the hydrological variable of interest (e.g. a flood quantile or a parameter of the LP3 distribution) denoted by y_i for a given site i can be described by a function of catchment characteristics (explanatory variables) with an additive error (Reis *et al*., 2005; Griffis and Stedinger, 2007):

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} + \delta_i, \;\; i = 1, 2, \ldots, n \tag{1}$$

where $X_{ij} (j = 1, \ldots, k)$ are explanatory variables, β_j is the regression coefficient, δ_i is the model error which is assumed to be normally and independently distributed with model error variance σ_δ^2 and n is the number of sites in the region. In all cases only an at-site estimate of y_i denoted as $\hat{y}_i$ is available. To account for the error in this data, a sampling error η_i must be introduced into the model so that:

$$\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta} + \boldsymbol{\delta} = \mathbf{X}\boldsymbol{\beta} + \varepsilon \text{ where } \hat{y}_i = y_i + \eta_i; \quad i = 1, 2, \ldots, n \tag{2}$$

Thus, the observed regression model error $\boldsymbol{\varepsilon}_i$ is the sum of the model error δ_i and the sampling errors η_i. The total error vector $(\Lambda(\sigma_\delta^2))$ has mean zero and a covariance matrix:

$$E\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\right] = \boldsymbol{\Lambda}(\sigma_\delta^2) = \sigma_\delta^2 \mathbf{I} + \sum(\hat{\mathbf{y}}) \tag{3}$$

where $\sum(\hat{\mathbf{y}})$ is the covariance matrix of the sampling errors in the sample estimators of the flood quantiles or the parameters of the LP3 distribution, $\mathbf{I}$ is a $(n \times n)$ identity matrix. The covariance matrix for η_i depends on the record length available at each site and the cross correlation among floods at different sites. Therefore, the observed regression model errors are a combination of time-sampling error η_i and an underlying model error δ_i.

In this regional framework, σ_δ^2 can be viewed as a heterogeneity measure. Madsen *et al*. (1997, 2002) showed that the regional average GLS estimator is a general extension of the record-length-weighted average commonly applied in the index-flood procedure; however, the record-length-weighted average estimator neglects intersite correlation and regional heterogeneity (Stedinger *et al*., 1993; Stedinger and Lu, 1995).

The GLS estimator of $\boldsymbol{\beta}$ and its respective covariance matrices for known σ_δ^2 are given by:

$$\hat{\beta}_{\mathrm{GLS}} = \left[\mathbf{X}^T \boldsymbol{\Lambda}(\sigma_\delta^2)^{-1} \mathbf{X}\right]^{-1} \mathbf{X}^T \Lambda(\sigma_\delta^2)^{-1} \hat{\mathbf{y}} \tag{4}$$

$$\Sigma[\hat{\beta}_{\mathrm{GLS}}] = \left[\mathbf{X}^T \boldsymbol{\Lambda}(\sigma_\delta^2)^{-1} \mathbf{X}\right]^{-1} \tag{5}$$

The model error variance σ_δ^2 can be estimated by either generalized method of moments (MOM) or maximum likelihood (ML) estimators as described by Stedinger and Tasker (1986). The MOM estimator is determined by iteratively solving Equation (6) along with the generalized residual mean square error equation:

$$(\hat{\mathbf{y}} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{GLS}})^T [\hat{\sigma}_\delta^2 \mathbf{I} + \sum(\hat{\mathbf{y}})]^{-1} (\hat{\mathbf{y}} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}) = n - (k + 1) \tag{6}$$

In some situations, the sampling covariance matrix explains all the variability observed in the data, which means the left-hand side of Equation (6) will be less than $n - (k + 1)$ even if $\hat{\sigma}_\delta^2$ is 0. In these circumstances, the MOM estimator of the model error variance is generally taken to be 0 (Stedinger and Tasker, 1985, 1986).

Bayesian GLS regression

Bayesian inference is an alternative to the classical statistical approach. In a Bayesian framework, the parameters of the model are considered to be random variables,

whose probability density function should be estimated. Reis *et al*. (2005) developed a Bayesian approach to estimate the regional model parameters and showed that the Bayesian approach can provide a realistic description of the possible values of the model error variance, especially in the case where sampling error tends to dominate over the model errors in the regional analysis.

With the Bayesian approach it is assumed here that there is no prior information on any of the β parameters; thus a multivariate normal distribution with mean zero and a large variance (e.g. greater than 100) is used as a prior for the regression coefficient parameters as suggested by Reis *et al*. (2005). This prior is considered to be almost non informative, which produces a probability distribution function that is generally flat in the region of interest. The prior information for the model error variance σ_δ^2 is represented by an informative one-parameter exponential distribution, which represents the reciprocal of the prior mean of the model error variance. Reis *et al*. (2005) discusse the derivation of the choice of a prior for the model error variance for regionalizing the skew. For the regionalization of skew, we employed a value for the prior mean of the model error variance equal to six following Reis *et al*. (2005).

To derive the prior distribution for the standard deviation, mean flood and quantiles of the LP3 distribution we used an informative one-parameter exponential distribution where the reciprocal of the residual error variance estimate taken from OLS regression is used as the prior mean of the model error variance. For the mean flood and flood quantiles, the model error variance tends to dominate the regional analysis. In this case a zero or negative value for the model error variance is highly unlikely.

SELECTION OF PREDICTOR VARIABLES

In the OLS regression, several statistics are used to justify the model selection such as the traditional coefficient of determination (R^2), F statistics, Durbin Watson Statistics, Akaike information criterion (AIC) and Bayesian information criteria (BIC); Gelman *et al*., (2004). Among these statistics, the AIC and BIC penalize for the extra complexity in the model, which means that an extra predictor variable must improve the model significantly to justify its inclusion. A brief discussion below presents the Bayesian GLS regression statistics that guided our model selection procedure.

Average variance of prediction

In RFFA, the objective is to make prediction at both gauged and ungauged sites; hence a statistic appropriate for evaluation of model selection is the variance of prediction, which in many cases depends on the explanatory variables at both a gauged and ungauged sites. Hence, Tasker and Stedinger (1989) suggested the use of the average variance prediction (AVP).

By using a GLS regression model one can predict a hydrological statistic on average over a new region. Thus, this becomes the average variance of prediction $\mathrm{AVP}_{\mathrm{new}}$ for a new site which is made up of the average sampling error and the average model error (Tasker and Stedinger, 1986). For Bayesian GLS analysis according to Gruber *et al*. (2007):

$$\mathrm{AVP}_{\mathrm{new}} = E[\sigma_\delta^2] + \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{z}[\beta|\hat{\mathbf{y}}]\mathbf{x}_i^T \tag{7}$$

Also, if the prediction is for a site that was used in the estimation of the regional regression model, the measure of prediction $\mathrm{AVP}_{\mathrm{old}}$ requires an additional term:

$$\mathrm{AVP}_{\mathrm{old}} = E[\sigma_\delta^2] + \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathrm{Var}[\beta|\hat{\mathbf{y}}]\mathbf{x}_i^T - 2\sigma_\delta^2 \mathbf{x}_i (\mathbf{X}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\mathbf{e}_i \tag{8}$$

where $\mathbf{e}_i$ is a unit column vector with 1 at the ith row and 0 otherwise.

Bayesian and Akaike information criteria

The Akaike information criterion (AIC) is given by Equation (9), where $\Pi(Y)$ is the log-likelihood maximized function with respect to the number of predictor variables, n is the number of sites in the region (sample size in the regression) and k is the number of predictor variables in the fitted regression model. Here, $\Pi(Y)$is the log-likelihood of Equation (1). The first term on the right-hand side of Equation (9) measures essentially the true lack of fit while the second term measures the estimation uncertainty which is due to the number of predictor variables.

$$\mathrm{AIC} = -2\Pi(Y) + 2k \tag{9}$$

In practice, after the computation of the AIC for all of the competing models, one selects the model with the minimum AIC value, $\mathrm{AIC}_{\mathrm{min}}$. The Bayesian information criterion (BIC) is very similar to AIC, but is developed in a Bayesian framework:

$$\mathrm{BIC} = -2\Pi(Y) + \ln(n)k \tag{10}$$

The BIC penalizes more heavily for small sample sizes and models with higher values of k. Since $\Pi(Y)$ depends on the sample size, the competing models can be compared using AIC and BIC only if fitted using the same sample, but having different parameter estimators, as done in this study.

REGRESSION DIAGNOSTICS

The assessment of the regional regression model is made by using a number of statistical diagnostics such as a pseudo-coefficient of determination and standard error of prediction. An analysis of variance for the Bayesian GLS models is undertaken to examine the sampling and model errors. The Cook's distance and the standardized residuals are used to identify outlier sites; absence of outlier in regression diagnostics indicates the overall

adequacy of the regional model. These statistics are described below.

Co-efficient of determination R^2 and analysis of variance

The traditional coefficient of determination (R^2) measures the degree to which a model explains the variability in the dependent variable. It uses the partitioning of the sum of squared deviations and associated degrees of freedom to describe the variance of the signal versus the model error. Traditionally, for OLS regression, the total-sum-of-squared deviations about the mean (SST) is divided into two separate terms, the sum-of-squared errors explained by the regression model (SSR) and the residual sum-of-squared errors (SSE), where SST = SSR + SSE.

Reis *et al*. (2005) proposed a pseudo co-efficient of determination ($\overline{R}^2_{\text{GLS}}$) appropriate for use with the GLS regression. For traditional R^2, both the SSE and SST include sampling and model error variances, and therefore this statistic can grossly misrepresent the true power of the GLS model to explain the actual variation in the y_i. Hence, for the GLS regression a more appropriate pseudo co-efficient of determination is defined by:

$$\overline{R}^2_{\text{GLS}} = \frac{n[\hat{\sigma}^2_\delta(0) - \hat{\sigma}^2_\delta(k)]}{n\hat{\sigma}^2_\delta(0)} = 1 - \frac{\hat{\sigma}^2_\delta(k)}{\hat{\sigma}^2_\delta(0)} \tag{11}$$

where $\hat{\sigma}^2_\delta(k)$ and $\hat{\sigma}^2_\delta(0)$ are the model error variances when k and no explanatory variables are used, respectively. Here, $\overline{R}^2_{\text{GLS}}$ measures the improvement of a GLS regression model with k explanatory variables against the estimated error variance for a model without any explanatory variable. If $\hat{\sigma}^2_\delta(k) = 0$, $\overline{R}^2_{\text{GLS}} = 1$ as it should, even though the model is not perfect because var$[\eta_i + \delta_i]$ is still not zero because var$[\eta_i] > 0$. A pseudo analysis of variance (ANOVA) table is used in GLS regression for error variance analysis as presented by Reis *et al*. (2005) and Griffis and Stedinger (2007).

Standard error of prediction

If the residuals have a nearly normal distribution, the standard error of prediction in percent (SEP) for the true flood quantile estimator is described by (Tasker *et al*., 1986):

$$\text{SEP}(\%) = 100 \times [\exp(\text{AVP}_{\text{new}}) - 1]^{0.5} \tag{12}$$

Cook's distance and Z-score

Tasker and Stedinger (1989) developed measures such as Cook's distance (D) from an OLS to GLS case. Tasker and Stedinger (1989) and Reis *et al*. (2005) suggested that influence is large when D is greater than $4/n$, where n is the number of sites in the region.

Analysis of residuals provides a means of assessing the model fit and identifying possible outliers. In this study, the standardized residual (r_{si}) is used, which is the residual r_i divided by the square root of its variance. This is calculated as:

$$r_{si} = \frac{r_i}{[\lambda_i - \mathbf{x}_i(\mathbf{X}^{\text{T}}\boldsymbol{\Lambda}^{-1}\mathbf{X})^{-1}\mathbf{x}_i^T]^{0.5}} \quad \text{where } \lambda_i \text{ is the diagonal of } \boldsymbol{\Lambda} \tag{13}$$

To assess the adequacy of the model in estimating flood quantiles, a Z score is used, which is defined as:

$$Z_{\text{score}} = \frac{\text{LNQ}_{ARI,i} - \text{LN}\hat{\text{Q}}_{\text{ARI},i}}{\sqrt{\sigma^2_{\text{ARI},i} + \hat{\sigma}^2_{\text{ARI},i}}} \tag{14}$$

Here the numerator is the difference between the at-site flood quantile and regional flood quantile (estimated from the developed regression equation) and the denominator is the square root of the sum of the variances of the at-site ($\sigma^2_{\text{ARI},i}$) and regional ($\hat{\sigma}^2_{\text{ARI},i}$) flood quantiles in natural logarithm space. It is reasonable to assume that the errors in the two estimators are independent because $Q_{\text{ARI},i}$ is an unbiased estimator of the true quantile estimators based upon the at-site data, whereas the error in $\hat{Q}_{\text{ARI},i}$ is mostly due to the failure of the best regional model to estimate accurately the true at-site flood quantile. The use of log space makes the difference approximately normally distributed and hence enables the use of standard statistical tests.

EVALUATION STATISTICS

We evaluated the overall performance of the Bayesian GLS regression method by using one-at-a-time cross validation. The site of interest was left out in building the model so it was in effect being treated as an ungauged site. This was repeated for all the sites considered in the study. The advantage of the one-at-a-time cross-validation procedure is that it generates quantile or moment estimates for the site of interest which are not computed using the record available for that site. To compare model adequacy we adopted a number of evaluation statistics (Equations (15) to (19)) being the mean percent relative error (MPRE), Nash-Sutcliffe coefficient of efficiency (CE), root mean square error (RMSE) in log space and the RMSE (%) and the mean ratio of the predicted flow to observed flow. These evaluation statistics were applied to assess the performances of the Bayesian GLS-QRT and the GLS-PRT methods.

$$\text{MPRE} = \frac{100}{n}\sum_{i=1}^{n} \text{abs}\left(\frac{Q_{\text{pred}} - Q_{\text{obs}}}{Q_{\text{obs}}}\right) \tag{15}$$

$$\text{CE} = 1 - \frac{\sum_{i=1}^{n}(Q_{\text{pred}} - Q_{\text{obs}})^2}{\sum_{i=1}^{n}(Q_{\text{obs}} - \overline{Q})^2} \tag{16}$$

$$\text{RMSE} = \sqrt{\frac{\sum(\log Q_{\text{pred}} - \log Q_{\text{obs}})^2}{n}} \tag{17}$$

$$\text{RMSE}(\%) = 100(e^{5.302\sigma_e^2} - 1)^{1/2} \quad (18)$$

where σ_e is the RMSE in log base 10 units (Tasker *et al.*, 1996).

$$\text{Mean Ratio} = \frac{1}{n}\sum_{i=1}^{n}\frac{Q_{\text{pred}}}{Q_{\text{obs}}} \quad (19)$$

where Q_{obs} is the observed flood quantile obtained from at-site flood frequency analysis, Q_{pred} is the predicted flood quantile obtained from the Bayesian GLS-QRT or Bayesian GLS-PRT based on the one-at-a-time cross validation approach, $\overline{Q}$ is the mean of the Q_{obs} values for a given ARI and n is the number of sites.

The RMSE (%) and MPRE provide an indication of the overall accuracy of a model. The CE provides an indication of how good a model is at predicting values in relation to the mean value. The CE ranges from $-\infty$ in the worst case to $+1$ for a perfect model. The average value of the $Q_{\text{pred}}/Q_{\text{obs}}$ ratio gives an indication of the degree of bias (i.e. systematic over- or under estimation), where a value of 1 indicates a good average agreement between the Q_{pred} and Q_{obs}. A $Q_{\text{pred}}/Q_{\text{obs}}$ ratio value in the range of 0·5 to 2 may be regarded as a 'desirable estimate', a value smaller than 0·5 may be regarded as 'gross underestimation' and a value greater than 2·0 may be regarded as 'gross overestimation'. It should be mentioned here that these are only arbitrary limits and would provide a reasonable guide about the relative accuracy of the methods as far as the practical application of the methods is concerned. In applying these evaluation statistics to compare the alternative models, factors such as data error (e.g. measurement error and error due to rating curve extrapolation) and the error associated with the at-site flood frequency analysis were not considered.

AT-SITE FLOOD FREQUENCY ANALYSIS AND QUANTILE AND PARAMETER REGRESSION TECHNIQUES

For at-site flood frequency analysis, the LP3 distribution was adopted based on the findings of previous studies (Haddad and Rahman, 2010; I. E. Aust., 1987). At-site flood quantiles for ARIs of 2, 5, 10, 20, 50 and 100 years were estimated using FLIKE (at-site flood frequency analysis software) with the LP3 distribution and Bayesian parameter estimation procedures as described in Kuczera (1999). No prior information was used in fitting the LP3 distribution. The parameters of the LP3 distribution were also extracted from FLIKE software.

To regionalize the flood quantiles the sampling covariance matrix (Σ) of the LP3 distribution is required. Tasker and Stedinger (1989) and Griffis and Stedinger (2007) provide the approximate estimator of the components of Σ matrix of the LP3 distribution. The skew and standard deviation in the Σ matrix are subject to estimation uncertainty. In this study to avoid correlation between the residuals and the fitted quantiles, the (i) inter site correlation between the concurrent annual maximum flood series (ρ_{i_1,i_2}) was estimated as a function of the distance between sites i_1 and i_2 (ii) the standard deviations σ_{i_1} and σ_{i_2} were estimated using a separate OLS/GLS regression using the explanatory variables used in the study (given in Section 4) and (iii) the regional skew was used in place of the population skew γ as suggested by Tasker and Stedinger (1989). This analysis above used the regional estimates of the standard deviation and skew obtained from Bayesian GLS regression. The detailed information on the covariance matrices associated with the standard deviation and skew can be found in Reis *et al.* (2005) and Griffis and Stedinger (2007).

For the PRT, we adopted the GLS regression (Tasker and Stedinger, 1989 and Griffis and Stedinger, 2007) using a Bayesian framework (Reis *et al.*, 2005) to develop regression equations for the parameters of the LP3 distribution (i.e. mean μ, standard deviation σ, and skew coefficient γ of the logarithms of the annual maximum flood series). The regional values of standard deviation and skew were taken from the Σ matrix of the flood quantile modelling as mentioned above. The covariance matrix for the mean flood was obtained following Stedinger and Tasker (1986)

DATA DESCRIPTION

A total of 53 catchments were selected from Tasmania for this study. Tasmania is the 26th largest island in the world with an area of 68 401 km^2. Tasmania's climate is different from the rest of Australia due to its latitude and exposure to the Southern Ocean. Although Tasmania is relatively low-lying (the highest point is about 1600 m), it is Australia's most mountainous state, with no truly flat terrain. The major mountain ranges lie along the Western half of the state, starting at the coast in the south-west and extending inland. Although the eastern half of the state is generally lower and flatter, there are several significant mountain ranges in the east. Almost all the major rivers in Tasmania begin in the central highlands and flow to the coast. Given the western orientation of both the weather and the topography, most of the precipitation falls across to its west coast. The eastern region of the state is much drier. In the east of the state, the largest rainfall events occur in the warmer spring and summer months when low pressure systems in the Tasman Sea can direct an easterly onshore flow over Tasmania. The heaviest rainfalls in the west of the state are due to the passage of fronts, sometimes associated with an intense extratropical cyclone with a westerly or southwesterly airstream.

The elevations of the 53 selected catchments range from 120 to 1300 m (average: 478 m, 21 catchments are over 500 m and 16 catchments are below 300 m). The mean annual rainfall ranges from 520 to 3014 mm (median: 1046 mm). The catchment areas are in the range of 1·3–1900 km^2 (median: 158 km^2 and mean: 323 km^2). The selected catchments are mainly unregulated and have not been affected by major land use

changes. The streamflow data of these stations were prepared following stringent procedures as described in Haddad *et al*. (2010). The lengths of the annual maximum flood series in the catchments vary between 19 and 74 years (median: 28 years, mean: 30 years). Some 35 catchments (66% overall) have record lengths in the range of 19 to 30 years, 12 (23% overall) in the range of 31 to 40 years and 6 (11% overall) in the range of 41 years and greater.

Based on the findings from previous studies (e.g. Rahman, 2005), a total of six explanatory variables were used, as outlined below: (i) catchment area expressed in km^2 (*area*); (ii) design rainfall intensities (mm/h) for the 2 years average recurrence interval (ARI) with 1 h duration ($^2I_1$) and 2 years ARI with 12 h durations ($^2I_{12}$) (range: 3·08 to 6·33 mm/h and median: 4·39 mm/h), 50 years ARI with 1 h duration ($^{50}I_1$) and 50 years ARI with 12 h duration ($^{50}I_{12}$), design rainfall intensity values $I_{\mathrm{ARI},tc}$ [where ARI = 2, 5, 10, 20, 50 and 100 years and t_c = time of concentration (h), estimated from $t_c = 0{\cdot}76(\mathrm{area})^{0.38}$]; (iii) mean annual rainfall expressed in mm/year (rain); (iv) mean annual evapo-transpiration expressed in mm/y (evap) (range: 757 to 1027 mm/year, median: 865 mm/year); (iv) stream density expressed in km/km^2 (*sden*; range: 0·15 to 1·94 km/km^2, median: 1·35 km/km^2); (v) main stream slope expressed in m/km (*S1085*; range: 0·10 to 87·15 m/km, median: 12·5 m/km); and (vi) forest cover expressed as a percentage (%) of catchment area (*forest*; range: 4 to 99%, median: 68%). It was found that a simple natural logarithmic transformation of the dependent and independent variables (explanatory variables) was sufficient for achieving near-linearity in regression analysis and thus was adopted here. A log-transformed explanatory variable was centred by subtracting its log-mean value so that the intercept term in the regression equation represents the mean of the logarithm of the observed dependent variable.

RESULTS

Selection of predictor variables

A total of 17 (for the mean, standard deviation and skew models) and 25 (for the flood quantiles) different combinations of explanatory variables were considered in selecting the best set of predictors. These combinations are listed in Table A1 in the Appendix. For each of these combinations, the model error variance σ_δ^2, standard error of the model error variance, $\overline{R}^2_{\mathrm{GLS}}$, BIC, AIC, $\mathrm{AVP}_{\mathrm{new}}$ (AVPN) and $\mathrm{AVP}_{\mathrm{old}}$ (AVPO) were examined to identify the best set of predictor variables in the GLS regression.

Figure 2A shows the model error variance, standard error of the model error variance and $\overline{R}^2_{\mathrm{GLS}}$ values for the mean flood model. Here combination 6 (*constant/area/*$^2I_{12}$) has the smallest model error variance ($\sigma_\delta^2 = 0{\cdot}35$ and standard error = 0·066) and the highest $\overline{R}^2_{\mathrm{GLS}}$. The posterior expected values for β_1 and β_2 were approximately 3 and 16 times of the posterior standard error values which support the inclusion of both of the predictor variables *area* and $^2I_{12}$ in the model. Figure 2B shows that combination 6 has the smallest BIC, AIC, AVPN and AVPO values. Based on these results, combination 6 was selected as the best set of predictor variables for the mean flood model. The same procedure was followed for selecting the best combination of predictor variables for the models of standard deviation, skew and flood quantiles. The best combination for the standard deviation model was combination 17 (*constant/rain*). For the skew model, combination 4 (*constant/area/*$^{50}I_1$) was found to be the best set of predictor variables. Figures A1 and A2 in the Appendix present the different statistics for the skew model.

The best set of predictor variables for each of the flood quantile models (ARIs of 2, 5, 10, 20, 50 and 100 years) was selected following the above procedure. The results for ARI of 20 years are presented below. Here, combination 6 (*constant/area/*$^{50}I_{12}$) showed the lowest model error variance ($\sigma_\delta^2 = 0{\cdot}30$ and standard error = 0·071). The posterior expected values for β_1 and β_2 were approximately 15 and 5 times of the posterior standard error values which support the inclusion of both of the predictor variables *area* and $^{50}I_{12}$ in the model. For all the other ARIs, combination 6 was found to be the best set of predictor variables.

Analysis of Variance

Tables I to IV present pseudo ANOVA for the Bayesian GLS regression models for the first three moments of the LP3 distribution and flood quantile of the ARI of 20 years. The ANOVA table describes how much of the variation in the observations can be attributed to the regional model, and how much of the residual variation can be attributed to the model error and sampling error, respectively. The pseudo ANOVA tables for the flood quantile models of ARIs of 2 and 100 years are presented in Tables A2 and A3 in the Appendix.

One can see that for the three moments of the LP3 distribution, the sampling error increases with the order

Figure 1. Location of study catchments in Tasmania, Australia

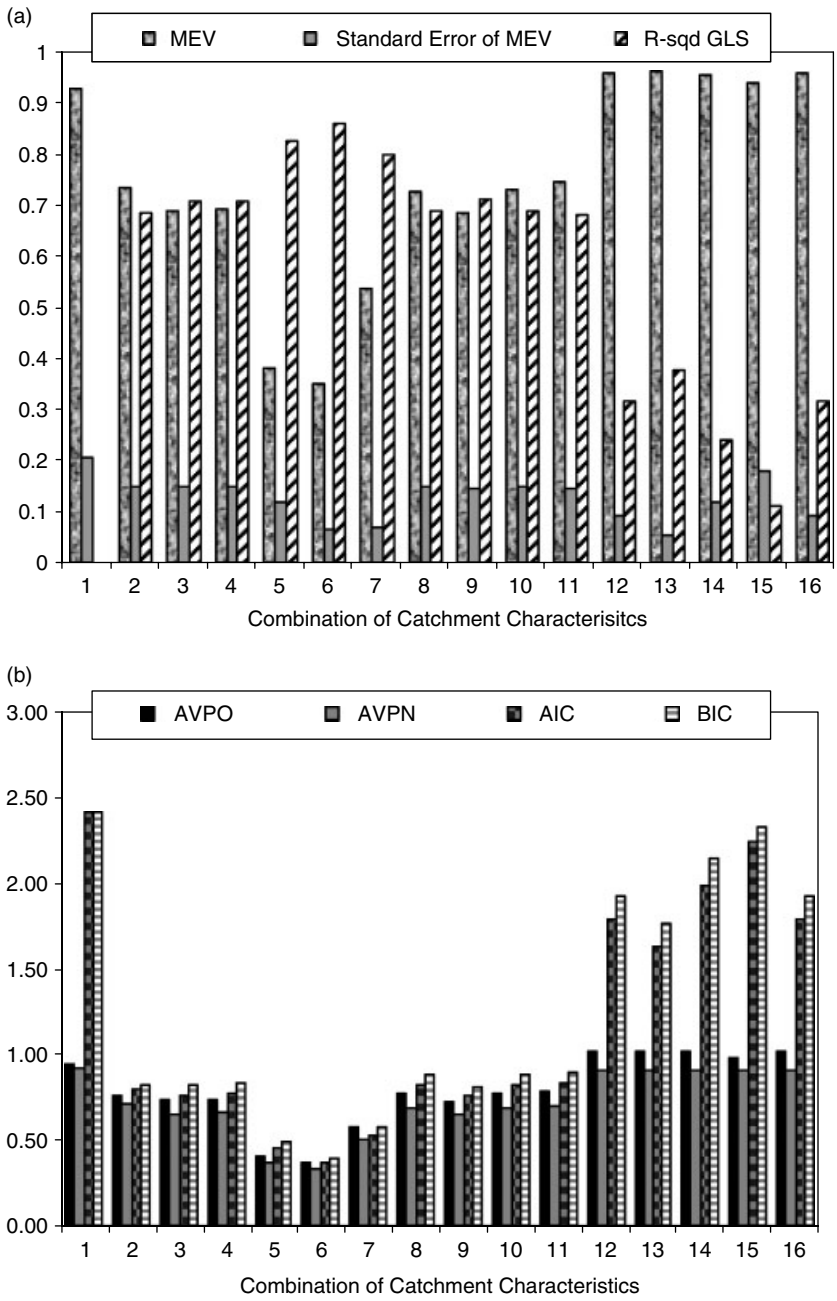


Figure 2. Selection of explanatory variables for the Bayesian GLS regression model for the mean flood. MEV = model error variance

of the moment (i.e. an increase in the error variance ratio, EVR, which is the ratio of the sampling error to the model error). For the mean flood and standard deviation models, the model error dominates the regional analysis. This is more pronounced for the mean flood model, where the sampling error variance is only 6% of the model error variance, while the sampling error variance for the standard deviation model is 54% of the model error variance. Based on the model error variance, the mean flood has the greatest heterogeneity associated with it as compared to the standard deviation and skew models. For the skew model, EVR = 9, which means that sampling error is nine times higher than the model error. This clearly shows that a Bayesian GLS is the right modelling choice for the skew rather than the OLS one. Another important observation is that if a method of moment estimator was used to estimate the model error variance σ_δ^2 for the skew model, the model error variance would have been underestimated as the sampling error has heavily dominated the regional analysis. In this case, the Bayesian procedure used in this study has provided a reasonably accurate estimate of the model error variance because it represents the values of σ_δ^2 by computing expectations over the entire posterior distribution.

The pseudo ANOVA tables of the flood quantiles show that sampling error variance increases with increasing ARIs as expected. The flood quantile of ARI of 2 years has an EVR of only 2%, i.e. it has the smallest sampling error variance and the highest model error variance. This means that based on the model error

Table I. Pseudo ANOVA table for the mean flood model (Combination 6)

Source	Degrees of freedom	Sum of squares
Model	$k = 3$	$n(\sigma_{\delta 0}^2 - \sigma_\delta^2) = 30{\cdot}5$
Model error δ	$n - k - 1 = 48$	$n(\sigma_\delta^2) = 17{\cdot}8$
Sampling error η	$n = 52$	$tr[\Sigma(\hat{y})] = 1{\cdot}13$
Total	$2n - 1 = 103$	Sum of the above $= 49{\cdot}44$
	EVR	0·06

Table II. Pseudo ANOVA table for standard deviation model (Combination 17)

Source	Degrees of freedom	Sum of squares
Model	$k = 2$	3·58
Model error δ	$n - k - 1 = 49$	3·60
Sampling error η	$n = 52$	1·94
Total	$2n - 1 = 103$	9·12
	EVR	0·54

Table III. Pseudo ANOVA table for skew model (Combination 4)

Source	Degrees of freedom	Sum of squares
Model	$k = 3$	0·62
Model error δ	$n - k - 1 = 48$	1·74
Sampling error η	$n = 52$	15·5
Total	$2n - 1 = 103$	17·84
	EVR	9·0

Table IV. Pseudo ANOVA table for flood quantile model (ARI = 20 years; Combination 6)

Source	Degrees of freedom	Sum of squares
Model	$k = 3$	34·30
Model error δ	$n - k - 1 = 48$	15·49
Sampling error η	$n = 52$	2·08
Total	$2n - 1 = 103$	51·9
	EVR	0·13

variance the flood quantile model of ARI of 2 years experiences the greatest regional heterogeneity as compared to flood quantile models of higher ARIs. For the 100 years ARI, the EVR is 17% which suggests that the Bayesian GLS model is quite appropriate for modelling the larger ARI flood quantiles for the study data set.

Regression diagnostics

To assess the adequacy of the Bayesian GLS regression models Cook's distance values were calculated. No outlier/influential sites were found for the mean, standard deviation, and flood quantile models. For the skew model (Figure 3), sites 8 and 50 were above the threshold value of 0·076 (i.e. 4/53, where 53 is the total number of sites). Site 8 showed the largest standardized residual value. The flow data, site history, and flood frequency plots of these two sites were examined. It was found that site 8 had a record length of 33 years (top 20%) and a very small annual maximum flow in 1968 which is not surprising as this was a drought year. This small flow caused a large negative skew of −1·60 for the site. Site 50 had record length of 46 years (fifth largest record length) and a skew value 1·15, and it did show the largest influence value (Figure 3). The regression analysis was repeated by removing these two sites. Indeed site 8 did influence the analysis with a decrease in the expected model error variance σ_δ^2 from 0·052 to 0·034. The AVPO and AVPN dropped notably from 0·073 and 0·067 to 0·053 and 0·049, respectively. The $\overline{R}^2_{\mathrm{GLS}}$ also increased from 36 to 53%, which seems to be a remarkable increase. The effective record length based on AVPN of 0·049 in this case is 122 years (Combination 4) which is nearly 4 times the average record length for Tasmania. Site 8 did therefore influence the results notably and was therefore removed from the database in subsequent analyses. The removal of site 50 resulted in little improvement in the skew model with a negligible increase in $\overline{R}^2_{\mathrm{GLS}}$ (55%) and a slightly smaller σ_δ^2 (0·032) and was therefore retained.

To assess that the underlying assumption of normality of the residuals, the standardized regression residuals (obtained from Equation 13) *versus* predicted values (obtained from one-at-a-time cross validation) were plotted. Figure 4 shows the plots for the skew model and flood quantiles for the ARI of 20 years. If the underlying assumption is satisfied, the standardized residuals should not be of greater magnitude than ±2. A reasonable assumption is that 95% of the standardized residuals should fall between ±2, which is the case in Figure 4. Also no specific pattern can be identified in the plot with the standardized residuals being nearly equally distributed below and above zero. Similar results were obtained for the mean, standard deviation, and other flood quantile models. Statistical hypothesis tests using both the Kolmogorov–Smirnov and Anderson–Darling tests for normality were applied at the 10% level of significance and it was found that the residuals were approximately normally distributed for all the ARIs (2 to 100 years for both QRT and PRT).

The QQ-plots of the standardized residuals (Figure 5 for skew model and flood quantile for ARI of 20 years) show that the assumption of normality for the standardized residuals is well satisfied with all the points closely following a straight line. If the standardized residuals were indeed normally and independently distributed with mean 0 and variance 1, the slope of the best fit line in the QQ-plot, which can be interpreted as the standard deviation of the normal score (or Z-score) of the quantile, should approach 1 and the intercept, which is the mean of the normal score of the quantile should approach 0 as the number of sites increases. Figure 5 indeed shows that the fitted lines passes through the origin (0, 0) and has a slope approximately equal to 1. These results indicate that the developed prediction equations satisfy the homogeneity and normality of the residual assumptions

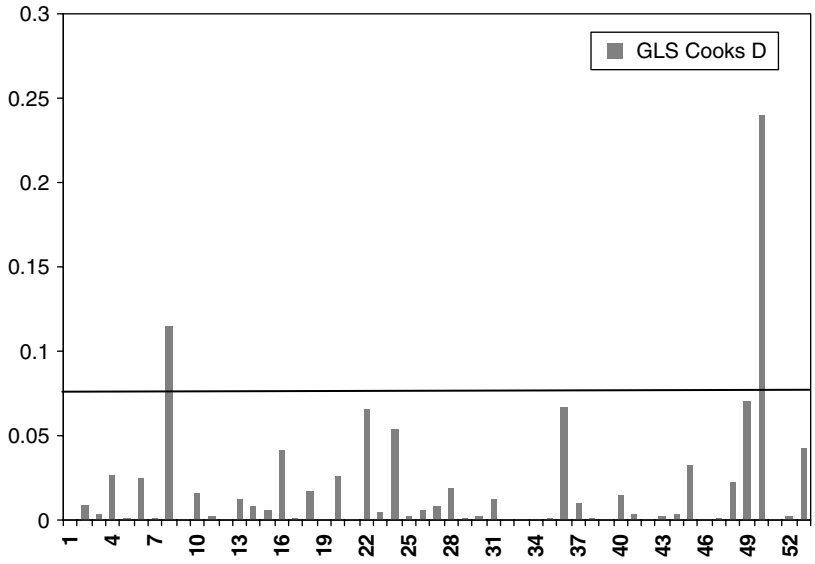


Figure 3. GLS Cook's distance (D) for locating outlier sites for skew model based on combination 4

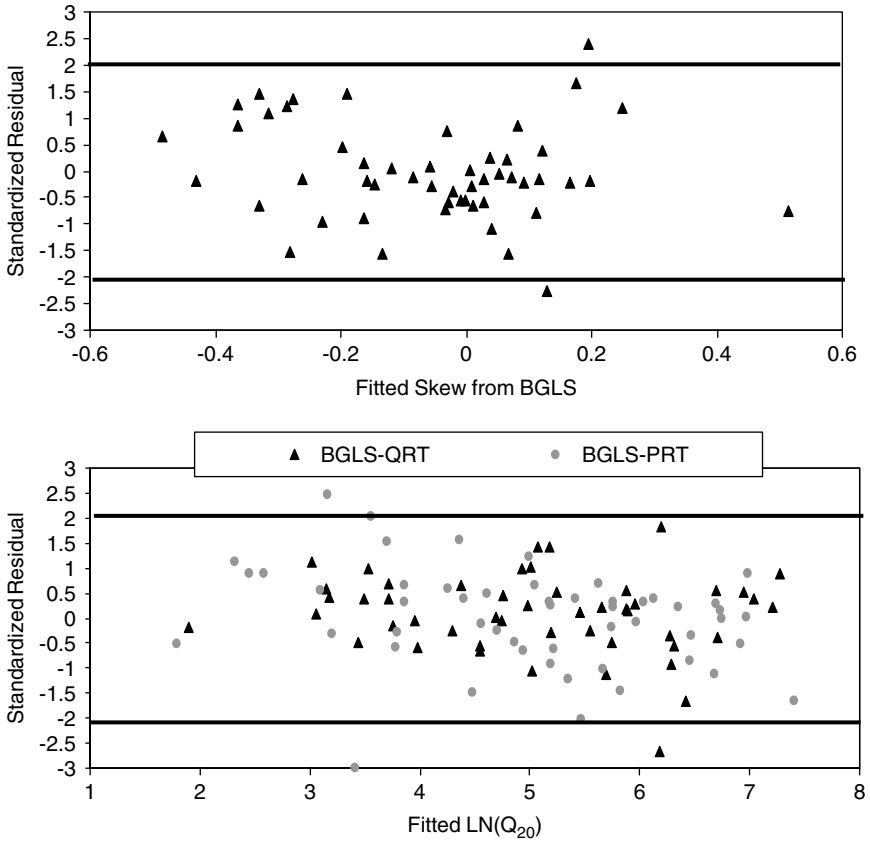


Figure 4. Standardized residuals vs. predicted values for the skew and flood quantiles of ARI = 20 years

quite well. Similar results were also found for the mean, standard deviation and other flood quantile models.

The coefficients of the regression models, $\overline{R}^2_{GLS}$ and SEP values are presented in Table V. It should be noted that to implement the QRT only two predictor variables are required whereas, PRT needs four predictor variables. The significance of the regression coefficient values provided in Table V was evaluated using the Bayesian plausibility value (BPV) as developed by Reis *et al.* (2005) and Gruber *et al.* (2007). The BPV allows one to perform the equivalent of a classical hypothesis *P*-value test within a Bayesian framework. The advantage of the BPV is that it uses the posterior distribution of each parameter, which also reflects the prior. The BPV values for the regression coefficients associated with the QRT in Table V over all the ARIs were between 2 and 8% for the variable area and 0·000% for design rainfall intensity $^{50}I_{12}$. This justifies the inclusion of predictor variables

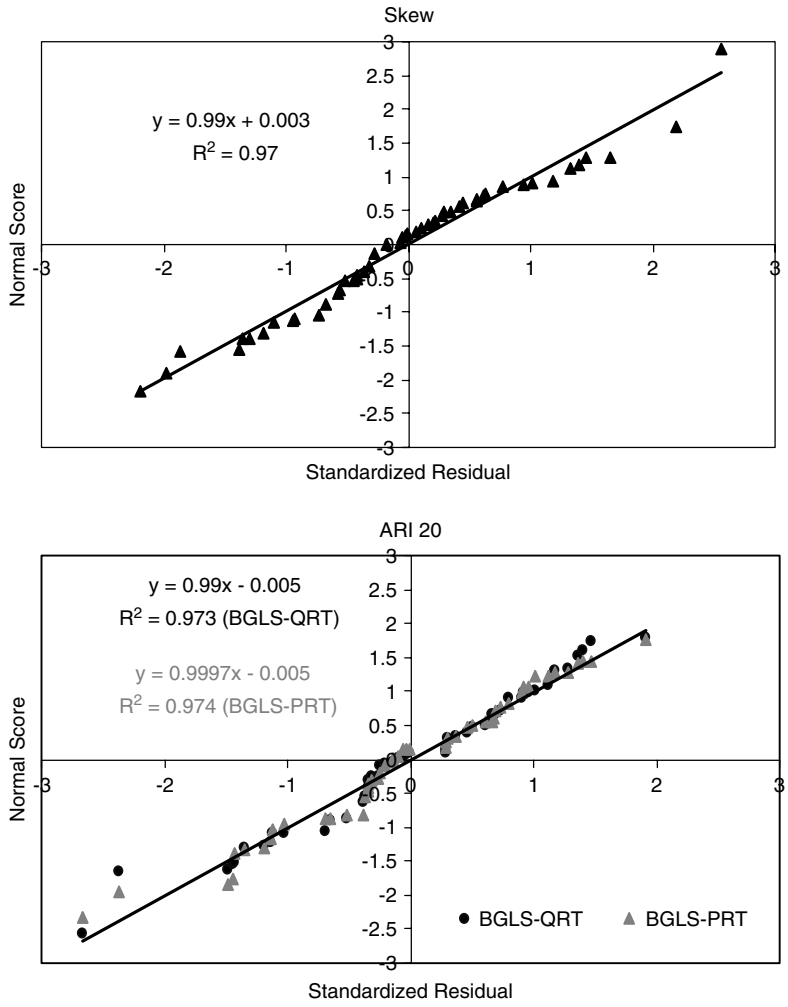


Figure 5. QQ-plot of the standardized residual (eq. (13)) vs. Z score of the fitted quantiles (eq. (14)) (skew model above and flood quantile of ARI = 20 years model below)

Table V. Regional Bayesian GLS regression equations for the mean (μ), standard deviation (σ), skew (γ) and flood quantiles

Model/ARI (years)	Coefficients			$\overline{R}^2_{GLS}$ (%)	SEP (%)
	β_0	β_1	β_2		
Parameter regression technique (PRT)					
Mean, μ	4·0	0·90	3·85	86	67
Standard deviation, σ	0·64	0·55	—	53	28
Skew, γ	−0·05	0·07	1·20	52	22
Quantile regression technique (QRT)					
2	4·18	0·91	3·35	76	83
5	4·59	0·89	2·80	82	61
10	4·87	0·85	2·57	84	58
20	5·09	0·84	2·39	83	58
50	5·45	0·84	2·23	82	62
100	5·48	0·82	2·02	79	66

The SEP% and $\overline{R}^2_{GLS}$ values are obtained from one-at-a-time validation.

$$\mu = \beta_0 + \beta_1 \ln(\text{area}) + \beta_2 \ln({}^{2}I_{12})$$

$$\sigma = \beta_0 + \beta_1 \ln(\text{rain})$$

$$\gamma = \beta_0 + \beta_1 \ln(\text{area}) + \beta_2 \ln({}^{50}I_{1})$$

$$\ln(Q_{ARI}) = \beta_0 + \beta_1 \ln(\text{area}) + \beta_2 \ln({}^{50}I_{12})$$

area and ${}^{50}I_{12}$ in the prediction equations for QRT. The BPVs for the skew model were 23 and 11% for area and ${}^{50}I_1$, respectively, indicating that these are not very good predictors for skew in this particular case. Indeed Figure A1 reveals that the model error variances among the different combinations are practically the same, this, and the results from the BPVs suggest a regional constant skew model may be adequate for Tasmania similar to other studies (e.g. Reis *et al*., 2005; Gruber and Stedinger, 2008). The BPVs for the mean model were close to 1% for both the predictor variables. For the standard deviation model, the BPV for the predictor variable rain was 1%.

The average SEP was calculated using Equation (12) for the flood quantiles and the first three moments of the LP3 distribution from one-at-a-time cross validation. Equation (12) is only valid if the residuals closely follow a normal distribution (this assumption has been largely satisfied as discussed above). The SEP values for the first three moments of the LP3 distribution are 67, 28, and 22% respectively. The mean flood has the highest SEP value (i.e. shows greater heterogeneity than the standard deviation and skew model). The SEP for the flood quantile model of ARI of 2 years is 83%, which

Table VI. Performance statistic results for one-at-a-time validation for Bayesian GLS-QRT and Bayesian GLS-PRT for MPRE (%), CE and RMSE (%) and mean ratio over the 52 sites

ARI (years)	MPRE (%)		CE		RMSE (%)		Mean ratio	
Model	PRT	QRT	PRT	QRT	PRT	QRT	PRT	QRT
2	33	38	0·93	0·60	70	106	1·22	1·52
5	35	34	0·85	0·71	69	74	1·14	1·19
10	34	30	0·77	0·78	69	69	1·13	1·17
20	36	27	0·68	0·81	73	71	1·14	1·16
50	39	27	0·53	0·75	81	77	1·15	1·15
100	49	33	0·41	0·57	86	73	1·17	1·11

is notably higher than that of the mean flood (Table V). For the rest of the ARIs, the SEP values are lower than that of the mean flood model except for the 100 years ARI for which the SEP values are very similar for these two cases. The lowest SEP values were found for ARI of 10 and 20 years (58%). The ARIs of 10 and 20 years also showed the highest $\overline{R}^2_{\mathrm{GLS}}$ values (84 and 83%, respectively) among all the ARIs. It may be noted here that the SEP values obtained here are relatively high, which is due to the model error dominating the total error in the regional analysis as shown in the ANOVA results (e.g. Tables I, II, and IV).

Evaluation statistics

Table VI summarizes the evaluation statistics for the Bayesian GLS-QRT and Bayesian GLS-PRT models. These values are based on the independent testing of the prediction equations from one-at-a-time cross validation approach. The MPRE value for the PRT model of ARI of 2 years is 5% smaller than that of the QRT. However, for the ARIs of 5 to 100 years the QRT shows relatively smaller (by 1–16%) MPRE values. In relation to CE, the PRT provides much higher values (i.e. better model performance) than the QRT for 2 and 5 years ARIs. For 10 to 100 years ARIs, the QRT shows higher CE values.

The RMSE (%) values for both the QRT and PRT are very similar for the 10 and 20 years ARIs. The RMSE (%) values for 2 and 5 years ARIs for the PRT model is 36 and 5% smaller than that of the QRT model. For 50 and 100 years ARIs, the RMSE (%) values for the QRT models are 4 and 13% smaller than the PRT models.

Finally the ratio $Q_{\mathrm{pred}}/Q_{\mathrm{obs}}$ values were assessed. Here Q_{pred} values were obtained from one-at-a-time cross validation. Based on the average values over the 52 sites, QRT shows much overestimation for 2 years ARI. For the other ARIs, the ratio values are quite similar. Based on the criteria mentioned in Section 2·5; out of the 312 cases (6 ARIs and 52 sites), QRT and PRT produce 254 and 232 cases, respectively, with a 'desirable estimation', which is equivalent to 81 and 74% of the cases, respectively. The PRT, however shows more 'acceptable estimation' for ARI of 2 and 5 years.

Overall, the above results demonstrate that the QRT performs slightly better than the PRT for ARIs of 10 to 100 years and PRT performs relatively better for ARIs of 2 and 5 years. It should be mentioned here that the adopted validation procedure favours the QRT over the PRT because in the QRT the quantiles are used directly to develop the prediction equations as well as in the validation purpose, however in the case of PRT the prediction equations are developed for the moments of the distribution and validation is made with the estimated quantiles from the fitted distribution. Nevertheless, the PRT has performed very similarly to the QRT, which shows that the PRT is a viable approach for RFFA as an alternative to QRT.

CONCLUSIONS

This article compares two regression-based RFFA methods that employ a Bayesian Generalized Least Squares framework: quantile regression technique (QRT) and parameter regression technique (PRT). It has been shown that using a method similar to stepwise regression, the best set of predictor variables in the Bayesian GLS regression can be identified by employing a number of statistics such as the model error variance, average variance of prediction, Bayesian information criterion and the Akaike information criterion. The method is applied to 53 catchments in Tasmania, Australia. It has been found that catchment area and design rainfall intensity are the most important explanatory variables for the flood quantile estimation using the QRT. For the PRT, four explanatory variables were employed for predicting the first three moments of the LP3 distribution. The developed regression models have satisfied the underlying model assumptions quite well. No outlier sites have been detected in the regression diagnostic plots associated with the adopted models. The flood quantiles obtained from the developed prediction equations (based on one-at-a-time cross validation approach) have been compared with the at-site flood frequency estimates using a number of evaluation statistics. These results have shown that for Tasmania the QRT performs slightly better than PRT for higher ARIs (10–100 years), while the PRT performs better for smaller ARIs (2 and 5 years). Overall, there is only a modest difference between the performances of the QRT and PRT, and hence it may be argued that the PRT is a viable alternative to the QRT in RFFA. The RFFA techniques presented here can easily be adapted to other Australian states and countries to derive more accurate regional flood predictions.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial contribution of the Federal Department of Climate Change Australia and Engineers Australia to undertake the study, Hydro Tasmania for providing the streamflow data and the Bureau of Meteorology for providing the climatic data CDs. The authors would also like to strongly acknowledge Professor George Kuczera and Dr Dirceu Reis for their helpful comments and suggestions on this work.

We also acknowledge Dr Tom Micevski, Dr Fiona Ling, Mr Erwin Weinmann, Associate Professor James Ball, Mr Mark Babister and Dr William Weeks for their suggestions and input to the project. This study is a part of revision of Australian Rainfall and Runoff (Project 5) which is aimed at the development of more precise regional flood estimation methods for Australia, which is being undertaken in association with Engineers Australia's National Committee on Water Engineering. Thanks to three anonymous reviewers, whose comments have helped to improve the paper notably.

REFERENCES

Benson MA. 1962. Evolution of methods for evaluating the occurrence of floods. *US Geological Survey Water Supply Paper* **1580-A**, 30.

Eng K, Tasker GD, Milly PCD. 2005. An analysis of region-of-influence methods for flood regionalization in the Gulf-Atlantic rolling plains. *Journal of the American Water Resources Association* **41**(1): 135–143.

Gelman A, Carlin JB, Stern HS, Rubin, DB. 2004. *Bayesian Data Analysis*. Chapman and Hall/CRC: Boca Raton.

Griffis VW, Stedinger JR. 2007. The use of GLS regression in regional hydrologic analyses. *Journal of Hydrology* **204**(:): 82–95.

Gruber AM, Reis DS Jr, Stedinger JR. 2007. Models of regional skew based on Bayesian GLS Regression. In *World Environmental & Water Resources Conference*, 15–18 May 2007.Tampa, Florida.

Gruber AM, Stedinger JR. 2008. Models of LP3 Regional Skew, Data Selection, and Bayesian GLS Regression. In *EWRI World Water & Environmental Resources Congress*, 13–16 May 2008, American Society of Civil Engineers, Honolulu, HI.

Gupta VK, Mesa, OJ, Dawdy DR. 1994. Multiscaling theory of flood peaks: regional quantile analysis. *Water Resources Research* **30**(12): 3405–3421.

Gupta VK, Dawdy DR. 2006. Physical interpretations of regional variations in the scaling exponents of flood quantiles. *Hydrological Processes* **9**(3–4): 347–361.

Hackelbusch A., Micevski M, Kuczera G, Rahman A, Haddad, K. 2009. Regional flood frequency analysis for eastern NSW: A region of influence approach using generalised least squares log-Pearson 3 parameter regression. In 32nd *Hydrology and Water Resources Symposium*, 30th Nov to 3rd Dec 2009, Newcastle.

Haddad K, Rahman A, Weinmann PE, Kuczera G, Ball JE. 2010. Streamflow data preparation for regional flood frequency analysis: Lessons from south-east Australia. *Australian Journal of Water Resources* **14**(1): 17–32.

Haddad K, Rahman A. 2010. Selection of the best fit flood frequency distribution and parameter estimation procedure—A case study for Tasmania in Australia. *Stochastic Environmental Research and Risk Assessment* DOI: 10.1007/s00477-010-0412-1.

Hosking JRM, Wallis JR. 1997. *Regional frequency analysis—an approach based on L-moments*. Cambridge University Press: New York, 224.

Kroll CN, Stedinger JR. 1998. Regional hydrologic analysis: ordinary and generalized least squares revisited. *Water Resources Research* **34**(1): 121–128.

Kuczera G. 1999. Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference. *Water Resources Research* **35**(5): 1551–1557.

Madsen H, Rosbjerg D. 1997. Generalized least squares and empirical Bayes estimation in regional partial duration series index-flood modelling. *Water Resources Research* **33**(4): 771–782.

Madsen H, Rosbjerg D, Harremoes P. 1995. Application of the Bayesian approach in regional analysis of extreme rainfalls. *Stochastic Hydrology and Hydraulics* **9**(:): 77–88.

Madsen H, Pearson, CP, Rosbjerg D. 1997. Comparison of annual maximum series and partial duration series for modelling extreme hydrologic events, 2. Regional modelling. *Water Resources Research* **33**(4): 759–769.

Madsen H, Mikkelsen PS, Rosbjerg D, Harremoes P. 2002. Regional estimation of rainfall intensity duration curves using generalized least squares regression of partial duration series statistics. *Water Resources Research* **38**(11): 1–11.

Micevski T, Kuczera, G. 2009. Combining site and regional flood information using a Bayesian Monte Carlo approach. *Water Resources Research* **45**(:): W04405, DOI: 10.1029/2008WR007173.

Nezhad MK, Chokmani K, Ouarda TBMJ, Barbet M, Bruneau P. 2010. Regional flood frequency analysis using residual kriging in physiographical space. *Hydrolological Processes* **24**(:): 2045–2055, DOI: 10.1002/hyp.7631.

Pandey GR, Nguyen VTV. 1999. A comparative study of regression based methods in regional flood frequency analysis. *Journal of Hydrology* **225**(:): 92–101.

Rahman A. 2005. A quantile regression technique to estimate design floods for ungauged catchments in south-east Australia. *Australian Journal of Water Resources* **9**(1): 81–89.

Reis DS Jr, Stedinger JR, Martins ES. 2005. Bayesian GLS regression with application to LP3 regional skew estimation. *Water Resources Research* **41**(:): W10419, DOI: 10.1029/2004WR00344.

Stedinger JR, Tasker GD. 1985. Regional hydrologic analysis, 1.Ordinary, weighted, and generalised least squares compared. *Water Resources Research* **21**(9): 1421–1432.

Stedinger JR, Tasker GD. 1986a. Correction to Regional hydrologic analysis, 1.Ordinary, weighted, and generalised least squares compared. *Water Resources Research* **22**(5): 844.

Stedinger JR, Tasker GD. 1986b. Regional hydrologic analysis, 2. Model error estimators, estimation of sigma and log—Pearson type 3 distributions. *Water Resources Research* **22**(10): . 1487–1499.

Stedinger JR, Vogel RM, Foufoula-Georgiou E. 1993. Frequency analysis of extreme events. In *Handbook of Hydrology* (Chapter 18). McGraw Hill Book Co.: New York, 18·1–18·66.

Stedinger JR, Lu, LH. 1995. Appraisal of regional and index flood quantiles estimators. *Stochastic Hydrology and Hydraulics* **9**(1): 49–75.

Tasker GD, Stedinger JR. 1989. An operational GLS model for hydrologic regression. *Journal of Hydrology* **111**(:): 361–375.

Tasker GD, Eychaner, JH, Stedinger JR. 1986. Application of generalised least squares in regional hydrologic regression analysis. *US Geological Survey Water Supply Paper* 2310: 107–115.

Tasker GD, Hodge SA. Barks CS. 1996. Region of Influence regression for estimating the 50-year flood at ungauged sites. *Water Resources Bulletin* **32**(1): 163–170.

Thomas DM, Benson MA. 1970. Generalization of streamflow characteristics from drainage basin characteristics. *US Geological Survey Water Supply Paper* 1975, 55.

APPENDIX

Table AI. Combinations of explanatory variables in Bayesian GLS regression (Const means the intercept term in the regression equation, the explanatory variables are described in Section 4)

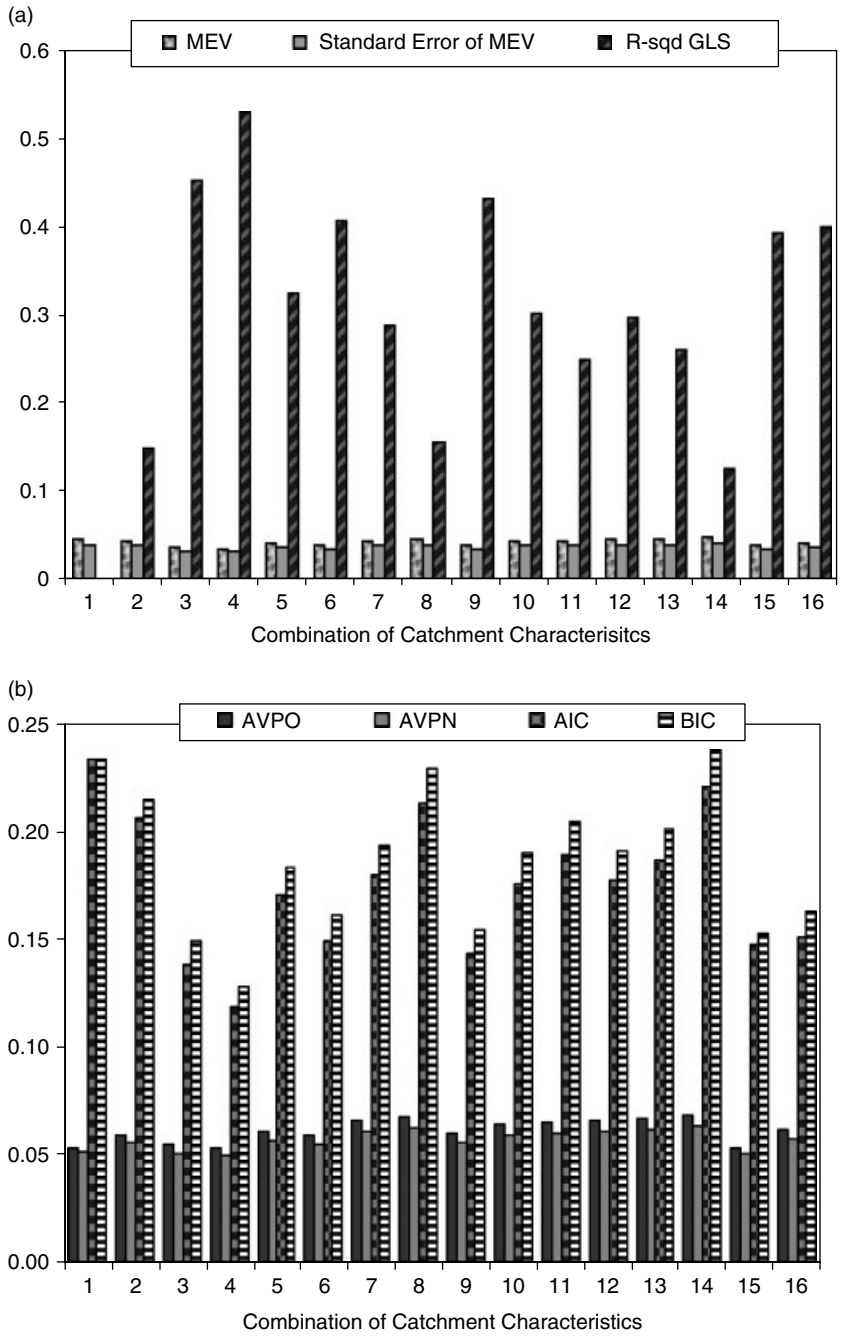
Combination	Combinations for mean, standard deviation and skew models	Combinations for flood quantile model
1	Const	Const
2	Const, area	Const, area
3	Const, area, $(^2I_1)$	Const, area, $(^2I_1)$
4	Const, area, $(^{50}I_1)$	Const, area, $(^2I_{12})$
5	Const, area, $(^{50}I_{12})$	Const, area, $(^{50}I_1)$
6	Const, area, $(^2I_{12})$	Const, area, $(^{50}I_{12})$
7	Const, area, rain	Const, area, rain
8	Const, area, for	Const, area, for
9	Const, area, evap	Const, area, for, evap
10	Const, area, S1085	Const, area, I_{ARI}
11	Const, area, sden	Const, area, evap
12	Const, sden, rain	Const, area, S1085
13	Const, for, rain	Const, area, sden
14	Const, S1085, for	Const, sden, rain
15	Const, evap	Const, for, rain
16	Const, rain, evap	Const, area, $(^{50}I_{12})$, rain
17	Const, rain	Const, area, $(^{50}I_{12})$, sden
18	—	Const, area, $(^{50}I_{12})$, rain, evap
19	—	Const, area, $(^{50}I_{12})$, I_{ARI}, evap
20	—	Const, area, $(^{50}I_{12})$, I_{ARI}, rain, evap
21	—	Const, area, $(^{50}I_{12})$, I_{ARI}, sden
22	—	Const, area, $(^{50}I_{12})$, I_{ARI}, S1085
23	—	Const, area, I_{ARI}, evap
24	—	Const, area, I_{ARI}, rain
25	—	Const, area, $(^2I_1)$, I_{ARI}

Table A2. Pseudo ANOVA table for flood quantile model (ARI = 2 years) (Combination 6)

Source	Degrees of Freedom	Sum of Squares
Model	$k = 3$	21·20
Model error δ	$n - k - 1 = 48$	28·2
Sampling error η	$n = 52$	0·91
Total	$2n - 1 = 103$	50·3
	EVR	0·03

Table A3. Pseudo ANOVA table for flood quantile model (ARI = 100 years) (Combination 6)

Source	Degrees of freedom	Sum of Squares
Model	$k = 3$	30·70
Model error δ	$n - \mathrm{k} - 1 = 48$	19·04
Sampling error η	$n = 52$	3·25
Total	$2n - 1 = 103$	52·99
	EVR	0·17



Figures A1 and A2. Various statistics for selecting the best set of predictor variables for the skew model