Journal of Hydrology 430–431 (2012) 142–161

Contents lists available at SciVerse ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

# Regional flood frequency analysis in eastern Australia: Bayesian GLS regression-based methods within fixed region and ROI framework – Quantile Regression vs. Parameter Regression Technique

Khaled Haddad [a], Ataur Rahman [b,*]

[a] School of Engineering, University of Western Sydney, Penrith, Sydney, Australia
[b] School of Engineering, University of Western Sydney, Building XB248, Locked Bag 1797, Penrith NSW 2751, Australia

ARTICLE INFO

*Article history:*
Received 10 June 2011
Received in revised form 9 November 2011
Accepted 5 February 2012
Available online 13 February 2012
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Attilio Castellarin, Associate Editor

*Keywords:*
Bayesian method
Generalised Least Squares Regression
Region-of-influence approach
Regional flood frequency analysis
Quantile Regression
Parameter Regression

SUMMARY

In this article, an approach using Bayesian Generalised Least Squares (BGLS) regression in a region-of-influence (ROI) framework is proposed for regional flood frequency analysis (RFFA) for ungauged catchments. Using the data from 399 catchments in eastern Australia, the BGLS-ROI is constructed to regionalise the flood quantiles (Quantile Regression Technique (QRT)) and the first three moments of the log-Pearson type 3 (LP3) distribution (Parameter Regression Technique (PRT)). This scheme firstly develops a fixed region model to select the best set of predictor variables for use in the subsequent regression analyses using an approach that minimises the model error variance while also satisfying a number of statistical selection criteria.

The identified optimal regression equation is then used in the ROI experiment where the ROI is chosen for a site in question as the region that minimises the predictive uncertainty. To evaluate the overall performances of the quantiles estimated by the QRT and PRT, a one-at-a-time cross-validation procedure is applied. Results of the proposed method indicate that both the QRT and PRT in a BGLS-ROI framework lead to more accurate and reliable estimates of flood quantiles and moments of the LP3 distribution when compared to a fixed region approach. Also the BGLS-ROI can deal reasonably well with the heterogeneity in Australian catchments as evidenced by the regression diagnostics. Based on the evaluation statistics it was found that both BGLS-QRT and PRT-ROI perform similarly well, which suggests that the PRT is a viable alternative to QRT in RFFA.

The RFFA methods developed in this paper is based on the database available in eastern Australia. It is expected that availability of a more comprehensive database (in terms of both quality and quantity) will further improve the predictive performance of both the fixed and ROI based RFFA methods presented in this study, which however needs to be investigated in future when such a database is available.


## 1. Introduction

Estimation of design floods is needed for the design and planning of water infrastructure, flood risk assessment and various regulatory purposes. To estimate design floods, observed streamflow data is ideally needed; however, in many instances, the observed flood records are quite short or unavailable (in the case of ungauged catchments). Under these situations, regional flood frequency analysis (RFFA) is usually carried out which in essence attempts to transfer flood characteristics information from a group of gauged catchments to an ungauged catchment. In the literature, many RFFA approaches have been proposed, applied and tested in different countries around the world (e.g. Burn 1990a, 1990b; Hosking and Wallis, 1993; Rosbjerg and Madsen, 1994; Zrinji and Burn, 1994; Stedinger and Tasker, 1985; Stedinger and Tasker, 1986a; 1986b; Bates et al., 1998; Tasker and Stedinger, 1989; Pandey and Nguyen, 1999; Ouarda et al., 2001; Chokmani and Ourada, 2004; Rahman, 2005; Reis et al., 2005; Griffis and Stedinger, 2007; Gruber and Stedinger, 2008; Chebana and Ouarda, 2008; Micevski and Kuczera, 2009; Nezhad et al., 2010). RFFA can also be used to create larger data samples (using historical, paleoflood or extreme floods occurring in ungauged catchments) to reduce the uncertainties on high return period quantiles in a region (e.g. Gaume et al., in press).

In most of these regional estimation methods, hydrological statistics of interest (i.e. flood quantiles, mean flood, etc.) are estimated at gauged sites with relatively good data and are then transferred to an ungauged site to estimate the same statistic of interest. In transferring the information of at-site data to an ungauged site it

* Corresponding author. Tel.: +61 2 4736 0154; fax: +61 2 4736 0833.
*E-mail address:* a.rahman@uws.edu.au (A. Rahman).

0022-1694/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2012.02.012

is assumed that the data presents similar hydrological characteristics and that the region forms an 'acceptably homogenous region' (Hosking and Wallis, 1993). Formation of homogeneous regions has traditionally been based on geographic and administrative boundaries or even on a region-of-influence (ROI) analysis that seeks to satisfy some criteria (e.g. minimise the number of stations or the root mean square error of estimates) (I.E. Aust., 1987, 2001).

The degree of homogeneity of a proposed region is often judged on the basis of a dimensionless coefficient of the annual maximum flood series, standardised flood quantiles and similar statistics (e.g. Dalrymple, 1960; Wiltshire, 1986a, 1986b; Chowdhury et al., 1991; Lu and Stedinger, 1992; Hosking and Wallis, 1993; Fill and Stedinger, 1995; Cunderlik and Burn, 2006; Castellarin et al., 2008). For an ungauged site, since recorded streamflow data is not available, its degree of homogeneity in relation to a proposed homogeneous region cannot be assessed directly, hence the proximity of the ungauged site in geographical or catchment attributes space is used to assess its similarity with the proposed homogeneous region. Regions purely based on geographic and administrative boundaries may lack in hydrological similarity/homogeneity (Burn et al., 1997). It has been shown in several studies (Burn, 1990a, 1990b; Zrinji and Burn, 1994; Tasker et al., 1996; Eng et al., 2005; Merz and Blöschl, 2005; Eng et al., 2007a, 2007b) that the ROI approach performs better than the fixed region approach. In the ROI approach, regions can be formed based on the proximity in geographical or catchment attributes space.

Since the inception of the ROI procedure, it has been found the ROI can result in improved flood quantile estimates in terms of root mean square error and that ROI gave the flexibility of variable regions (Zrinji and Burn, 1996). They went on further to refine the initial ROI approach into a hierarchal ROI approach. The hierarchical ROI approach was found to perform well for the estimation of higher order moments (i.e. skewness), this is the case where more sites are needed to form a region. It was found in this study that the hierarchical ROI approach improved flood estimates in the extreme range. Tasker et al. (1996) compared five different methods for developing regional regression models to estimate flood quantiles at ungauged sites in Arkansas, US. The methods looked at traditional flood estimation regression approaches, multivariate techniques of cluster and discriminant analysis and a ROI approach based on geographical and catchment attribute space where the $n$ gauging sites with the smallest distance make up the ROI for site $i$. The results concluded that the ROI approach (based on catchment attributes space) outperformed the other methods based on the lowest root mean square error.

Eng et al. (2005) used different ROI approaches for estimating the 50 years average recurrence interval (ARI) flood quantile at ungauged sites in a case study for the Gulf Atlantic Rolling Plains of the southeastern United States. Ordinary Least Squares regression (OLS) was used to regress flood statistics against catchment characteristics for each ungauged site based on data from ROI containing the $n$ closest gauging sites in both geographical (GROI) and catchment attributes space (CROI). Model performance was based on the prediction errors from independent testing. From this testing, it was shown for the two ROI approaches using the $n$ closest gauging sites (based on geographical distance) was better than using a distance measure in catchment attributes space. They also found that GROI produced lower errors than CROI.

Merz and Blöschl (2005) examined the predictive performance of several flood regionalisation methods. They performed the assessment using a jackknife comparison of at-site estimated regionalised flood quantiles for 575 Austrian catchments. The ROI methods that only used catchment attributes performed relatively poorer to the methods that used geographical proximity. The ROI used in this study was then combined with multiple regression. Merz and Blöschl (2005) were able to demonstrate that when spatial dependency was incorporated, the ROI showed less random errors.

Eng et al. (2007a) proposed a hybrid ROI (HROI) which combined the GROI and CROI in a Generalised Least Squares (GLS) regression framework. They applied this method to 1091 catchments in the southeastern part of the United States to estimate the 50 years ARI flood quantile. Their study was able to show that the HROI yielded smaller root mean square estimation errors while also producing fewer extreme errors often found in either GROI or CROI. From this study it was concluded that for the 50 years ARI flood quantile, the similarity with respect to catchment attributes was important, however it was incomplete and that the consideration of the geographical proximity of the sites provided a useful surrogate for characteristics that were not included in the analysis. Eng et al. (2007b) went onto also present an enhanced GLS regression and ROI framework that is based on a leverage-guided ROI. This procedure used two newly defined ROI leverage and influence metrics. They applied their method to 996 catchments in the southeastern part of the United States. This new leverage-guided ROI regression provided improvements in terms of lower root mean square errors while also eliminating all the influential observations.

In Australia the Index Flood method has been researched (e.g. Bates et al., 1998; Rahman et al., 1999; Ishak et al., 2011). Bates et al. (1998) and Rahman et al. (1999) applied the Index Flood method to the state of Victoria (VIC) and part of New South Wales (NSW). It was found that no 'acceptably homogenous region' could be identified based on the Hosking and Wallis (1993) test and a number of different grouping methods. The methods by Bates et al. (1998) and Rahman et al. (1999) involved the assignment of ungauged catchments to a particular homogenous group identified (through the use of $L$-moments) on the basis of catchment characteristics as opposed to geographical proximity. The relationships sought were developed by statistical procedures such as canonical correlation analysis, tree based modelling and other multivariate statistical techniques. The results of this method also depended upon the correct assignment of an ungauged catchment to a homogenous group, thus any wrong assignment would greatly increase error in quantile estimation. Ishak et al. (2011) recently presented a study for the state of NSW where it was identified that a simple scaling approach to flood estimation is feasible, however no homogenous region could be found for the application of the Index Flood method. With the existence of large predictive uncertainty and the heterogeneity that plagues Australian catchments, an approach is needed that can deal with heterogeneity and predictive uncertainty in an efficient manner. For instance a method is needed that may perform reasonably well in the case of slight to medium heterogeneity.

The ROI method in Australia has recently received much attention (Hackelbusch et al., 2009; Rahman et al., 2009) because of its flexible and easy integration with a variety of RFFA methods and that it may deal effectively with the highly heterogeneous Australian catchment conditions. In the ROI approach, the site of interest is assumed to form its own 'unique' region. The ROI may be applied in a variety of ways such as by geographical distance or even in multi dimensional catchment attributes space defined by catchment slope, rainfall intensity, catchment area or other catchment and climatic attributes. After applying this delineation, estimation techniques such as the USGS Quantile Regression Technique (QRT) (Thomas and Benson, 1970) may be used.

A more efficient approach would be the application of GLS regression (Kuczera, 1983; Stedinger and Tasker, 1986a, 1986b; Tasker and Stedinger, 1989; Reis et al., 2005; Griffis and Stedinger, 2007) which accounts for correlated flood data, different record lengths and moreover distinguishes between sampling error and model error. The use of regional GLS regression (Tasker et al., 1986; Tasker and Stedinger, 1989; Pandey and Nguyen, 1999; Madsen and Rosbjerg, 1997; Madsen et al., 2002) and moreover Bayesian GLS estimation methods have been shown to be more accurate in estimating flood quantiles and statistics than using at-site flood frequency analysis alone (Reis et al., 2005; Micevski and Kuczera, 2009).

While both the ROI and GLS regression have been applied before in a QRT framework (Eng et al., 2007b), we are unaware of any comprehensive comparison between ROI and fixed regions in a BGLS framework, moreover there has been no solid comparison between the estimation of quantiles and the parameters of distributions in a ROI framework. Regionalising the parameters of a probability distribution (which is referred to as Parameter Regression Technique (PRT) in this study) offers three significant advantages over the QRT:

1. It ensures flood quantiles increase smoothly with increasing ARI, an outcome that may not always be achieved with the QRT. The flood quantiles obtained from the PRT may also be used to determine whether the flood quantiles derived from the QRT provides similar and consistent results.
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. It permits quantiles to be estimated for any ARI within the limits of the developed RFFA method.

The aim of this paper therefore is twofold: comparison of the predictive performance of (i) fixed regions and ROI; and (ii) the QRT and PRT. The ROI method used in this paper improves on the current ROI approaches (e.g. Tasker et al., 1996) where we seek to minimise the regression models predictive error variance rather than selecting or assuming a fixed number of sites to minimise a distance metric.

We focus on the ungauged catchment case for which frequency regionalisation is more challenging. The selection of the final set of predictor variables for use with GLS regression usually involves a stepwise variable selection search based on OLS regression (e.g. Tasker et al., 1996). This paper improves upon predictor variable selection by searching for the set of variables that minimises the model error variance, rather than the total (sampling plus model) error variance and by satisfying a number of statistical diagnostic metrics.

We use a comprehensive data set of small to medium sized catchments in south east Australia covering the states of VIC, NSW and Queensland (QLD). To make a thorough comparison of predictive performance between the fixed regions and ROI for both the QRT and PRT we adopt a one-at-a-time validation approach where we compare regional flood estimates (by QRT and PRT) to at-site flood frequency estimates. The one-at-time validation gives us an independent measure of how well each of the methods would perform for the ungauged catchment case.

We also present a residual analysis in a GLS framework where three sources of uncertainty (model error, sampling error and the uncertainty due to the unknown coefficients being estimated) are used. The residual analysis presented also provides insight into the overall regional performance of the methods and identifies any major outliers that may be affecting model consistency.

## 2. Study area and data

The catchments used in this paper are all located in the eastern coast of Australia. The physiography ranges from the lowlands in the western part of Victoria (VIC) with mean catchment elevations of less than 300 metres above sea level (ASL), up to higher catchments in the eastern part of VIC and New South Wales (NSW) with a mean catchment elevation of about 800 ASL. In Queensland (QLD) the catchments are mainly low to medium lying with mean catchment elevations in the order of 600 ASL.

The mean annual rainfall ranges from 400 mm/year in the north west of VIC to 3500 mm/year along the eastern parts of QLD. Winter dominated rainfall is common in VIC; while summer dominated rainfalls are more common for northern NSW and QLD. The locations of the gauged catchments are shown in Fig. 1.

The analysis undertaken in this paper makes use of (i) observed annual maximum flood series of catchments ranging in area from 3 to 1010 km$^2$, and (ii) climatic and catchment characteristics data. As a preliminary step, the annual maximum flood series of candidate catchments were chosen based on (i) catchment area (ii) record length (iii) regulation of catchment (iv) urbanisation of catchment (v) landuse change (vi) quality of data and (vii) climate variability and change. Further reading on the details of these methods can be found in Haddad et al. (2010a, 2010b).

Missing data points in the annual maximum flood series were in filled where possible by two methods. Method 1 involved comparing the monthly instantaneous maximum data (IMD) with monthly maximum mean daily data (MMD) at the same station. If a missing month of IMD flow corresponded to a month of very low MMD flow, then that was taken to show that the annual maximum did not occur during that missing month. Method 2 involved a simple linear regression of the annual MMD flow against the annual IMD series of the same station. It must be mentioned that the regression equations developed were used for filling gaps in the IMD record, but not to extend the overall period of record.

Rating curve extrapolation errors were identified by using a rating ratio test and treated using the in-built procedure 'rating curve error' case in at-site flood frequency analysis software (FLIKE) (Kuczera, 1999). Outliers were identified using the Grubbs and Beck (1972) method, which is also recommended in Bulletin 17B by the US Water Resources Council (IACWD, 1982). Low outliers were censored using the FLIKE software (that is, the information that there was no flood in that year was taken into account). High outliers were found only in few cases, these however were retained as there was no evidence of these points being a data error. The selected stations did not show any trend.

The finally selected data set consists of 399 catchments (Fig. 1) with annual maximum flow record lengths ranging from 25 to 94 years (maximum record length for NSW: 75 years, mean and standard deviation: 37 and 11 years respectively; maximum record length for VIC: 52 years, mean and standard deviation: 33 and 5 years respectively and maximum record length for QLD: 94 years, mean and standard deviation: 40 and 15 years respectively). Based on the findings from previous studies in Australia (e.g. Rahman, 2005), a total of 14 explanatory variables were used, as outlined below:

(i) catchment area in km$^2$ (*area*);
(ii) design rainfall intensities for the 2 years ARI with 1 and 12 h durations ($^2I_1$, $^2I_{12}$), 50 years ARI with 1 h duration ($^{50}I_1$) and 50 years ARI with 12 and 72 h duration ($^{50}I_{12}$, $^{50}I_{72}$), all expressed in mm/h;
(iii) mean annual rainfall in mm/y (rain);
(iv) mean annual evapo-transpiration expressed in mm/y (evap);
(v) design rainfall intensity values in mm/h $I_{ARI,tc}$ (where ARI = 2, 5, 10, 20, 50 and 100 years and $t_c$ = time of concentration (hour), estimated from $t_c = 0.76(\text{area})^{0.38}$);
(vi) stream density expressed in km/km$^2$ (*sden*);
(vii) main stream slope expressed in m/km (*S1085*);
(viii) stream length expressed in km;

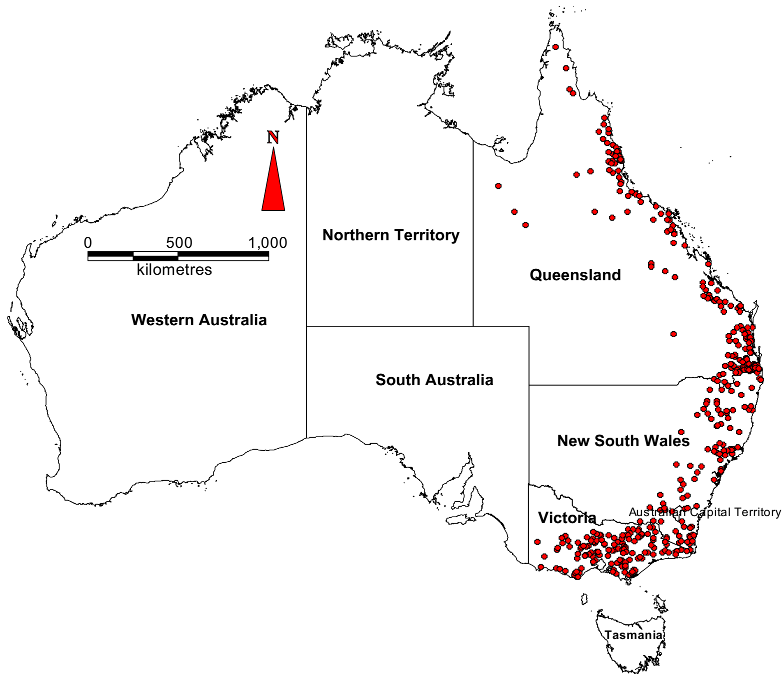


**Fig. 1.** Location of gauged catchments in NSW, VIC and QLD.

(ix) forest cover expressed as a percentage (%) of catchment area (*forest)* and
(x) quaternary sediment area expressed as a percentage of catchment area (*QSA*, used in VIC only).

To reduce possible biases due to highly skewed log-transformed explanatory variables, all explanatory variables data were centred by subtracting its log-mean value, in this case the intercept term in the regression equation represents the mean of the logarithm of the observed dependent variable data.

## 3. Methodology

### *3.1. At-site flood frequency analysis and quantile and parameter regression techniques*

At-site flood quantiles for ARIs of 2, 5, 10, 20, 50 and 100 years were estimated by the at-site flood frequency analysis software FLIKE (Kuczera, 1999) using the log-Pearson type 3 (LP3) distribution with the Bayesian parameter estimation procedure as described in Kuczera (1999). The LP3 distribution was chosen based on two reasons. Firstly the LP3 distribution is the currently recommended at-site flood frequency distribution in Australian Rainfall and Runoff (National guideline for flood estimation). Secondly the LP3 has shown consistently better results in the past studies for Australian catchments (Haddad et al., 2010a, 2010b; Haddad et al., 2009) and thus is adopted for this study. No prior information was used in fitting the LP3 distribution. The parameters of the LP3 distribution (i.e. mean, standard deviation and skewness) were also extracted from the FLIKE software.

To regionalise the flood quantiles the sampling covariance matrix ($\Sigma$) of the LP3 distribution is required. Tasker and Stedinger (1989) and Griffis and Stedinger (2007) (p. 84, Eq. (4)) provide the approximate estimator of the components of $\Sigma$ matrix of the LP3 distribution. The skew and standard deviation in the $\Sigma$ 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 ($\rho_{ij}$) is estimated as a function of the distance between sites *i* and *j*;
(ii) the standard deviations (of the logarithms of annual maximum flood series) $\sigma_i$ and $\sigma_j$ are estimated using a separate Ordinary Least Squares (OLS) and Generalised Least Squares (GLS) regression using the explanatory variables used in the study (given in Section 2); and
(iii) the regional skew (of the logarithms of annual maximum flood series) is used in place of the population skew $\gamma$ as suggested by Tasker and Stedinger (1989). This analysis above used the regional estimates of the standard deviation and skew obtained from Bayesian GLS (BGLS) 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 Parameter Regression Technique (PRT), we adopted the GLS regression (Tasker and Stedinger, 1989; 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, standard deviation, and skew coefficient of the logarithms of the annual maximum flood series). The regional values of standard deviation and skew were taken from the $\Sigma$ matrix of the flood quantile modelling as mentioned above. The covariance matrix for the mean flood was obtained following Stedinger and Tasker (1985, 1986a, 1986b).

### *3.2. Generalised Least Squares Regression*

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:

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

where $X_{ij}(j = 1, \ldots, k)$ are explanatory variables, $\beta_j$ are the regression coefficients, $\delta_i$ is the model error which is assumed to be normally and independently distributed with model error variance $\sigma_\delta^2$ 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 $\eta_i$ must be introduced into the model so that:

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

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

$$E\lfloor \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T \rfloor = \boldsymbol{\Lambda}(\sigma_\delta^2) = \sigma_\delta^2 \mathbf{I} + \Sigma(\hat{\mathbf{y}}) \tag{3}$$

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. The variance of $\eta_i$ depends on the record length available at each site and the cross correlation of the sites flood data. Therefore the observed regression model errors are a combination of time-sampling error $\eta_i$ and an underlying model error $\delta_i$.

In this regional framework, $\sigma_\delta^2$ can be viewed as a heterogeneity measure. Madsen and Rosbjerg (1997) and Madsen et al. (2002) showed that the regional average GLS estimator is a general extension of the record-length-weighted average commonly applied in the Index Flood method; however the record-length-weighted average estimator neglects inter-site correlation and regional heterogeneity (Stedinger et al., 1993; Stedinger and Lu, 1995).

The GLS estimator of $\beta$ and its respective covariance matrices for known $\sigma_\delta^2$ are given by:

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

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

The model error variance $\sigma_\delta^2$ can be estimated by either generalised method of moments (MOM) or maximum likelihood (ML) estimators as described by Stedinger and Tasker (1985, 1986a, 1986b). The MOM estimator is determined by iteratively solving Eq. (6) along with the generalised residual mean square error equation:

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

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

### 3.3. Bayesian GLS regression

In a Bayesian framework, the parameters of the model are considered to be random variables, whose probability density function should be estimated. The Bayesian approach combines any data with prior information (if available) about the parameters being estimated. This information usually is established from other data sets, previous studies or specific knowledge about the behavior of the system being analysed. Parameter estimation is made through the posterior distribution which is developed using Bayes' rule: (Zellner, 1971):

$$p(\theta|\hat{y}) = \frac{p(\hat{y}|\theta)\zeta(\theta)}{\int p(\hat{y}|\theta)\zeta(\theta)d\theta} \tag{7}$$

Here, $p(\theta|\hat{y})$ is the posterior distribution of the parameter vector $\theta$ given the information $\hat{y}$ in the available data set, $p(\hat{y}|\theta)$ is the likelihood function for the data, and $\zeta(\theta)$ is the prior distribution of $\theta$. The denominator is a normalising constant which ensures that the area under the posterior pdf equals one. Reis et al. (2005) developed a Bayesian approach to estimate the regional flood model parameters and showed that the Bayesian approach can provide a realistic description of the possible values of the modal error variance, especially in the case where sampling error tend 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 $\beta$ coefficients 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 coefficients as suggested by Reis et al. (2005). This prior is considered to be almost non-informative, which produces a probability distribution function (pdf) that is generally flat in the region of interest. The prior information for the model error variance $\sigma_\delta^2$ is represented by an informative one-parameter ($\lambda$) exponential distribution, which represents the reciprocal of the prior mean of the model error variance, which is given by Eq. (8) and can be found in Reis et al. (2005):

$$\zeta(\sigma_\delta^2) = (1/\lambda)e^{-\sigma_\delta^2/\lambda}, \quad \text{where } \sigma_\delta^2 > 0 \tag{8}$$

The likelihood function for the data as suggested by Reis et al. (2005) is considered to be a multivariate normal distribution, so that

$$L(\beta, \sigma_\delta^2|I) = (2\pi)^{-n/2}\frac{1}{|\Lambda|^{1/2}} \exp\left[-0.5(\hat{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})^T\Lambda^{-1}((\hat{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})\right] \tag{9}$$

where $n$ is the number of sites in a region.

The marginal posterior distribution of the model error variance can be computed by integrating the joint posterior distribution over the possible values of the $\beta$ coefficients (Reis et al., 2005), to obtain:

$$f(\sigma_\delta^2|I) = \int f(\beta, \sigma_\delta^2|I)d\beta \propto \int f(I|\beta, \sigma_\delta^2)\zeta(\beta, \sigma_\delta^2)d\beta \tag{10}$$

where $f(\beta, \sigma_\delta^2|I$ is the joint posterior of the parameters, $f(I|\beta, \sigma_\delta^2)$ is the likelihood function, and $\zeta(\beta, \sigma_\delta^2)$ is the joint prior for $\beta$ and $\sigma_\delta^2$. If one uses a relatively non-informative prior on the $\beta$ coefficients as was done in this study, the marginal posterior distribution for the model error variance, except for the normalizing constant, is:

$$f(\sigma_\delta^2|I) \propto \left[|\boldsymbol{\Lambda}||\mathbf{X}^T\boldsymbol{\Lambda}^{-1}\mathbf{X}|\right]^{-1/2} \times \exp\left[-0.5(\hat{\mathbf{y}} - \mathbf{X}\hat{\beta})^T\boldsymbol{\Lambda}^{-1}\left(\hat{\mathbf{y}} - \mathbf{X}\hat{\beta}\right)\right]\zeta(\sigma_\delta^2) \tag{11}$$

where $|\boldsymbol{\Lambda}|$ signifies the determinant of matrix $\boldsymbol{\Lambda}$.

Using Eq. (11) the marginal pdf, mean, and variance of $\sigma_\delta^2$ can be computed numerically. It also follows that the posterior moments of the $\beta$ coefficients can be computed numerically as well.

Reis et al. (2005) discusses the derivation of the choice of a prior for the model error variance for regionalising the skew. For the regionalisation of skew, we employed a value for the prior mean of the model error variance equal to 6 following Reis et al. (2005).

A negative model error variance is unrealistic as noted by Reis et al. (2005) which can happen in the case of GLS regression, in particular for the skew model when sampling error dominates over the model error. This was observed in previous GLS based regional frequency analysis applications with Australian data (e.g. Haddad et al., 2010a, 2010b, 2011). In this situation Eq. (6) may introduce further uncertainty into the regional model. A Bayesian estimator of the model error variance (Eq. (11)) as discussed above may be used to safeguard against this happening, as adopted in this study. Further details can be found in Reis et al. (2005) and Micevski and Kuczera (2009). In summary, the Bayesian estimator offers a better way of dealing with the model error variance and quantifying associated uncertainty about it.

To derive the prior distribution for the standard deviation, mean flood and flood 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.

### 3.4. Selection of predictor variables

This section describes the approach adopted for selecting the predictor variables that should be included in the prediction equations. The approach for selecting predictor variables used in this paper provides improvements over current methods used to justify model selection in the GLS regression. Provided below is a brief discussion on the BGLS regression statistics that guided our model selection.

We use a procedure similar to forward stepwise regression utilising all the sites for each state (separate regression for each state) and initially adopting just a constant term in the regression equation. The model error variance and its standard error are noted. We then add predictor variables starting with *area* followed by different combinations of other variables. In all, 16 different combinations of predictor variables were used for the mean, standard deviation and skew models, while 25 combinations were trialled for the flood quantile models.

The choice for the preferred regional BGLS model was the combination that best satisfied all the following statistical measures:

(i) Minimum model error variance (MEV), and not on the total error variance (model plus sampling error as with current approaches), also the use of the expected MEV from BGLS ensures non-negative and zero values,
(ii) The minimum average variance of prediction (AVP) for a new and old station (AVPN) and (AVPO) (Gruber and Stedinger, 2008). Here AVPN refers to AVP for a new site within the region of interest which has not been used to develop the regression model. The AVPO refers to the AVP for a site which has been used to develop the regression model. As we are interested in making predictions at ungauged sites the AVP penalises the inclusion of extra independent variables because it accounts for the sampling variances of the regression coefficients.
(iii) The significance of the regression coefficient values ($\beta$) obtained was evaluated using the Bayesian plausibility value (BPV) as developed by Reis et al. (2005) and Gruber and Stedinger (2008). 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 in this study was carried out at the 5% significance level.
(iv) The Akaike and Bayesian information criteria (AIC and BIC) penalise more heavily a model with a greater number of predictors (i.e. the inclusion of a predictor variable must significantly improve the model if it is to be included). In practice, after the computation of the posterior mean of the AIC and BIC for all of the competing models, one selects the model with the minimum AIC and BIC value.
(v) The highest Pseudo $R^2$ value ($\overline{R}^2_{GLS}$). Reis et al. (2005) proposed a pseudo co-efficient of determination ($\overline{R}^2_{GLS}$) appropriate for use with the GLS regression. For the traditional $R^2$, both the Sum-of-Squared Errors (SSE) and the Total-Sum-of-Squared deviations about the mean (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 $y_i$.
(vi) A predictor variable having an estimated coefficient (other than the constant) that was less than two posterior standard deviations away from zero was rejected (this shows the relative importance of the predictor) (Hackelbusch et al., 2009). In all the cases the simplest model was preferred.

### 3.5. Formation of regions, fixed and ROI

The fixed region BGLS regression analysis as above identifies the catchment characteristics that best account for heterogeneity by minimising the model error variance. However, it is assumed that there remains a possible spatial structure in the model error residuals. With this in mind the model error variance therefore within possible sub regions of the fixed region should be less than the fixed region model error variance. This is investigated further in this paper (see Section 4)

It is in this framework that the ROI approach was applied to the parameters (i.e. mean, standard deviation and skew) and flood quantiles of the LP3 distribution to further reduce the heterogeneity unaccounted for by the fixed region BGLS model.

The ROI approach in this paper uses the distance between sites as the distance metric (i.e. geographic proximity). We apply the ROI within the state boundaries in the following way. In the first iteration, the 15 nearest stations to the site of interest are selected and a regional BGLS regression is performed and the predictive variance (Eqs. (6) and (11)) is noted. The second iteration proceeds with the next five closest stations being added to the ROI and repeating the regression. This procedure terminates when all eligible sites have been included in the ROI. The ROI for the site of interest is then selected as the one which yields the lowest predictive variance.

This approach is fundamentally different to that of Tasker et al. (1996) in that it seeks to minimise

(i) the regression model's predictive error variance rather than selecting or assuming a fixed number of sites that minimise a distance metric in catchment characteristic space;
(ii) the ROI criterion of Tasker et al. (1996) cannot guarantee minimum predictive variance; and
(iii) moreover, the selection of sites that are minimally different in catchment characteristic space may result in greater uncertainty in the estimated regression coefficients.

It should be noted that the predictive error variance has two terms associated with it:

(i) the model error variance; and
(ii) the predictive variance arising from uncertainty in the estimated regression coefficients.

The first term is the posterior expected value of the model error variance estimated using the approach of Reis et al. (2005), see Section 3.3 and Eq. (11) – this is always non-zero and guards against situations where the most likely value of the model error variance is zero. The second term effectively guards against the ROI favouring fewer sites to minimise the model error variance; indeed as the number of sites is reduced the model error variance is likely to be offset by an increase in uncertainty in the estimated regression coefficients (i.e. $\boldsymbol{\beta}$).

### 3.6. 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 (as discussed already in Section 3.4) and the standard error of prediction. An analysis of variance for the BGLS models is undertaken to examine which portion of the total error (sampling or model) dominates the regional analysis for both the fixed region and ROI methods. We also present the standardised residuals and $Z$ score analysis in a GLS framework which is used to identify outlier sites; absence of outlier in regression diagnostics indicates the overall adequacy of the regional model. These statistics are described below.

If the standardised residuals have a nearly normal distribution (to be determined in the residual analysis, see below), the standard error of prediction in percent (SEP) (Tasker et al., 1986) for the true flood quantile or parameter estimator is described by:

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

Important to this study is the assessment of the adequacy of the regional regression model in its application to ungauged catchments. The measure of the raw residual ($r_i$), which is the difference between the sample (at-site estimate) and regional estimates of the LP3 parameter or flood quantile can be assessed initially for major deviations. However, interpreting the raw residual may be misleading as the raw residual has three sources of uncertainty: model error, sampling error and uncertainty due to regression coefficients being unknown.

In this study we use a standardised residual $r_{si}$, which is the raw residual divided by its standard deviation defined as the square root of the sum of the predictive variance of the LP3 parameter or flood quantile and its sampling variance given by the appropriate diagonal element of the sampling covariance matrix. This yields the definition

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

To assess the adequacy of the estimated LP3 parameters and flood quantiles from QRT and PRT, standardised residuals, referred to as $Z$-scores were used. For site $i$ and a given ARI, the $Z$-score is

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

Here the numerator is the difference between the at-site flood quantile and regional flood quantile (estimated from the developed prediction equation) and the denominator is the square root of the sum of the variances of the at-site ($\sigma^2_{ARI,i}$) and regional ($\hat{\sigma}^2_{ARI,i}$) flood quantiles in natural logarithm space.

It is reasonable to assume that the errors in the two estimators are independent because $Q_{ARI,i}$ is an unbiased estimator of the true quantile estimators based upon the at-site data, whereas the error in $\hat{Q}_{ARI,i}$ 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.

### 3.7. Evaluation statistics

A one-at-a-time cross validation procedure was applied to assess the performance of the RFFA methods. The site that is left out in building the model is in effect being treated as an ungauged site. Since all the sites in the database are being treated as ungauged for ROI this automatically satisfies the one-at-a-time validation approach. The following performance statistics were calculated from the fixed and ROI analysis: absolute (abs) relative median error ($\text{RE}_r$) in % over $n$ sites and the relative root mean square error ($\text{RMSE}_r$) in % as described below.

$$\text{RE}_r = Median\left[abs\left(\frac{Q_{pred_i} - Q_{obs_i}}{Q_{obs_i}}\right)\right]_{i=1}^{n} \tag{15}$$

$$\text{RMSE}_r = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{Q_{pred_i} - Q_{obs_i}}{Q_{obs_i}}\right)^2} \tag{16}$$

where $Q_{obs_i}$ is the observed flood quantile at site $i$ obtained from at-site flood frequency analysis estimated using FLIKE (Kuczera, 1999), $Q_{pred_i}$ is the predicted flood quantile at site $i$ from the regional prediction equation from QRT and PRT and $n$ is the number of sites in the region. The $\text{RE}_r$ (%) and $\text{RMSE}_r$ (%) provide an indication of the overall accuracy of the regional model. The model with minimum $\text{RE}_r$ is always preferred. For $\text{RMSE}_r$ the smallest value between the two competing models with the same number of parameters is generally preferred.

It should be noted here that both the $Q_{pred}$ and $Q_{obs}$ values have uncertainties associated with them, and in particular, the $Q_{obs}$ values are subject to errors due to the annual maximum flood record length, rating curve extrapolation errors, selection of probability distribution and associated parameter estimation procedures. The above error statistics thus give some guidance about the relative accuracy of the method and should not be taken as the true uncertainty associated with the method.

## 4. Results

### 4.1. Selection of predictor variables

The stepwise procedure for selecting the best set of catchment characteristics resulted in the following equations for the LP3 mean ($\mu$), standard deviation ($\sigma$), skewness ($\gamma$) and the flood quantiles ($Q_{ARI}$) for each state NSW, VIC and QLD. The regression equations are presented in general form below, while the final results of the equations for NSW are provided in Table 1. The final results of VIC and QLD can be seen in Appendix A.

$$\mu = \beta_0 + \beta_1(area) + \beta_2({}^2I_{12}) \text{ for NSW, VIC and QLD} \tag{17}$$

$$\sigma = \beta_0 - \beta_1(rain) - \beta_2(S1085) \text{ for NSW} \tag{18}$$

$$\gamma = -\beta_0 - \beta_1(area) - \beta_2(forest) \text{ for NSW} \tag{19}$$

$$\sigma = \beta_0 - \beta_1(rain) + \beta_2(evap) \text{ for VIC} \tag{20}$$

$$\gamma = -\beta_0 + \beta_1(rain) - \beta_2(evap) \text{ for VIC} \tag{21}$$

$$\sigma = \beta_0 - \beta_1(area) - \beta_2({}^2I_1) \text{ for QLD} \tag{22}$$

$$\gamma = -\beta_0 - \beta_1({}^{50}I_{72}) + \beta_2(rain) \text{ for QLD} \tag{23}$$

$$\ln(Q_{ARI}) = \beta_0 + \beta_1(area) + \beta_2(I_{tc,ARI}) \text{ for NSW, VIC and QLD} \tag{24}$$

Tables 2a and 2b summarizes the model error variance (MEV) as expressed by its posterior mean value, for the regional models of the three LP3 parameters and the flood quantiles $Q_2$, $Q_{10}$ and $Q_{100}$ for each combination of catchment characteristics for NSW. Also provided in Tables 2a and 2b is the summary of the statistical measures used i.e. average variance of prediction for an old site (AVPO) and new site (AVPN), Akaike and Bayesian information criteria's (AIC) and (BIC), Bayesian plausibility value (BPV) and Pseudo $R^2$ ($\overline{R}^2_{GLS}$) to assess the best combination of catchment characteristics to predict the three parameters and flood quantiles of the LP3 distribution. Appendix A shows the table of the final results for VIC and QLD.

Fig. 2 shows the MEV, standard error of the MEV and $\overline{R}^2_{GLS}$ values for the skew model. Combination 9 with a constant and two predictor variables *area* and *forest* showed the lowest MEV and the highest $\overline{R}^2_{GLS}$ as well as the lowest AIC and BIC. However the lowest AVPO and AVPN were found for combination 1 (a constant value – see Fig. 2).

**Table 1**
Summary of the final BGLS regression results for NSW.

| GLS regression model (NSW) | Regression coefficient | Posterior moment | |
|---|---|---|---|
| | | Mean | Standard deviation |
| Mean $\mu$ | $\sigma_\delta^2$ | 0.29 | 0.051 |
| | $\beta_0$ (constant) | 4.09 | 0.092 |
| | $\beta_1$ (*area*) | 0.67 | 0.053 |
| | $\beta_2$ ($^2I_{12}$) | 2.31 | 0.21 |
| Standard deviation $\sigma$ | $\sigma_\delta^2$ | 0.067 | 0.013 |
| | $\beta_0$ (constant) | 1.25 | 0.12 |
| | $\beta_1$ (*rain*) | −0.61 | 0.11 |
| | $\beta_2$ (*S1085*) | −0.13 | 0.040 |
| Skewness $\gamma$ | $\sigma_\delta^2$ | 0.0125 | 0.012 |
| | $\beta_0$ (constant) | −0.42 | 0.072 |
| | $\beta_1$ (*area*) | −0.092 | 0.048 |
| | $\beta_2$ (*forest*) | −0.094 | 0.053 |
| Flood quantiles | $\sigma_\delta^2$ | 0.31 | 0.055 |
| $Q_{ARI=2}$ | $\beta_0$ (constant) | 4.06 | 0.13 |
| | $\beta_1$ (*area*) | 1.26 | 0.086 |
| | $\beta_2$ ($I_{tc,ARI=2}$) | 2.42 | 0.24 |
| $Q_{ARI=5}$ | $\sigma_\delta^2$ | 0.23 | 0.042 |
| | $\beta_0$ (constant) | 5.11 | 0.092 |
| | $\beta_1$ (*area*) | 1.19 | 0.072 |
| | $\beta_2$ ($I_{tc,ARI=5}$) | 2.08 | 0.20 |
| $Q_{ARI=10}$ | $\sigma_\delta^2$ | 0.23 | 0.045 |
| | $\beta_0$ (constant) | 5.56 | 0.10 |
| | $\beta_1$ (*area*) | 1.14 | 0.074 |
| | $\beta_2$ ($I_{tc,ARI=10}$) | 1.93 | 0.21 |
| $Q_{ARI=20}$ | $\sigma_\delta^2$ | 0.25 | 0.050 |
| | $\beta_0$ (constant) | 5.91 | 0.11 |
| | $\beta_1$ (*area*) | 1.09 | 0.078 |
| | $\beta_2$ ($I_{tc,ARI=20}$) | 1.79 | 0.22 |
| $Q_{ARI=50}$ | $\sigma_\delta^2$ | 0.35 | 0.060 |
| | $\beta_0$ (constant) | 6.55 | 0.13 |
| | $\beta_1$ (*area*) | 1.01 | 0.081 |
| | $\beta_2$ ($I_{tc,ARI=50}$) | 1.73 | 0.24 |
| $Q_{ARI=100}$ | $\sigma_\delta^2$ | 0.35 | 0.075 |
| | $\beta_0$ (constant) | 6.47 | 0.34 |
| | $\beta_1$ (*area*) | 0.97 | 0.12 |
| | $\beta_2$ ($I_{tc,ARI=100}$) | 1.50 | 0.29 |

The BPV were used to carry out a hypothesis test (at the 5% significance level) on the predictors of combination 9. The BPVs were found to be 6% and 7% for *area* and *forest* respectively, while this showed the variables not to be significant, these values are not considered overly high. Both the posterior coefficients $\beta_1$ and $\beta_2$ were less than two posterior standard deviations away from zero supporting the results from the BPV test that these variables are not significant.

In this case it may be possible to adopt a regional skew value for NSW without using any prediction equation/predictor variable. This finding is consistent with Gruber and Stedinger (2008) who found that a constant model for a regional skewness was the best model for a large region in the southeastern part of the United States. This is also supported by the fact that there was only a modest difference in the MEV values. Combinations 9 and 1 however were both adopted and tested in this study with the PRT approach.

A similar outcome was observed for the standard deviation model where the MEVs were very similar for combinations 12 and 1 (figure not shown). Combination 12 was adopted that had *slope* and *rain* as predictor variables. Indeed AVPO, AVPN, BIC and AIC were the lowest for this combination. Both the posterior coefficients $\beta_1$ and $\beta_2$ were well identified in the regression equations being more than two times the posterior standard deviation away from zero. The BPVs were 2% indicating the relative significance of these variables.

For the mean flood, combination 6 (constant, area and $^2I_{12}$) had the smallest MEV. The posterior coefficients of $\beta_1$ and $\beta_2$ in this combination, were at least 5 and 11 times the posterior standard deviation away from zero, which shows that $\beta_1$ and $\beta_2$ are well identified in the prediction equation. All the statistical criteria were found to be in favour of combination 6.

Fig. 3 shows an example plot of the statistics used in selecting the best set of predictor variables for the fixed region flood quantile model. According to the MEV, combinations 19, 18, 20, 23, 16, 6, 4, 25 and 10 were potential sets of predictor variables for the $Q_{10}$ model. Combinations 18, 19, 20 and 23 contained 3–4 predictor variables while combinations 16, 6, 4, 25 and 10 contained 2 predictor variables with similar MEVs and $\bar{R}^2_{GLS}$.

The AVPO and AVPN and the AIC and BIC values favoured combination 10, and hence this was finally selected as the best set of predictor variables for the $Q_{10}$ model which includes area and design rainfall intensity $I_{tc,10}$. Both posterior coefficients $\beta_1$, and $\beta_2$ were found to be 9 times the posterior standard deviation away from zero suggesting these two variables are well defined in the prediction equation. Combination 10 was selected for all the flood quantile prediction equations (ARI = 2–100 years). The BPVs for the regression coefficients associated with the variable *area* and design rainfall intensity $I_{tc,ARI}$ for the QRT over all the ARIs were found to be significant with values smaller than 0.01%.

### *4.2. Region of influence BGLS regression vs. fixed regions for parameter and quantile regression techniques*

#### *4.2.1. Regression diagnostics – Pseudo Analysis of Variance*

The Pseudo Analysis of Variance (ANOVA) tables for the $Q_{20}$ model and the parameters of the LP3 distribution (mean and skew shown only) are presented in Tables 3–5 for the fixed regions and ROI for NSW, VIC and QLD. The Pseudo ANOVA table describes how the total variation among the $\hat{y}_i$ values (predicted values) can be apportioned between that explained by the model error and sampling error. This is an extension of the ANOVA in the OLS regression which does not recognise and correct for the expected sampling variance (Reis et al., 2005). An error variance ratio (EVR) is used in Pseudo ANOVA, which is the ratio of sampling error variance to model error variance. An EVR greater than 0.20 may indicate that the sampling variance is not negligible when compared to the model error variance, which suggests the need for a GLS regression analysis (Gruber et al., 2007).

For the LP3 parameters, the sampling error (i.e. EVR) increases as the order of moment increases, this can be clearly seen for all the states in Tables 3 and 4. For example, for NSW the EVR for the mean flood model for ROI is 0.3 (i.e. the sampling error is only 0.3 times of the model error) (Table 3), the corresponding EVR value for the skew model (Table 4) is 18 (i.e. the sampling error is 18 times of the model error). The ROI shows a reduced model error variance for all the three states (i.e. a reduced heterogeneity), in particular for the mean flood model, as compared to the fixed regions. For example, for NSW (Table 3) the model error variances for the fixed region and ROI are 27.7 and 16.5, respectively. It was found that the model error dominated the regional analysis for the mean flood and the standard deviation models (results not shown) for both the fixed regions and ROI for all the states. For the ROI, the mean flood model also shows a much higher model error variance than those of the standard deviation and skew models. These results based on the model error variance alone indicate that the mean flood has the greater level of heterogeneity associated with its regionalisation as compared to the standard deviation and skew. The ROI, however shows a higher EVR than the fixed regions e.g. for the mean flood model

**Table 2a**
Summary of the catchment characteristics and statistical measures used in the forward stepwise regression for the parameters of the LP3 distribution for NSW.

| Combination | Catchment characteristics[a] | LP3 parameter | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean | | | | | | | Standard deviation | | | | | | | Skewness | | | | | | |
| | | $\sigma_\delta^2$ | AVPO | AVPN | AIC | BIC | BPV% | $\overline{R}^2_{GLS}$ (%) | $\sigma_\delta^2$ | AVPO | AVPN | AIC | BIC | BPV% | $\overline{R}^2_{GLS}$ (%) | $\sigma_\delta^2$ | AVPO | AVPN | AIC | BIC | BPV% | $\overline{R}^2_{GLS}$ (%) |
| 1 | Const | 0.92 | 0.94 | 0.92 | 1.22 | 1.22 | 0 | 0 | 0.099 | 0.10 | 0.10 | 0.13 | 0.13 | 0 | 0 | 0.0135 | 0.019 | 0.018 | 0.156 | 0.156 | <0.1 | 0 |
| 2 | Const, area | 0.69 | 0.71 | 0.68 | 0.76 | 0.78 | 0, 0 | 39 | 0.098 | 0.10 | 0.10 | 0.13 | 0.13 | 0, 10 | 4 | 0.0132 | 0.021 | 0.021 | 0.080 | 0.082 | <0.1, 3 | 50 |
| 3 | Const, area, ${}^{2}I_{1}$ | 0.36 | 0.38 | 0.35 | 0.34 | 0.35 | 0, 0, 0 | 74 | 0.097 | 0.10 | 0.10 | 0.13 | 0.13 | 0,13, 19 | 6 | 0.0131 | 0.025 | 0.024 | 0.079 | 0.083 | <0.1, 3, 68 | 52 |
| 4 | Const, area, ${}^{50}I_{1}$ | 0.34 | 0.36 | 0.34 | 0.38 | 0.40 | 0, 0, 0 | 70 | 0.096 | 0.10 | 0.10 | 0.13 | 0.13 | 0,10, 20 | 6 | 0.0131 | 0.025 | 0.024 | 0.079 | 0.083 | <0.1, 3, 72 | 52 |
| 5 | Const, area, ${}^{50}I_{12}$ | 0.30 | 0.31 | 0.29 | 0.32 | 0.34 | 0, 0, 0 | 75 | 0.094 | 0.10 | 0.09 | 0.12 | 0.13 | 0,13, 10 | 8 | 0.0132 | 0.025 | 0.024 | 0.080 | 0.084 | <0.1, 3, 72 | 51 |
| 6 | Const, area, ${}^{2}I_{12}$ | 0.28 | 0.30 | 0.28 | 0.31 | 0.32 | 0, 0, 0 | 76 | 0.091 | 0.10 | 0.09 | 0.12 | 0.13 | 0,14, 6 | 10 | 0.0133 | 0.025 | 0.024 | 0.082 | 0.086 | <0.1, 3, 86 | 50 |
| 7 | Const, area, S1085 | 0.63 | 0.66 | 0.62 | 0.70 | 0.74 | 0, 0, 0.4 | 45 | 0.091 | 0.10 | 0.09 | 0.12 | 0.13 | 0,29, 8 | 8 | 0.0135 | 0.024 | 0.023 | 0.083 | 0.087 | <0.1, 4, 92 | 49 |
| 8 | Const, area, sden | 0.60 | 0.63 | 0.59 | 0.54 | 0.57 | 0, 0, 0.6 | 58 | 0.099 | 0.10 | 0.10 | 0.13 | 0.14 | 0,14, 58 | 4 | 0.0134 | 0.024 | 0.023 | 0.083 | 0.088 | <0.1, 4, 81 | 49 |
| 9 | Const, area, forest | 0.69 | 0.72 | 0.68 | 0.78 | 0.82 | 0, 0, 60 | 39 | 0.091 | 0.10 | 0.09 | 0.12 | 0.13 | 0,5, 7 | 9 | 0.0126 | 0.024 | 0.023 | 0.057 | 0.060 | <0.1, 6, 7 | 65 |
| 10 | Const, area, evap | 0.34 | 0.35 | 0.33 | 0.39 | 0.41 | 0, 0, 0.1 | 69 | 0.098 | 0.10 | 0.10 | 0.13 | 0.13 | 0,14, 26 | 6 | 0.0133 | 0.026 | 0.025 | 0.076 | 0.080 | <0.1, 2, 49 | 53 |
| 11 | Const, area, rain | 0.29 | 0.31 | 0.29 | 0.31 | 0.33 | 0, 0, 0.1 | 76 | 0.078 | 0.08 | 0.08 | 0.10 | 0.10 | 0,40, 1 | 26 | 0.0134 | 0.025 | 0.024 | 0.082 | 0.087 | <0.1, 2, 87 | 49 |
| 12 | Const, rain, S1085 | 0.92 | 0.96 | 0.90 | 1.24 | 1.31 | 0,37, 16 | 2 | 0.066 | 0.07 | 0.07 | 0.09 | 0.09 | 0,2, 1 | 35 | 0.0140 | 0.025 | 0.025 | 0.148 | 0.156 | 0,74, 87 | 10 |
| 13 | Const, sden, S1085 | 0.91 | 0.94 | 0.89 | 1.15 | 1.21 | 0,0.8, 82 | 9 | 0.090 | 0.09 | 0.09 | 0.12 | 0.13 | 0,60, 5 | 8 | 0.0139 | 0.025 | 0.024 | 0.140 | 0.148 | 0,74, 51 | 14 |
| 14 | Const, evap, sden | 0.88 | 0.92 | 0.86 | 1.05 | 1.11 | 0,0.1, 36 | 18 | 0.098 | 0.10 | 0.10 | 0.13 | 0.14 | 0,27, 61 | 3 | 0.0137 | 0.026 | 0.025 | 0.135 | 0.143 | 0,50, 38 | 17 |
| 15 | Const, forest | 0.91 | 0.94 | 0.90 | 1.17 | 1.21 | 0, 3 | 6 | 0.093 | 0.10 | 0.09 | 0.13 | 0.13 | 0, 11 | 4 | 0.0127 | 0.021 | 0.020 | 0.078 | 0.080 | 0, 4 | 51 |
| 16 | Const, S1085, forest | 0.91 | 0.95 | 0.89 | 1.18 | 1.24 | 0, 17, 2 | 7 | 0.088 | 0.09 | 0.09 | 0.12 | 0.13 | 0, 7, 32 | 9 | 0.0127 | 0.024 | 0.023 | 0.065 | 0.069 | 0, 17, 2 | 60 |

[a] Const is a constant term. Refer to text in Section 2 for a full description of the catchment characteristics.

**Table 2b**
Summary of the catchment characteristics and statistical measures used in the forward stepwise regression for the flood quantiles of the LP3 distribution (ARIs = 2, 10 and 100 years) for NSW.

| Combination | Catchment Characteristics[a] | LP3 flood quantiles | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | ARI = 2 | | | | | | | ARI = 10 | | | | | | | ARI = 100 | | | | | | |
| | | $\sigma_\delta^2$ | AVPO | AVPN | AIC | BIC | BPV% | $\overline{R}^2_{GLS}$ (%) | $\sigma_\delta^2$ | AVPO | AVPN | AIC | BIC | BPV% | $\overline{R}^2_{GLS}$ (%) | $\sigma_\delta^2$ | AVPO | AVPN | AIC | BIC | BPV% | $\overline{R}^2_{GLS}$ (%) |
| 1 | Const | 0.94 | 0.96 | 0.94 | 1.26 | 1.26 | 0 | 0 | 0.89 | 0.91 | 0.89 | 1.16 | 1.16 | 0 | 0 | 0.87 | 0.89 | 0.87 | 1.21 | 1.21 | 0 | 0 |
| 2 | Const, area | 0.73 | 0.75 | 0.72 | 0.78 | 0.81 | 0, 0, 0 | 39 | 0.54 | 0.56 | 0.53 | 0.52 | 0.53 | 0, 0, 0 | 56 | 0.52 | 0.54 | 0.52 | 0.64 | 0.66 | 0, 0, 0 | 48 |
| 3 | Const, area, $^{2}I_{1}$ | 0.35 | 0.37 | 0.34 | 0.38 | 0.40 | 0, 0, 0 | 71 | 0.23 | 0.25 | 0.24 | 0.26 | 0.28 | 0, 0, 0 | 78 | 0.35 | 0.38 | 0.36 | 0.42 | 0.45 | 0, 0, 0 | 67 |
| 4 | Const, area, $^{2}I_{12}$ | 0.31 | 0.33 | 0.31 | 0.33 | 0.35 | 0, 0, 0 | 75 | 0.23 | 0.24 | 0.23 | 0.26 | 0.27 | 0, 0, 0 | 78 | 0.35 | 0.37 | 0.35 | 0.36 | 0.38 | 0, 0, 0 | 72 |
| 5 | Const, area, $^{50}I_{1}$ | 0.34 | 0.36 | 0.34 | 0.36 | 0.38 | 0, 0, 0 | 73 | 0.25 | 0.27 | 0.25 | 0.28 | 0.29 | 0, 0, 0 | 77 | 0.35 | 0.38 | 0.36 | 0.42 | 0.44 | 0, 0, 0 | 67 |
| 6 | Const, area, $^{50}I_{12}$ | 0.31 | 0.33 | 0.31 | 0.33 | 0.35 | 0, 0, 0 | 75 | 0.22 | 0.24 | 0.23 | 0.25 | 0.27 | 0, 0, 0 | 79 | 0.35 | 0.38 | 0.36 | 0.41 | 0.43 | 0, 0, 0 | 68 |
| 7 | Const, area, S1085 | 0.74 | 0.77 | 0.73 | 0.80 | 0.85 | 0, 0, 69 | 39 | 0.54 | 0.57 | 0.53 | 0.52 | 0.55 | 0, 0, 34 | 56 | 0.52 | 0.55 | 0.52 | 0.65 | 0.69 | 0, 0, 63 | 48 |
| 8 | Const, area, sden | 0.66 | 0.69 | 0.65 | 0.72 | 0.76 | 0, 0, 0.3 | 45 | 0.46 | 0.49 | 0.46 | 0.55 | 0.58 | 0, 0, 0.2 | 55 | 0.49 | 0.52 | 0.49 | 0.63 | 0.66 | 0, 0, 0.5 | 50 |
| 9 | Const, area, sden, forest | 0.65 | 0.68 | 0.63 | 0.72 | 0.78 | 0, 0, 1, 9 | 46 | 0.48 | 0.51 | 0.47 | 0.56 | 0.61 | 0, 0, 1, 90 | 54 | 0.49 | 0.52 | 0.48 | 0.63 | 0.69 | 0, 0, 1, 20 | 51 |
| 10 | Const, area, $I_{tc,ARI}$ | 0.29 | 0.33 | 0.31 | 0.33 | 0.35 | 0, 0, 0 | 75 | 0.23 | 0.24 | 0.23 | 0.26 | 0.27 | 0, 0, 0 | 79 | 0.35 | 0.38 | 0.36 | 0.44 | 0.46 | 0, 0, 0 | 65 |
| 11 | Const, area, forest | 0.69 | 0.72 | 0.67 | 0.76 | 0.80 | 0, 0, 2 | 42 | 0.54 | 0.57 | 0.54 | 0.51 | 0.54 | 0, 0, 40 | 57 | 0.53 | 0.56 | 0.52 | 0.65 | 0.69 | 0, 0, 59 | 48 |
| 12 | Const, area, evap | 0.61 | 0.64 | 0.60 | 0.65 | 0.69 | 0, 0, 0.2 | 50 | 0.38 | 0.40 | 0.38 | 0.38 | 0.40 | 0, 0, 0 | 69 | 0.45 | 0.48 | 0.45 | 0.59 | 0.63 | 0, 0, 0.4 | 53 |
| 13 | Const, area, rain | 0.34 | 0.36 | 0.34 | 0.36 | 0.38 | 0, 0, 0.2 | 73 | 0.35 | 0.37 | 0.35 | 0.43 | 0.45 | 0, 0, 0 | 64 | 0.40 | 0.43 | 0.41 | 0.50 | 0.53 | 0, 0, 0.1 | 61 |
| 14 | Const, rain, S1085 | 0.90 | 0.94 | 0.88 | 1.06 | 1.12 | 0, 0, 4 | 19 | 0.86 | 0.90 | 0.85 | 1.07 | 1.12 | 0, 6, 1 | 11 | 0.85 | 0.89 | 0.83 | 1.17 | 1.23 | 0, 36, 0.7 | 8 |
| 15 | Const, sden, S1085 | 0.93 | 0.97 | 0.91 | 1.21 | 1.28 | 0, 15, 2 | 8 | 0.88 | 0.91 | 0.86 | 1.10 | 1.16 | 0, 25,0.1 | 9 | 0.85 | 0.89 | 0.84 | 1.16 | 1.22 | 0, 27,0.1 | 8 |
| 16 | Const, area, $^{50}I_{12}$, S1085 | 0.37 | 0.39 | 0.36 | 0.23 | 0.25 | 0, 0, 0, 40 | 83 | 0.22 | 0.24 | 0.22 | 0.26 | 0.28 | 0, 0, 0, 35 | 79 | 0.35 | 0.38 | 0.35 | 0.42 | 0.46 | 0, 0, 0, 62 | 67 |
| 17 | Const, area, $^{50}I_{12}$, rain | 0.29 | 0.31 | 0.29 | 0.32 | 0.35 | 0, 0, 0, 0.4 | 76 | 0.23 | 0.25 | 0.23 | 0.26 | 0.28 | 0, 0, 0, 22 | 79 | 0.35 | 0.38 | 0.35 | 0.42 | 0.46 | 0, 0, 0, 28 | 67 |
| 18 | Const, area, $^{50}I_{12}$, S1085, forest | 0.37 | 0.39 | 0.36 | 0.25 | 0.28 | 0, 0, 0, 48, 79 | 72 | 0.21 | 0.24 | 0.22 | 0.25 | 0.28 | 0, 0, 0, 55, 75 | 80 | 0.35 | 0.38 | 0.35 | 0.33 | 0.37 | 0, 0, 0, 55, 79 | 75 |
| 19 | Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, forest | 0.37 | 0.39 | 0.35 | 0.22 | 0.25 | 0, 0,15, 16,70 | 74 | 0.21 | 0.24 | 0.21 | 0.25 | 0.28 | 0, 0,22, 43,70 | 80 | 0.34 | 0.38 | 0.35 | 0.33 | 0.37 | 0, 0,10, 80,90 | 75 |
| 20 | Const, area, $^{50}I_{12}$, $I_{tc,ARI}$, S1085, forest | 0.37 | 0.40 | 0.35 | 0.24 | 0.28 | 0, 0, 15, 18, 70,78 | 73 | 0.22 | 0.24 | 0.22 | 0.26 | 0.30 | 0, 0, 23, 44, 95,90 | 80 | 0.35 | 0.39 | 0.35 | 0.36 | 0.42 | 0, 0, 27, 90, 95,90 | 73 |
| 21 | Const, area, $I_{tc,ARI}$, rain | 0.30 | 0.32 | 0.29 | 0.32 | 0.35 | 0, 0, 0, 2 | 76 | 0.23 | 0.25 | 0.23 | 0.26 | 0.29 | 0, 0, 0, 76 | 78 | 0.35 | 0.38 | 0.35 | 0.44 | 0.48 | 0, 0, 0, 81 | 66 |
| 22 | Const, area, $I_{tc,ARI}$, evap | 0.32 | 0.34 | 0.31 | 0.34 | 0.37 | 0, 0, 0, 86 | 74 | 0.23 | 0.25 | 0.23 | 0.26 | 0.29 | 0, 0, 0, 80 | 79 | 0.35 | 0.39 | 0.36 | 0.45 | 0.49 | 0, 0, 0, 95 | 65 |
| 23 | Const, area, $I_{tc,ARI}$, forest | 0.37 | 0.39 | 0.36 | 0.23 | 0.25 | 0, 0, 0, 98 | 73 | 0.22 | 0.24 | 0.22 | 0.25 | 0.27 | 0, 0, 0, 8 | 79 | 0.35 | 0.38 | 0.35 | 0.40 | 0.43 | 0, 0, 0, 98 | 69 |
| 24 | Const, area, $I_{tc,ARI}$, S1085 | 0.37 | 0.39 | 0.36 | 0.23 | 0.25 | 0, 0, 0, 92 | 73 | 0.23 | 0.25 | 0.23 | 0.26 | 0.29 | 0, 0, 0, 50 | 79 | 0.35 | 0.38 | 0.35 | 0.45 | 0.49 | 0, 0, 0, 95 | 65 |
| 25 | Const, area, $^{2}I_{1}$, $I_{tc,ARI}$ | 0.32 | 0.34 | 0.31 | 0.35 | 0.38 | 0, 0, 46, 0 | 74 | 0.23 | 0.25 | 0.23 | 0.26 | 0.28 | 0, 0, 59, 1 | 79 | 0.35 | 0.38 | 0.35 | 0.43 | 0.47 | 0, 0, 49, 0 | 67 |

[a] Const is a constant term. Refer to text in Section 2 for a full description of the catchment characteristics.

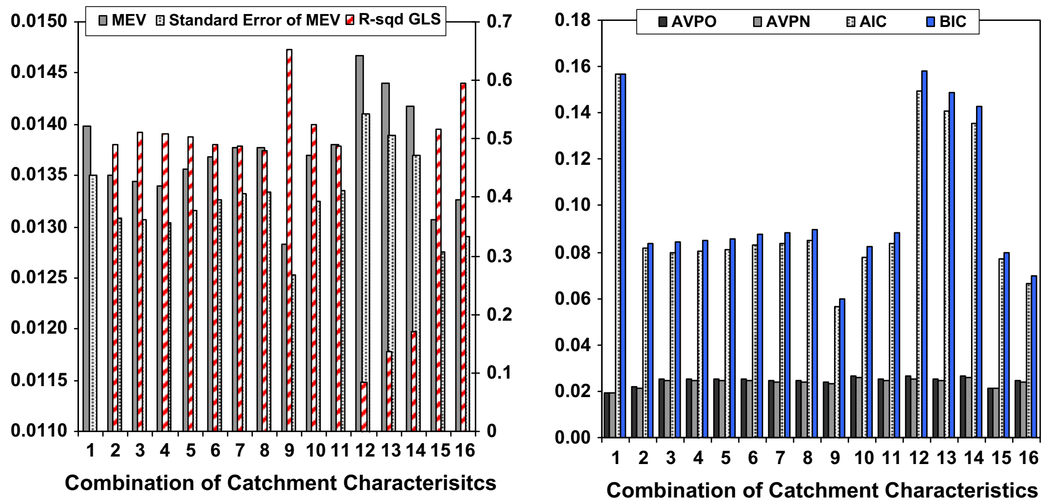


**Fig. 2.** Selection of explanatory variables for the BGLS regression model for the skew (note that $\overline{R}^2_{GLS}$ uses the right-hand axis).

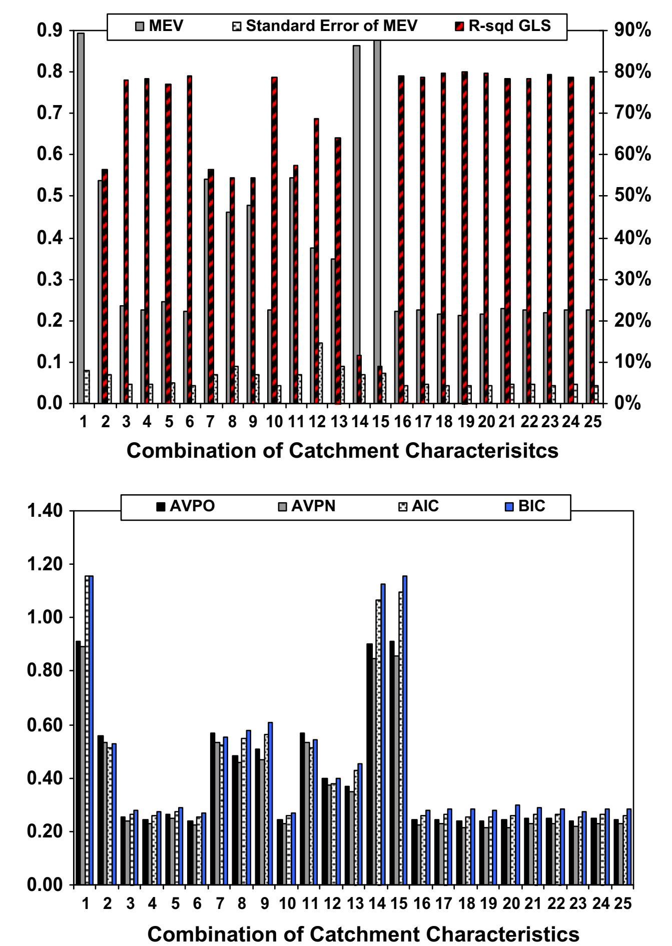


**Fig. 3.** Selection of predictor variables for the BGLS regression model for $Q_{10}$ model (note that $\overline{R}^2_{GLS}$ uses the right-hand axis), (QRT, fixed region NSW), MEV = model error variance, AVPO = average variance of prediction (old), AVPN = average variance of prediction (new) AIC = Akaike information criterion, BIC = Bayesian information criterion.

for NSW the EVR is 0.30 for the ROI and 0.17 for the fixed region (see Table 3), Table 3 also provides the EVR results for VIC and QLD which show a similar outcome to NSW. For the standard deviation model for NSW the EVR is 0.77 for the ROI and 0.35 for the fixed region, again similar results were found for VIC and QLD.

The EVR values for the skew models of NSW, VIC and QLD are shown in Table 4. It can be observed from Table 4 that the EVR values range from 8 to 19 and 9.5 to 19 for the fixed regions and ROI respectively (Table 4), which are much higher than the recommended limit of 0.20. Two important points are noted below:

(i) This clearly indicates that the GLS regression is the preferred modeling choice over the OLS one for the skew model. An OLS model for the skew would have clearly given misleading results as it does not distinguish between the model and sampling errors as found in similar previous studies (e.g. Haddad et al., 2010a, 2010b).

(ii) Importantly what is clear is that if a method of moment's estimator was used to estimate the model error variance $\sigma^2_\delta$ for the skew model, the model error variance would have been grossly underestimated as the sampling error has heavily dominated the regional analysis (for example see Haddad et al., 2010a, 2010b). A more reasonable estimate of the model error variance has been achieved with the Bayesian procedure as it represents the values of $\sigma^2_\delta$ by computing expectations over the entire posterior distribution. Similar results have been found by Reis et al. (2005) and Gruber and Stedinger (2008). As far as the ROI is concerned there is little change in the EVR as compared to the fixed region for all the states as the skew model tends to include more stations in the regional analysis.

The pseudo ANOVA tables were also prepared for all the flood quantile models. The results for the $Q_{20}$ for all the three states are shown in Table 5. Here the ROI shows a higher EVR than the fixed region and that the sampling error generally increases with increasing ARIs. The reduction in the model error variance as seen in Table 5 for all the states is due to the fact that ROI has found an optimum number of sites based on the minimum model error variance which naturally uses fewer sites than that of the fixed region approach. This indeed suggests that sub regions may exist in larger regions.

The flood quantile $Q_2$ was found to experience the lowest EVR for NSW and QLD for both the fixed region and ROI as compared to the $Q_{20}$ and $Q_{100}$ models results. This reflects the much greater spatial variability of the mean which is dominated by local factors

**Table 3**
Pseudo ANOVA table for the mean flood model (PRT, fixed region and ROI, NSW, VIC and QLD) (here $n$ = number of sites in the region, $k$ = number of predictors in the regression equation, EVR = error variance ratio, $\sigma_{\delta 0}^2$ = model error variance when no explanatory variable is used in the regression model, $\sigma_\delta^2$ = model error variance when explanatory variable is used in the regression model and $tr[\sum(\hat{y})]$ = sum of the diagonals of the sampling covariance matrix).

| Source | Degrees of freedom | | Sum of squares | | |
|---|---|---|---|---|---|
| | Fixed region | ROI | | Fixed region | ROI |
| *NSW* | | | | | |
| Model | $k = 3$ | $k = 3$ | $n(\sigma_{\delta 0}^2 - \sigma_\delta^2)$ | 61.5 | 61.2 |
| Model error $\delta$ | $n - k - 1 = 92$ | $n - k - 1 = 32$ | $n(\sigma_\delta^2$ | 27.7 | 16.5 |
| Sampling error | $n = 96$ | $n = 36$ | $tr[\sum(\hat{y})]$ | 5 | 4.5 |
| Total | $2n - 1 = 191$ | $2n - 1 = 71$ | Sum of the above | 94 | 83 |
| | | | EVR | 0.17 | 0.3 |
| *VIC* | | | | | |
| Model | $k = 3$ | $k = 3$ | | 46 | 45 |
| Model error $\delta$ | $n - k - 1 = 127$ | $n - k - 1 = 39$ | | 37.5 | 28 |
| Sampling error $\eta$ | $n = 131$ | $n = 43$ | | 6.1 | 6 |
| Total | $2n - 1 = 261$ | $2n - 1 = 85$ | Sum of the above | 90 | 79 |
| | | | EVR | 0.16 | 0.2 |
| *QLD* | | | | | |
| Model | $k = 3$ | $k = 3$ | | 105 | 102 |
| Model error $\delta$ | $n - k - 1 = 168$ | $n - k - 1 = 34$ | | 39 | 22 |
| Sampling error $\eta$ | $n = 172$ | $n = 38$ | | 10.2 | 9 |
| Total | $2n - 1 = 343$ | $2n - 1 = 75$ | Sum of the above | 155 | 133 |
| | | | EVR | 0.26 | 0.40 |

**Table 4**
Pseudo ANOVA table for the skew model (PRT, fixed region and ROI, NSW, VIC and QLD) (variables are explained in Table 3 caption).

| Source | Degrees of freedom | | Sum of squares | | |
|---|---|---|---|---|---|
| | Fixed region | ROI | | Fixed region | ROI |
| *NSW* | | | | | |
| Model | $k = 3$ | $k = 3$ | $n(\sigma_{\delta 0}^2 - \sigma_\delta^2)$ | 0.1 | 0.1 |
| Model error δ | $n - k - 1 = 92$ | $n - k - 1 = 91$ | $n(\sigma_\delta^2$ | 1.22 | 1.21 |
| Sampling error $\eta$ | $n = 96$ | $n = 95$ | $tr[\sum(\hat{y})]$ | 24 | 23 |
| Total | $2n - 1 = 191$ | $2n - 1 = 189$ | Sum of the above | 25 | 23 |
| | | | EVR | 19 | 18 |
| *VIC* | | | | | |
| Model | $k = 3$ | $k = 3$ | | 6.5 | 7.3 |
| Model error δ | $n - k - 1 = 127$ | $n - k - 1 = 113$ | | 4.5 | 3.7 |
| Sampling error $\eta$ | $n = 131$ | $n = 117$ | | 38 | 35 |
| Total | $2n - 1 = 261$ | $2n - 1 = 233$ | Sum of the above | 49 | 48 |
| | | | EVR | 8.4 | 9.5 |
| *QLD* | | | | | |
| Model | $k = 3$ | $k = 3$ | | 0.11 | 0.65 |
| Model error δ | $n - k - 1 = 168$ | $n - k - 1 = 146$ | | 2.6 | 2.1 |
| Sampling error $\eta$ | $n = 172$ | $n = 150$ | | 45 | 40 |
| Total | $2n - 1 = 343$ | $2n - 1 = 299$ | Sum of the above | 48 | 43 |
| | | | EVR | 17 | 19 |

(as compared to the higher moments). This is reflected in the $Q_2$ flood as it is mostly dominated by the mean flood.

The $Q_{20}$ shows an EVR of 0.43, 0.3 and 0.97 respectively for NSW, VIC and QLD (see Table 5) for ROI which suggests that the BGLS combined with ROI should be the preferred procedure when modelling the larger ARI quantiles, even though in this particular case the ROI has been impacted by the relatively large model error variances that have dominated the regional flood quantile modelling results.

### 4.2.2. Regression diagnostics – model adequacy and outlier analysis

To assess the underlying model assumptions (i.e. the normality of residuals), the plots of the standardised residuals (Eq. (13)) vs. fitted quantiles were examined for all the flood quantiles (estimated from QRT and PRT) and the parameters of the LP3 distribution for all the states. The predicted values were obtained from the one-at-a-time cross validation procedure. Fig. 4 shows the plot for the $Q_{20}$ model for NSW.

If the underlying model assumption is satisfied to a large extent the standardised residual values should not exceed the ±2 limits; in practice, 95% of the standardised residuals should fall between ±2. The result in Fig. 4 reveals that the developed flood quantiles from the prediction equations satisfy the normality of residual assumption quite satisfactorily for both the fixed and ROI methods. Also no specific pattern (heteroscedasicity) can be identified with the standardised values being almost equally distributed below and above zero. What is noteworthy is that ROI is clearly providing less genuine outliers for both the quantiles estimated by QRT and PRT than the fixed region approach demonstrating its superiority to a fixed region regression. Overall similar results were observed for the states of VIC and QLD.

The QQ-plots of the standardised residuals (Eq. (13)) vs. normal score (Eq. (14)) for the fixed region (based on one-at-a-time cross validation) and ROI were examined. The results for the $Q_{20}$ model for NSW are shown in Fig. 5, which reveals that all the points closely follow a straight line; this is especially the case for the ROI approach for both the QRT and PRT methods. This indicates that the assumption of normality and the homogeneity of variance of the standardised residuals are better approximated with the ROI approach. Overall, no genuine outliers can be

**Table 5**
Pseudo ANOVA table for $Q_{20}$ model (QRT, fixed region and ROI, NSW, VIC and QLD) (variables are explained in Table 3 caption).

| Source | Degrees of freedom | | Sum of squares | | |
|---|---|---|---|---|---|
| | Fixed region | ROI | | Fixed region | ROI |
| *NSW* | | | | | |
| Model | $k = 3$ | $k = 3$ | $n(\sigma_{\delta 0}^2 - \sigma_\delta^2)$ | 61.1 | 61.1 |
| Model error δ | $n - k - 1 = 92$ | $n - k - 1 = 48$ | $n(\sigma_\delta^2$ | 23.5 | 17.3 |
| Sampling error $\eta$ | $n = 96$ | $n = 52$ | $tr[\sum(\hat{y})]$ | 7.6 | 7.0 |
| Total | $2n - 1 = 191$ | $2n - 1 = 103$ | Sum of the above | 92 | 86 |
| | | | EVR | 0.32 | 0.43 |
| *VIC* | | | | | |
| Model | $k = 3$ | $k = 3$ | | 45.2 | 45.2 |
| Model error δ | $n - k - 1 = 127$ | $n - k - 1 = 48$ | | 55.2 | 24.4 |
| Sampling error $\eta$ | $n = 131$ | $n = 52$ | | 7.4 | 7.2 |
| Total | $2n - 1 = 261$ | $2n - 1 = 103$ | Sum of the above | 108 | 77 |
| | | | EVR | 0.13 | 0.30 |
| *QLD* | | | | | |
| Model | $k = 3$ | $k = 3$ | | 59 | 46 |
| Model error δ | $n - k - 1 = 168$ | $n - k - 1 = 77$ | | 25 | 12 |
| Sampling error $\eta$ | $n = 172$ | $n = 81$ | | 13 | 12 |
| Total | $2n - 1 = 343$ | $2n - 1 = 161$ | Sum of the above | 97 | 70 |
| | | | EVR | 0.53 | 0.97 |

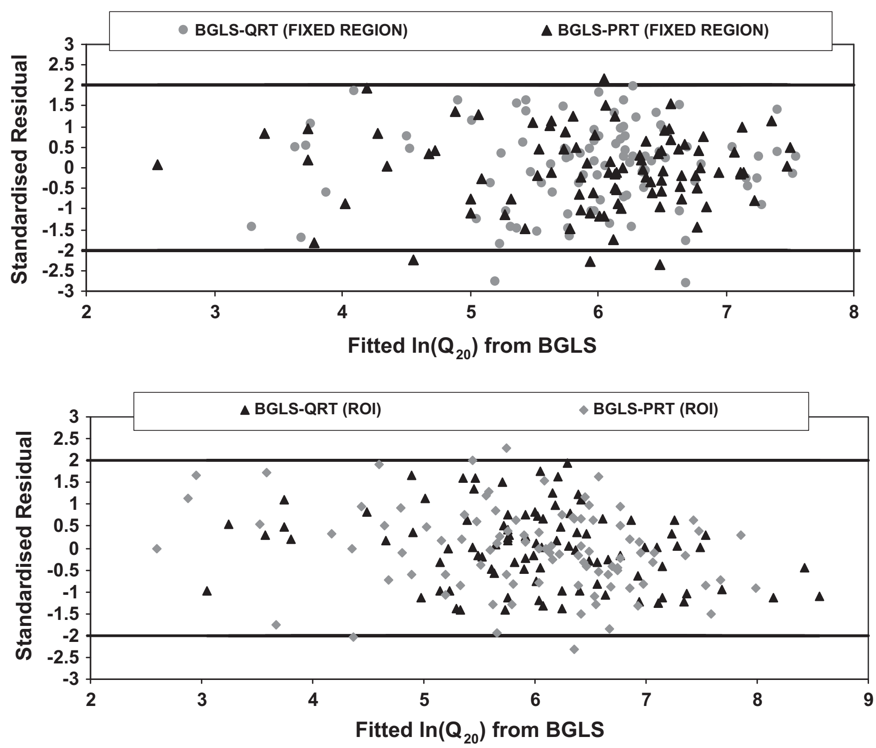


**Fig. 4.** Plots of the standardised residuals vs. predicted values for ARI of 20 years (QRT and PRT, fixed region and ROI, NSW).

detected for the flood quantiles estimated by the QRT and PRT on a regional scale.

If the standardised residuals are indeed normally and independently distributed $N(0,1)$ with mean 0 and variance 1 then the slope of the best fit line in the *QQ*-plot, which can be interpreted as the standard deviation of the normal score (*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. Fig. 5 indeed shows that the fitted lines for the developed models pass approximately through the origin (0, 0) and have a slope approximately equal to one. It can be seen that the results of the ROI approach satisfy the model assumptions relatively better than the fixed region approach. The superiority of the ROI approach again here is demonstrated. Similar results were observed for VIC and QLD. The assumption of the normality of the residuals for all the states (NSW, VIC and QLD) could not be rejected at the 10% level of significance using the Anderson–Darling and Kolmogorov–Smirnov tests for normality.

Below we present the residual analysis results of the ROI method for the Parameter Regression Technique (PRT) using a weighted regional average SD and skew which is weighted by the error covariance matrix (i.e. no predictor variables in the regression equation) for the state of NSW only. The main aspect of this analysis is to determine if there is any reasonable loss in accuracy and efficiency especially in the estimation of the mid to higher ARIs (i.e. 20–100 years) when using a weighted regional average standard deviation and skew (obtained as above) as compared to ones with explanatory variables. It should be stressed here that this

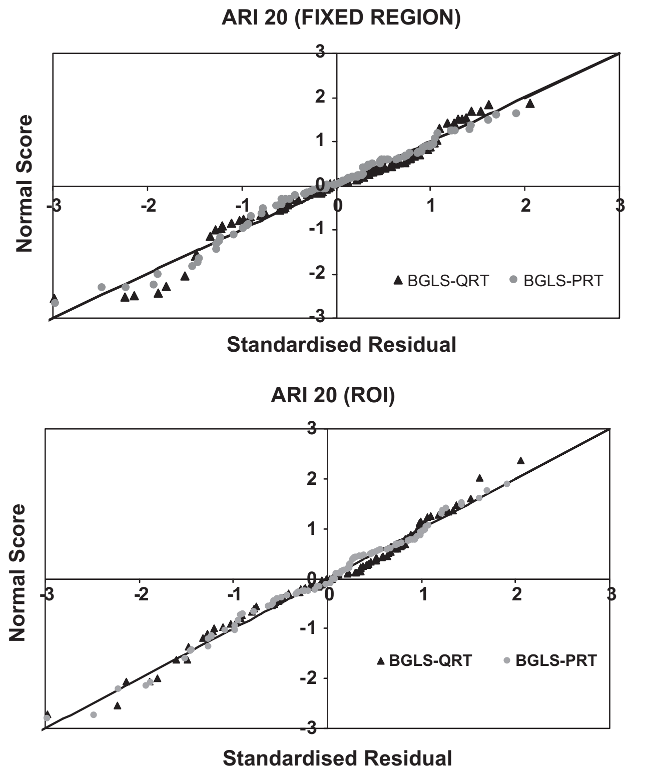


**Fig. 5.** *QQ*-plot of the standardised residuals vs. Z score for ARI of 20 years (QRT and PRT, fixed region, ROI, NSW).

weighted regional average SD and skew do vary from site to site as each site has a unique ROI.

We present the standardised residuals vs. the fitted quantile plot of $Q_{20}$ in Fig. 6a that superimposes the estimate made by the QRT-ROI, PRT-ROI and the PRT-ROI that uses a weighted regional average standard deviation and skew estimate. Indeed one can observe that the PRT-ROI estimate of $Q_{20}$ with the weighted regional average standard deviation and skew performs equally well as the competing models. Nearly all the standardised residuals fall within the ±2 limits, suggesting that the use of explanatory variables does not really add any more meaningful information to the analysis. The *QQ*-plot (Fig. 6b) of the competing models shows that the use of a weighted regional average standard deviation and skew does not result in any major gross errors. The residual analysis also reveals that the major assumptions of the regression have been largely satisfied (i.e. normality of the residuals). The results based on the evaluation statistics are given in the evaluation statistics section of the results (Section 4.3).

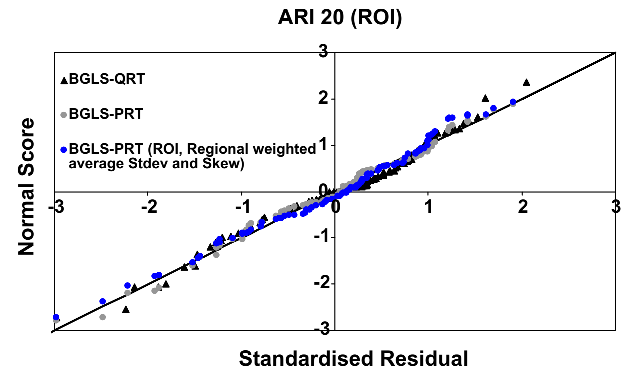


**Fig. 6b.** *QQ*-plot of the standardised residuals vs. *Z* score for ARI of 20 years (QRT and PRT, ROI, and PRT ROI with weighted average standard deviation and skew, NSW).

### 4.2.3. Diagnostic statistics

The summary of the various regression diagnostics (as described in Section 3.6 and Eq. (12) is provided in Table 6 for NSW, VIC and QLD. This shows that for the mean flood model (for all the states), the model error variance (MEV) and average standard error of prediction (SEP) are much higher than those of the standard deviation and skew models. This indicates that the mean flood model exhibits a higher degree of heterogeneity than the standard deviation and skew models. This result supports the Pseudo ANOVA results. Indeed the issue here is that sampling error becomes larger as the order of the moment increases, therefore in case of the skew the spatial variation is a second order effect that it not really detectable, this is apparent in both the fixed region and ROI analysis cases.

For the mean flood model (all the three states), the ROI shows a MEV which is smaller than the fixed region analysis for NSW, VIC and QLD respectively. The lower MEV inturn also provides the lower AVP values as can be seen in Table 6. Also, the $\overline{R}^2_{GLS}$ value for the mean flood model (all the three states) with the ROI is 8%, 1% and 1% higher than the fixed region analysis for NSW, VIC and QLD respectively. These results indicate that the ROI should be preferred over the fixed region for developing the mean flood model.

For the standard deviation model, ROI shows 2% smaller and 9% higher SEP and $\overline{R}^2_{GLS}$ values, respectively for NSW. The best result is found for QLD, here ROI shows a 14% smaller and 12% higher SEP and $\overline{R}^2_{GLS}$ values, respectively. This indicates that the ROI is preferable than the fixed region for the standard deviation model. The SEP and $\overline{R}^2_{GLS}$ values for the skew model are the same for the fixed region and ROI for NSW and QLD respectively (see Table 6). This can be explained by the fact that the number of sites for the skew

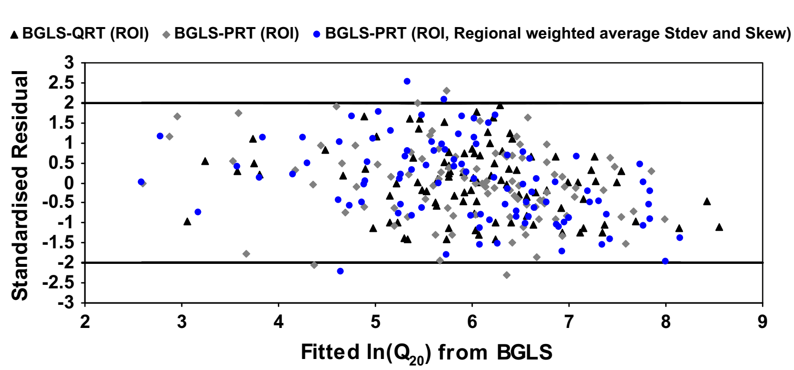


**Fig. 6a.** Plots of the standardised residuals vs. predicted values for ARI of 20 years (QRT and PRT, ROI and PRT-ROI with weighted average standard deviation and skew, NSW).

**Table 6**
Regression diagnostics for the fixed region and ROI for NSW, VIC and QLD.

| Model | Fixed region | | | | ROI | | | |
|---|---|---|---|---|---|---|---|---|
| | MEV | AVP | SEP (%) | $\overline{R}^2_{GLS}$ (%) | MEV | AVP | SEP (%) | $\overline{R}^2_{GLS}$ (%) |
| *NSW* | | | | | | | | |
| Mean | 0.29 | 0.31 | 60 | 76 | 0.19 | 0.23 | 51 | 84 |
| St. dev | 0.058 | 0.062 | 25 | 37 | 0.046 | 0.054 | 23 | 46 |
| Skew | 0.013 | 0.024 | 16 | 65 | 0.013 | 0.023 | 16 | 65 |
| $Q_2$ | 0.31 | 0.33 | 63 | 77 | 0.20 | 0.24 | 52 | 84 |
| $Q_5$ | 0.23 | 0.24 | 52 | 79 | 0.16 | 0.20 | 47 | 85 |
| $Q_{10}$ | 0.23 | 0.24 | 52 | 79 | 0.16 | 0.20 | 46 | 85 |
| $Q_{20}$ | 0.25 | 0.27 | 55 | 76 | 0.18 | 0.22 | 49 | 83 |
| $Q_{50}$ | 0.35 | 0.37 | 66 | 70 | 0.25 | 0.28 | 56 | 74 |
| $Q_{100}$ | 0.35 | 0.38 | 68 | 65 | 0.29 | 0.34 | 63 | 70 |
| *VIC* | | | | | | | | |
| Mean | 0.29 | 0.31 | 60 | 62 | 0.21 | 0.23 | 46 | 63 |
| St. dev | 0.044 | 0.049 | 22 | 65 | 0.041 | 0.050 | 21 | 65 |
| Skew | 0.034 | 0.040 | 20 | 70 | 0.028 | 0.037 | 19 | 73 |
| $Q_2$ | 0.27 | 0.28 | 57 | 63 | 0.20 | 0.23 | 51 | 65 |
| $Q_5$ | 0.29 | 0.31 | 60 | 61 | 0.20 | 0.23 | 50 | 64 |
| $Q_{10}$ | 0.35 | 0.37 | 67 | 57 | 0.23 | 0.26 | 54 | 61 |
| $Q_{20}$ | 0.35 | 0.37 | 67 | 57 | 0.19 | 0.22 | 48 | 66 |
| $Q_{50}$ | 0.47 | 0.49 | 80 | 49 | 0.27 | 0.32 | 61 | 61 |
| $Q_{100}$ | 0.59 | 0.60 | 91 | 45 | 0.29 | 0.35 | 64 | 54 |
| *QLD* | | | | | | | | |
| Mean | 0.23 | 0.24 | 52 | 77 | 0.14 | 0.15 | 40 | 78 |
| St. dev | 0.13 | 0.14 | 38 | 34 | 0.056 | 0.061 | 24 | 46 |
| Skew | 0.015 | 0.024 | 16 | 44 | 0.014 | 0.026 | 16 | 44 |
| $Q_2$ | 0.26 | 0.27 | 56 | 75 | 0.15 | 0.18 | 43 | 79 |
| $Q_5$ | 0.17 | 0.18 | 44 | 79 | 0.08 | 0.11 | 34 | 83 |
| $Q_{10}$ | 0.18 | 0.19 | 45 | 74 | 0.07 | 0.11 | 33 | 79 |
| $Q_{20}$ | 0.15 | 0.16 | 41 | 77 | 0.07 | 0.13 | 36 | 80 |
| $Q_{50}$ | 0.17 | 0.19 | 45 | 72 | 0.10 | 0.14 | 39 | 77 |
| $Q_{100}$ | 0.20 | 0.22 | 49 | 72 | 0.12 | 0.16 | 40 | 73 |

model in the ROI approach was very close to the fixed region approach.

Interestingly one can see from Table 6 that the SEP values for all the flood quanitle models for NSW, VIC and QLD respectively are 5–11%, 6–27% and 5–13% smaller for the ROI cases than the fixed region analysis. Also, the $\overline{R}^2_{GLS}$ values for ROI cases for NSW, VIC and QLD respectively are 4–7%, 2–12% and 1–5% higher than the fixed region analysis. These results show the relative advantage of the ROI method coupled with BGLS regression over a fixed region BGLS regression analysis where further improvements have been achieved overall.

Table 7 shows the number of sites in a region, the associated model error variances and their % differences for the ROI against the fixed region models for NSW, VIC and QLD. This shows that the ROI mean flood model for all states has fewer sites on average (36 out of 96 i.e. 37% of the available sites for NSW, 32% for VIC and 24% for QLD) than the standard deviation and skew models. The ROI skew model for each state has the highest number of sites which includes nearly all the sites in the respective states. The model error variances for all the flood quantile ROI models are smaller than the fixed region models with differences in order of 50–60%. This shows that the fixed region models experience a greater heterogeneity than the ROI. If the fixed region models are made too big, the model error will be inflated by heterogeneity unaccounted for by the catchment characteristics. Two important notes can be made here is that spatial proximity (physical distance) may become a surrogate for unknown processes in RFFA and that catchment and climatic variables available at the regional scale may not always be good indicators of regional flood behaviour.

Fig. 7 plots the spatial variation of the minimum model error variances (grouped in classes according to numerical values as specified in the legend) for the mean flood model (Fig. 7a) and how the model error variance varies with the number of sites within the ROI, for a typical site (Fig. 7b) for the state of NSW. The plot reveals the relative advantage of the ROI approach. It can be seen that there are distinct spatial variations illustrating the heterogeneity of the mean flood that would be often ignored in a fixed region approach. Similar results were observed in both VIC and QLD.

The spatial variation in the model error for the skew model captures the entire study area mostly (figure not shown) for NSW, VIC and QLD. Similar results were found by Hackelbusch et al. (2009). The significance of this finding is that if any spatial variations exist in the hydrologic statistic of interest, they are most likely to be captured by the ROI.

### 4.3. Evaluation statistics

An objective assessment of the model estimation methods can be obtained by using the numerical evaluation statistics given is Eqs. (15) and (16), in which $RMSE_r$ is the relative root mean squared error and $RE_r$ is the absolute median relative error. The $RMSE_r$ is associated with the predictive error variance, where as $RE_r$ is related mostly with prediction bias. Using the model predicted flood quantiles (estimated by QRT and PRT, fixed and ROI) using the one-at-a-time cross validation, the evaluation statistics were calculated and are given in Table 8.

**Table 7**
Model error variances $\hat{\sigma}^2_\delta$ associated with the fixed region and ROI for NSW, VIC and QLD ($n$ = number of sites needed for the LP3 parameters and flood quantiles).

| State | Parameter/ARI | Mean | St. dev | Skew | $Q_2$ | $Q_5$ | $Q_{10}$ | $Q_{20}$ | $Q_{50}$ | $Q_{100}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| NSW | ROI($n$)/ | 36 | 47 | 95 | 31 | 42 | 48 | 52 | 53 | 55 |
| | $\hat{\sigma}^2_\delta$ | 0.19 | 0.046 | 0.013 | 0.20 | 0.16 | 0.16 | 0.18 | 0.25 | 0.29 |
| | Fixed region ($n$)/ | 96 | 96 | 96 | 96 | 96 | 96 | 96 | 96 | 96 |
| | $\hat{\sigma}^2_\delta$ | 0.29 | 0.058 | 0.013 | 0.21 | 0.23 | 0.23 | 0.25 | 0.35 | 0.35 |
| | (%) Diff in $\hat{\sigma}^2_\delta$ | 34% | 21% | 0% | 5% | 30% | 30% | 28% | 29% | 17% |
| VIC | ROI ($n$)/ | 43 | 83 | 117 | 41 | 45 | 52 | 52 | 57 | 57 |
| | $\hat{\sigma}^2_\delta$ | 0.21 | 0.041 | 0.028 | 0.20 | 0.20 | 0.23 | 0.19 | 0.27 | 0.29 |
| | Fixed region ($n$)/ | 131 | 131 | 131 | 131 | 131 | 131 | 131 | 131 | 131 |
| | $\hat{\sigma}^2_\delta$ | 0.29 | 0.044 | 0.034 | 0.27 | 0.29 | 0.35 | 0.35 | 0.47 | 0.59 |
| | (%) Diff in $\hat{\sigma}^2_\delta$ | 28% | 7% | 18% | 26% | 31% | 34% | 46% | 43% | 51% |
| QLD | ROI ($n$)/ | 42 | 65 | 150 | 60 | 65 | 74 | 80 | 88 | 90 |
| | $\hat{\sigma}^2_\delta$ | 0.15 | 0.056 | 0.014 | 0.14 | 0.08 | 0.07 | 0.07 | 0.10 | 0.12 |
| | Fixed region ($n$)/ | 172 | 172 | 172 | 172 | 172 | 172 | 172 | 172 | 172 |
| | $\hat{\sigma}^2_\delta$ | 0.23 | 0.14 | 0.015 | 0.26 | 0.17 | 0.18 | 0.15 | 0.17 | 0.20 |
| | (%) Diff in $\hat{\sigma}^2_\delta$ | 35% | 60% | 7% | 46% | 53% | 61% | 53% | 41% | 40% |

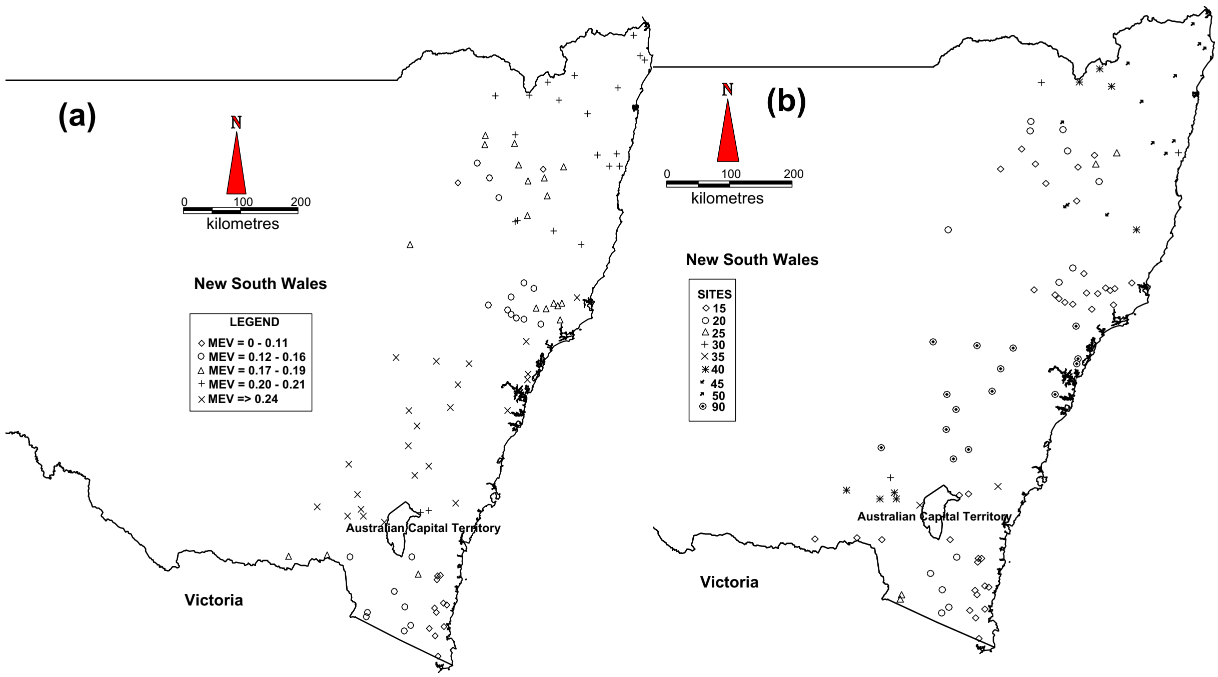


**Fig. 7.** Spatial variations of the grouped minimum model error variances for (a) mean flood model and (b) number of sites which produced the lowest predictive variance for the mean flood model.

Numerical values of these statistics show the relative advantage of the ROI method (for both the QRT and PRT) for all the three states (i.e. NSW, VIC and QLD). The flood quantile estimates obtained from the fixed models (QRT and PRT) are more biased (i.e. higher $RE_r$) and are of a lesser accuracy (i.e. higher $RMSE_r$), this is observed for all the three states examined.

Among QRT and PRT fixed quantile estimation methods (over all the ARIs) it can be observed from Table 8 that there is not much difference in accuracy ($RMSE_r$) for NSW, VIC and QLD. Indeed, in relation to bias ($RE_r$) both QRT and PRT fixed models were found to be very similar for the three states.

For QRT and PRT-ROI quantile estimation methods (over all the ARIs) a similar result was found where there was no notable difference in accuracy ($RMSE_r$) between the competing models. For the bias ($RE_r$) both QRT and PRT ROI models achieved very similar values as seen in Table 8. While Table 8 does show overall slightly better accuracy and bias for QRT over PRT, a point needs to be bought out to clarify this result.

There is some underlying bias involved with the validation of the QRT (fixed and ROI) in that the predicted quantiles are being compared to the quantiles used in the regression analysis. Thus the result mostly seems to be slightly in favour of the QRT (see Table 8). How to compensate for this bias in the validation process needs further effort, which has not been done in this paper. On the other hand the validation procedure for the PRT is more stringent in that the parameters of the distribution are used in the regression and quantiles are then independently estimated and compared to the at-site flood quantiles. The results from the evaluation statistics therefore indicate that the PRT is indeed a viable approach for RFFA as an alternative to the commonly applied QRT method in an ungauged catchment situation.

We now present the results based on the evaluation statistics (i.e. Eqs. (15) and (16)) to compare the flood quantiles from PRT-ROI using a weighted regional average standard deviation and skew to the PRT-ROI using a standard deviation and skew as a function of predictor variables for NSW state only. The evaluation statistics (see Table 8 – values in bracket) from the validation reveal that there is no real loss of accuracy (as compared to at-site flood quantiles) if a weighted regional average standard deviation and skew model is adopted to estimate the flood quantiles up to the 20 year ARI.

The results at the higher ARIs (50 and 100 years) show that using a weighted regional average standard deviation and skew may slightly affect the outcome of the analysis (i.e. lesser accuracy and greater bias). The larger ARI estimation may require further information which may be provided by having explanatory variables (such as catchment area, design rainfall intensity, forest

**Table 8**
Evaluation statistics ($RMSE_r$ and $RE_r$) from one-at-a-time cross validation for NSW (results NSW for PRT using the weighted regional average standard deviation and skew models, i.e. no predictor variables given in *brackets*), VIC and QLD NSW.

| Model | $RMSE_r$ (%) | | | | $RE_r$ (%) | | | |
|---|---|---|---|---|---|---|---|---|
| | PRT | | QRT | | PRT | | QRT | |
| | Fixed region | ROI | Fixed region | ROI | Fixed region | ROI | Fixed region | ROI |
| *NSW* | | | | | | | | |
| $Q_2$ | 73 | 62 (*63*) | 68 | 59 | 46 | 38 (*37*) | 44 | 40 |
| $Q_5$ | 65 | 54 (*59*) | 70 | 59 | 37 | 30 (*32*) | 38 | 36 |
| $Q_{10}$ | 67 | 56 (*60*) | 74 | 55 | 37 | 29 (*33*) | 37 | 36 |
| $Q_{20}$ | 72 | 57 (*63*) | 83 | 53 | 36 | 34 (*34*) | 35 | 31 |
| $Q_{50}$ | 81 | 70 (*77*) | 100 | 67 | 38 | 34 (*35*) | 36 | 32 |
| $Q_{100}$ | 90 | 75 (*85*) | 100 | 72 | 40 | 36 (*39*) | 38 | 35 |
| *VIC* | | | | | | | | |
| $Q_2$ | 56 | 55 | 77 | 68 | 38 | 37 | 37 | 37 |
| $Q_5$ | 69 | 68 | 87 | 68 | 38 | 36 | 35 | 35 |
| $Q_{10}$ | 82 | 80 | 107 | 69 | 37 | 37 | 36 | 35 |
| $Q_{20}$ | 96 | 92 | 112 | 74 | 41 | 40 | 38 | 33 |
| $Q_{50}$ | 115 | 110 | 113 | 95 | 41 | 40 | 41 | 40 |
| $Q_{100}$ | 130 | 127 | 140 | 120 | 46 | 45 | 44 | 44 |
| *QLD* | | | | | | | | |
| $Q_2$ | 82 | 69 | 61 | 56 | 39 | 35 | 39 | 39 |
| $Q_5$ | 68 | 60 | 48 | 44 | 33 | 34 | 34 | 32 |
| $Q_{10}$ | 69 | 60 | 52 | 47 | 34 | 30 | 32 | 31 |
| $Q_{20}$ | 72 | 65 | 50 | 44 | 35 | 33 | 31 | 29 |
| $Q_{50}$ | 78 | 68 | 53 | 49 | 37 | 36 | 32 | 31 |
| $Q_{100}$ | 85 | 79 | 58 | 53 | 41 | 40 | 36 | 31 |

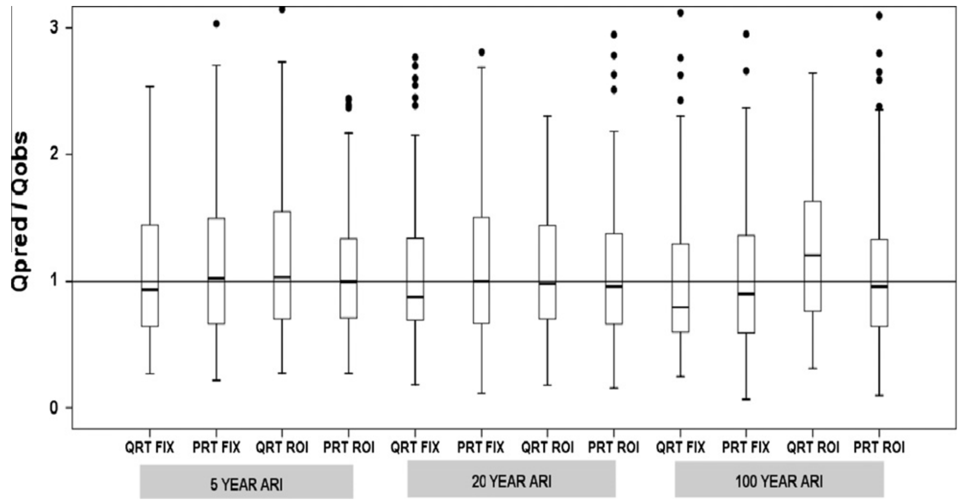


**Fig. 8.** Boxplots of $Q_{pred}/Q_{obs}$ for NSW for QRT and PRT, fixed regions and ROI.

and mean annual rainfall) for the standard deviation model as found in this study. This issue deserves further investigation before estimating larger ARI flood quantiles based on a weighted average standard deviation and skew estimates that do not use any predictor variables.

The evaluation statistics presented above related to a particular aspect of the model validation over all six ARIs for all the states. We now look at the overall performances of the different models (QRT and PRT, fixed and ROI) based on a ratio statistic and case score analysis. The ratio is defined as $Q_{pred}/Q_{obs}$ and gives an indication of the degree of bias (i.e. systematic over- or under estimation), where a value of 1 indicates good average agreement between the $Q_{pred}$ and $Q_{obs}$. Here $Q_{pred}$ values were obtained from one-at-a-time cross validation (fixed and ROI).

The distributions of the $Q_{pred}/Q_{obs}$ ratio values for the state of NSW are shown in Fig. 8 for 5, 20 and 100 years ARIs. Here, for 5 years ARI, PRT-ROI shows the best results as the median ratio is the closest to the line corresponding to $Q_{pred}/Q_{obs} = 1$ (1-line) and the overall spread of the ratio values is the smallest. For 20 years ARI, QRT-ROI median ratio is closer to the 1-line as compared to the PRT-ROI, however, the overall spread of the ratio values for both the QRT-ROI and PRT-ROI is very similar. For 100 years ARI, QRT-ROI shows noticeable overestimation and PRT-ROI shows some underestimation as the median ratio value is located just below the 1-line.

Considering all the three states, a case score analysis of the $Q_{pred}/Q_{obs}$ ratio values is presented below. A $Q_{pred}/Q_{obs}$ 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 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 as both the $Q_{pred}$ and $Q_{obs}$ values have uncertainties associated with them (in fact these are random variables).

The models are assessed based on which one receives the most desirable estimation on average over all the cases (i.e. 6 ARIs and 399 catchments (2394 cases for each PRT and QRT), combining NSW, VIC and QLD).

Based on the criteria set out above from the 2394 cases, the QRT and PRT fixed methods produce 1881 and 1829 cases respectively with a 'desirable estimation', which is equivalent to 78% and 76% of the cases respectively. The QRT and PRT fixed methods show that 11% and 13% of cases respectively have a 'gross underestimation'. The 'gross overestimation' for QRT and PRT fixed methods achieves 11% of the cases each.

The QRT and PRT ROI methods provide 83% and 80% of cases with a 'desirable estimation'. The 'gross underestimation' is associated with 9% of cases for both QRT and PRT, respectively. The 'gross overestimation' sites for QRT and PRT ROI are 8% and 11% of the cases, respectively. It can be seen that in both the fixed and ROI methods there are cases where the results do not have a very high degree of accuracy. Such results are typical of RFFA methods and are somewhat as expected due to simplistic nature of RFFA models, which involve many simplified assumptions. For example, addition of a greater number of predictor variables and/or use of a complex model form may increase accuracy marginally, but they are not generally significant as far as practical application of the method is concerned (e,g. see Rahman et al., 1999). Also, the error in at-site flood frequency analysis estimates (which is the base case for comparison) needs to be kept in perspective. While we see improvements in the ROI approach for QRT and PRT, the fact is that there remain a few cases where estimations are not of high accuracy. On average, however, only modest differences can be found for the QRT and PRT-ROI estimates for the majority of the cases (Table 8).

In looking at the cases where most of the 'gross overestimation' and 'gross underestimation' happened, it was found that the PRT in some cases under estimated the at-site flood quantles for the larger ARIs (50 and 100 years). Interestingly it was also found that the QRT overestimated in many cases the lower ARI (2 and 5 years) at-site flood quantile. These results were found for a range of catchments sizes over all the states.

What can be concluded overall from this evaluation is that the PRT does not provide significantly less accurate estimates than the commonly applied QRT method. In fact the PRT is a useful way to check the results from QRT to make sure estimates make sense, especially in the case where the QRT results may not increase smoothly with ARI.

## 5. Conclusions

The main objective of this study was to compare the Bayesian Generalised Least Squares (BGLS) regression approaches using a fixed and region-of-influence (ROI) framework that seeks to minimise the Bayesian model error variance (predictive uncertainty). For this purpose, data from 399 small to medium sized catchments in eastern Australia were used. Prediction equations were developed for the flood quantiles of average recurrence intervals (ARI) of 2–100 years using Quantile Regression Technique (QRT) and for the first three moments of the log-Pearson type 3 (LP3) distribution (Parameter Regression Technique (PRT). Using a method similar to forward stepwise regression and adopting a number of statistical selection criteria we were able to identify the optimal regression models to use in the ROI approach.

It was found that area and design rainfall intensity were significant for the estimation of the flood quantiles in the region using QRT, while area, design rainfall intensity, mean annual evaporation, mean annual rain, slope and forest were relatively significant in the estimation of the second and third parameters of the LP3 distribution. One-at-a-time cross validation indicated that the ROI based on the minimisation of the predictive uncertainty leads to more efficient and accurate flood quantiles estimates by both the QRT and PRT. The regression diagnostics revealed that the catchment variables alone may not pick up all the heterogeneity in the regional model. Both BGLS QRT-ROI and PRT-ROI showed improvements in regional heterogeneity with an increase in the average Pseudo $R^2$ GLS and a decrease in the model error variance, average variance of prediction and the average standard error of prediction.

Both the standardised residual and QQ-plots of the ROI analysis satisfied the underlying model assumptions better than the fixed region regression. It was shown that both BGLS QRT-ROI and PRT-ROI produce smaller average $RMSE_r$ and $RE_r$ when compared to the fixed region regression approach. Based on the evaluation statistics overall it was found that there are only modest differences between the BGLS QRT-ROI and PRT-ROI which suggests that the PRT is a viable alternative to QRT in RFFA for the ungauged catchment case.

The RFFA methods developed in this paper is based on the database available in eastern Australia. It is expected that availability of a more comprehensive database (in terms of both quality and quantity) will further improve the predictive performance of both the fixed and ROI based RFFA methods presented in this study, which however needs to be investigated in future when such a database is available.

## Acknowledgements

This work was supported by Project 5, Regional flood methods for Australia, in relation to the upgrade of Australian Rainfall and Runoff. The authors would like to acknowledge the financial contribution of the Department of Climate Change and Engineers Australia to carry out the study and Prof. George Kuczera (The University of Newcastle) and Prof. Jery Stedinger (Cornell University) for their guidance in the statistical analyses and Assoc. Prof. James Ball, Prof. Ashish Sharma, Mr. Erwin Weinmann, Dr. Tom Micevski, Mr. Mark Babister and Dr. William Weeks for their comments and suggestions on the work. The authors would like to also thank the two anonymous reviewers of this manuscript whose comments and suggestions have helped to improve the paper notably.

**Table A-1**
Summary of the final BGLS regression results for VIC.

| GLS regression model (VIC) | Regression coefficient | Posterior moment | | Statistics | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | St. dev | AVPO | AVPN | AIC | BIC | BPV% | $\overline{R}^2_{GLS}$ (%) |
| Mean $\mu$ | $\sigma^2_\delta$ | 0.29 | 0.042 | | | | | | |
| | $\beta_0$ (constant) | 3.22 | 0.10 | | | | | 0 | |
| | $\beta_1$ (*area*) | 0.61 | 0.040 | 0.31 | 0.29 | 0.31 | 0.32 | 0 | 63 |
| | $\beta_2$ ($^2I_{12}$) | 1.50 | 0.28 | | | | | 0 | |
| Standard deviation $\sigma$ | $\sigma^2_\delta$ | 0.043 | 0.012 | | | | | | |
| | $\beta_0$ (constant) | 1.16 | 0.10 | | | | | 0 | |
| | $\beta_1$ (*rain*) | −0.83 | 0.10 | 0.048 | 0.046 | 0.074 | 0.077 | 1 | 65 |
| | $\beta_2$ (*evap*) | 1.49 | 0.65 | | | | | 2 | |
| Skewness $\gamma$ | $\sigma^2_\delta$ | 0.034 | 0.027 | | | | | | |
| | $\beta_0$ (constant) | −0.65 | 0.051 | | | | | 0 | |
| | $\beta_1$ (*rain*) | 0.74 | 0.15 | 0.042 | 0.040 | 0.113 | 0.118 | 1 | 70 |
| | $\beta_2$ (*evap*) | −3.25 | 1.26 | | | | | 1 | |
| Flood quantiles | $\sigma^2_\delta$ | 0.27 | 0.039 | | | | | | |
| $Q_{ARI=2}$ | $\beta_0$ (constant) | 3.38 | 0.099 | | | | | 0 | |
| | $\beta_1$ (*area*) | 0.90 | 0.089 | 0.28 | 0.27 | 0.29 | 0.30 | 0 | 63 |
| | $\beta_2$ ($I_{tc,ARI=2}$) | 1.35 | 0.32 | | | | | 0 | |
| $Q_{ARI=5}$ | $\sigma^2_\delta$ | 0.29 | 0.043 | | | | | | |
| | $\beta_0$ (constant) | 4.17 | 0.10 | | | | | 0 | |
| | $\beta_1$ (*area*) | 0.92 | 0.098 | 0.31 | 0.30 | 0.32 | 0.33 | 0 | 61 |
| | $\beta_2$ ($I_{tc,ARI=5}$) | 1.32 | 0.35 | | | | | 0 | |
| $Q_{ARI=10}$ | $\sigma^2_\delta$ | 0.35 | 0.039 | | | | | | |
| | $\beta_0$ (constant) | 4.55 | 0.11 | | | | | 0 | |
| | $\beta_1$ (*area*) | 0.94 | 0.055 | 0.37 | 0.35 | 0.38 | 0.39 | 0 | 57 |
| | $\beta_2$ ($I_{tc,ARI=10}$) | 1.42 | 0.35 | | | | | 0 | |
| $Q_{ARI=20}$ | $\sigma^2_\delta$ | 0.35 | 0.036 | | | | | | |
| | $\beta_0$ (constant) | 4.82 | 0.12 | | | | | 0 | |
| | $\beta_1$ (*area*) | 0.97 | 0.066 | 0.37 | 0.35 | 0.40 | 0.41 | 0 | 57 |
| | $\beta_2$ ($I_{tc,ARI=20}$) | 1.50 | 0.36 | | | | | 0 | |
| $Q_{ARI=50}$ | $\sigma^2_\delta$ | 0.47 | 0.050 | | | | | | |
| | $\beta_0$ (constant) | 5.17 | 0.14 | | | | | 0 | |
| | $\beta_1$ (*area*) | 0.99 | 0.073 | 0.49 | 0.47 | 0.53 | 0.56 | 0 | 49 |
| | $\beta_2$ ($I_{tc,ARI=50}$) | 1.62 | 0.42 | | | | | 4 | |
| $Q_{ARI=100}$ | $\sigma^2_\delta$ | 0.59 | 0.067 | | | | | | |
| | $\beta_0$ (constant) | 5.24 | 0.17 | | | | | 0 | |
| | $\beta_1$ (*area*) | 0.98 | 0.075 | 0.60 | 0.60 | 0.60 | 0.64 | 0 | 45 |
| | $\beta_2$ ($I_{tc,ARI=100}$) | 1.63 | 0.46 | | | | | 5 | |

**Table A-2**
Summary of the final BGLS regression results for QLD.

| GLS regression model (QLD) | Regression coefficient | Posterior moment | | Statistics | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | St. dev | AVPO | AVPN | AIC | BIC | BPV% | $\bar{R}^2_{GLS}$ (%) |
| Mean $\mu$ | $\sigma^2_\delta$ | 0.23 | 0.032 | | | | | | |
| | $\beta_0$ (constant) | 4.71 | 0.074 | | | | | 0 | |
| | $\beta_1$ (*area*) | 0.74 | 0.043 | 0.24 | 0.23 | 0.27 | 0.28 | 0 | 77 |
| | $\beta_2$ ($^2I_{12}$) | 1.97 | 0.15 | | | | | 0 | |
| Standard deviation $\sigma$ | $\sigma^2_\delta$ | 0.13 | 0.015 | | | | | | |
| | $\beta_0$ (constant) | 1.37 | 0.10 | | | | | 0 | |
| | $\beta_1$ (*area*) | −0.025 | 0.032 | 0.13 | 0.13 | 0.20 | 0.20 | 42 | 35 |
| | $\beta_2$ ($^2I_{12}$) | −1.41 | 0.13 | | | | | 2 | |
| Skewness $\gamma$ | $\sigma^2_\delta$ | 0.015 | 0.014 | | | | | | |
| | $\beta_0$ (constant) | −0.63 | 0.066 | | | | | 0 | |
| | $\beta_1$ ($^{50}I_{72}$) | −0.32 | 0.19 | 0.026 | 0.025 | 0.18 | 0.18 | 8 | 46 |
| | $\beta_2$ (*rain*) | 0.36 | 0.18 | | | | | 4 | |
| Flood quantiles | $\sigma^2_\delta$ | 0.26 | 0.036 | | | | | | |
| $Q_{ARI=2}$ | $\beta_0$ (constant) | 4.80 | 0.079 | | | | | 0 | |
| | $\beta_1$ (*area*) | 1.35 | 0.078 | 0.27 | 0.26 | 0.28 | 0.29 | 0 | 75 |
| | $\beta_2$ ($I_{tc,ARI=2}$) | 2.57 | 0.19 | | | | | 0 | |
| $Q_{ARI=5}$ | $\sigma^2_\delta$ | 0.17 | 0.026 | | | | | | |
| | $\beta_0$ (constant) | 5.77 | 0.080 | | | | | 0 | |
| | $\beta_1$ (*area*) | 1.16 | 0.075 | 0.18 | 0.17 | 0.17 | 0.18 | 0 | 79 |
| | $\beta_2$ ($I_{tc,ARI=5}$) | 1.95 | 0.17 | | | | | 0 | |
| $Q_{ARI=10}$ | $\sigma^2_\delta$ | 0.18 | 0.028 | | | | | | |
| | $\beta_0$ (constant) | 6.25 | 0.079 | | | | | 0 | |
| | $\beta_1$ (*area*) | 1.00 | 0.058 | 0.19 | 0.18 | 0.19 | 0.20 | 0 | 74 |
| | $\beta_2$ ($I_{tc,ARI=10}$) | 1.67 | 0.13 | | | | | 0 | |
| $Q_{ARI=20}$ | $\sigma^2_\delta$ | 0.14 | 0.025 | | | | | | |
| | $\beta_0$ (constant) | 6.59 | 0.10 | | | | | 0 | |
| | $\beta_1$ (*area*) | 0.99 | 0.065 | 0.16 | 0.15 | 0.18 | 0.19 | 0 | 77 |
| | $\beta_2$ ($I_{tc,ARI=20}$) | 1.42 | 0.17 | | | | | 0 | |
| $Q_{ARI=50}$ | $\sigma^2_\delta$ | 0.17 | 0.029 | | | | | | |
| | $\beta_0$ (constant) | 6.97 | 0.094 | | | | | 0 | |
| | $\beta_1$ (*area*) | 0.91 | 0.073 | 0.19 | 0.18 | 0.21 | 0.22 | 0 | 72 |
| | $\beta_2$ ($I_{tc,ARI=50}$) | 1.19 | 0.19 | | | | | 0 | |
| $Q_{ARI=100}$ | $\sigma^2_\delta$ | 0.20 | 0.033 | | | | | | |
| | $\beta_0$ (constant) | 7.23 | 0.099 | | | | | 0 | |
| | $\beta_1$ (*area*) | 0.86 | 0.078 | 0.22 | 0.21 | 0.25 | 0.26 | 0 | 72 |
| | $\beta_2$ ($I_{tc,ARI=100}$) | 1.01 | 0.20 | | | | | 0 | |

## Appendix A

See Tables A-1 and A-2.

## References


Bates, B.C., Rahman, A., Mein, R.G., Weinmann, P.E., 1998. Climatic and physical factors that influence the homogeneity of regional floods in south-eastern Australia. Water Resour. Res. 34 (12), 3369–3382.

Burn, D.H., 1990a. An appraisal of the "region of influence" approach to flood frequency analysis. Hydrol. Sci. J. 35 (2), 149–165.

Burn, D.H., 1990b. Evaluation of regional flood frequency analysis with a region of influence approach. Water Resour. Res. 26 (10), 2257–2265.

Burn, D.H., Zrinji, Z., Kovvalchuk, M., 1997. Regionalisation of catchments for regional flood frequency analysis. J. Hydrol. Eng. 2 (2), 76–82.

Castellarin, A., Burn, D.H., Braith, A., 2008. Homogeneity testing: how homogenous do heterogeneous cross-correlated regions seem. J. Hydrol. 360 (1–4), 67–96.

Chebana, F., Ouarda, T.B.M.J., 2008. Depth and homogeneity in regional flood frequency analysis. Water Resour. Res. 44 (11), W11422. doi:10.1029/2007WR006771.

Chokmani, K., Ourada, T., 2004. Physiographical space-based kriging for regional flood frequency estimation at ungauged sites. Water Resour. Res. 40 (12) (art. No. W12514).

Chowdhury, J.U., Stedinger, J.R., Lu, L.-H., 1991. Goodness of fit tests for regional flood distributions. Water Resour. Res. 27 (7), 1765–1776.

Cunderlik, J.M., Burn, D.H., 2006. Site-focused nonparameteric test of regional homogeneity based on flood regime. J. Hydrol. 318 (1–4), 276–291.

Dalrymple, T., 1960. Flood Frequency Analysis, Water Supply Paper 1543-A. US Geological Survey, Reston, VA.

Eng, K., Tasker, G.D., Milly, P.C.D., 2005. An analysis of region-of-influence methods for flood frequency regionalisation in the gulf-atlantic rolling plains. J. Am. Water Resour. Assoc. 41 (1), 135–143.

Eng, K., Milly, P.C.D., Tasker, G.D., 2007a. Flood regionalization: a hybrid geographic and predictor-variable region-of-influence regression method. J. Hydrol. Eng. 12 (6), 585–591.

Eng, K., Stedinger, J.R., Gruber, A.M., 2007b. Regionalisation of streamflow characteristics for the gulf-atlantic rolling plains using leverage-guided region-of-influence regression. In: EWRI World Water & Environmental Resources Congress, American Society of Civil Engineers, 2007.

Fill, H.D., Stedinger, J.R., 1995. Homogeneity tests based upon Gumbel distribution and a critical appraisal of Dalrymple's test. J. Hydrol. 166 (1–2), 81–105.

Gaume, E., Gaál, L., Viglione, A., Szolgay, J., Kohnová, S., Blöschl., G., in press. Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites. J. Hydrol. doi:10.1016/j.jhydrol.2010.01.008.

Griffis, V.W., Stedinger, J.R., 2007. The use of GLS regression in regional hydrologic analyses. J. Hydrol. 204, 82–95.

Grubbs, F.E., Beck, G., 1972. Extension of sample sizes and percentage points for significance tests of outlying observations. Technometrics 4 (14), 847–853.

Gruber, A.M., Reis Jr., D.S., Stedinger, J.R., 2007. Models of regional skew based on Bayesian GLS regression. In: World Environmental & Water Resources Conference, Tampa, Florida, May 15–18, 2007.

Gruber, A.M., Stedinger, J.R., 2008. Models of LP3 regional skew, data selection, and Bayesian GLS regression. In: EWRI World Water & Environmental Resources Congress, American Society of Civil Engineers, Honolulu, HI, May 13–16, 2008.

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, Newcastle, 30th November to 3rd December 2009.
Haddad, K., Pirozzi, J., McPherson, G., Rahman, A., Kuczera, G., 2009. Regional flood estimation technique for NSW: application of generalised least squares quantile regression technique. In: 32nd Hydrology and Water Resources Symposium, 30 November to 3 December, Newcastle, Australia, pp. 829–840.
Haddad, K., Zaman, M., Rahman, A., 2010a. Regionalisation of skew for flood frequency analysis: a case study for eastern NSW. Aust. J. Water Resour. 14 (1), 33–41.
Haddad, K., Rahman, A., Weinmann, P.E., Kuczera, G., Ball, J.E., 2010b. Streamflow data preparation for regional flood frequency analysis: lessons from south-east Australia. Aust. J. Water Resour. 14 (1), 17–32.
Haddad, K., Rahman, A., Green, J., 2011. Design rainfall estimation in Australia: a case study using L moments and generalised least squares regression. Stoch. Environ. Res. Risk Assess. 25 (6), 815–825. doi:10.1007/s00477-010-0443-7.
Hosking, J.R.M., Wallis, J.R., 1993. Some statistics useful in regional frequency analysis. Water Resour. Res. 29 (2), 271–281.
Institution of Engineers Australia (I. E. Aust.) 1987, 2001. In: Pilgrim, D.H. (Ed.), Australian Rainfall and Runoff: A Guide to Flood Estimation, vol. 1, I. E. Aust., Canberra.
Interagency Advisory Committee on Water Data (IAWCD), 1982. Guidelines for Determining Flood Flow Frequency: Bulletin17-B (Revised and Corrected). Hydrol. Subcomm., Washington, DC, March 1982, p. 28.
Ishak, E., Haddad, K., Zaman, M., Rahman, A., 2011. Scaling property of regional floods in New South Wales Australia. Nat. Haz. J.. doi:10.1007/s11069-011-9719-6.
Kuczera, G., 1983. Effect of sampling uncertainty and spatial correlation on an empirical Bayes procedure for combining site and regional information. J. Hydrol. 65, 373–398.
Kuczera, G., 1999. Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference. Water Resour. Res. 35 (5), 1551–1557.
Lu, L.-H., Stedinger, J.R., 1992. Sampling variance of normalized GEV/PWM quantile estimators and a regional homogeneity test. J. Hydrol. 138 (1–2), 223–245.
Madsen, H., Rosbjerg, D., 1997. Generalised least squares and empirical Bayes estimation in regional partial duration series index-flood modelling. Water Resour. Res. 33 (4), 771–782.
Madsen, H., Mikkelsen, P.S., Rosbjerg, D., Harremoes, P., 2002. Regional estimation of rainfall intensity duration curves using generalised least squares regression of partial duration series statistics. Water Resour. Res. 38 (11), 1–11.
Merz, R., Blöschl, G., 2005. Flood frequency regionalisation—spatial proximity vs. catchment attributes. J. Hydrol. 302, 283–306.
Micevski, T., Kuczera, G., 2009. Combining site and regional flood information using a Bayesian Monte Carlo approach. Water Resour. Res. 45, W04405. doi:10.1029/2008WR007173.
Nezhad, M.K., Chokmani, K., Ouarda, T.B.M.J., Barbet, M., Bruneau, P., 2010. Regional flood frequency analysis using residual kriging in physiographical space. Hydrol. Process. 24, 2045–2055. doi:10.1002/hyp. 7631.
Ouarda, T.B.M.J., Girard, C., Cavadias, G.S., Bobée, B., 2001. Regional flood frequency estimation with canonical correlation analysis. J. Hydrol. 254, 157–173.
Pandey, G.R., Nguyen, V.T.V., 1999. A comparative study of regression based methods in regional flood frequency analysis. J. Hydrol. 225, 92–101.
Rahman, A., 2005. A quantile regression technique to estimate design floods for ungauged catchments in south-east Australia. Aust. J. Water Resour. 9 (1), 81–89.
Rahman, A., Bates, B.C., Mein, R.G., Weinmann, P.E., 1999. Regional flood frequency analysis for ungauged basins in south-eastern Australia. Aust. J. Water Resour. 3 (2), 199–207.
Rahman, A., Haddad, K., Kuczera, G. and Weinmann, P.E., 2009. Regional Flood Methods for Australia: Data Preparation and Exploratory Analysis. Australian Rainfall and Runoff Revision Projects, Project 5 Regional Flood Methods, Stage I Report No. P5/S1/003, Nov 2009, Engineers Australia, Water Engineering, 181pp.
Reis Jr., D.S., Stedinger, J.R., Martins, E.S., 2005. Bayesian GLS regression with application to LP3 regional skew estimation. Water Resour. Res. 41, W10419. doi:10.1029/2004WR00344.
Rosbjerg, D., Madsen, H., 1994. Uncertainty measures of regional flood frequency analysis estimators. J. Hydrol. 167, 209–224.
Stedinger, J.R., Lu, L.H., 1995. Appraisal of regional and index flood quantiles estimators. Stoch. Hydrol. Hydraul. 9 (1), 49–75.
Stedinger, J.R., Tasker, G.D., 1985. Regional hydrologic analysis, 1. Ordinary, weighted, and generalised least squares compared. Water Resour. Res. 21 (9), 1421–1432.
Stedinger, J.R., Tasker, G.D., 1986a. Regional hydrologic analysis, 2. Model error estimators, estimation of sigma and log – Pearson type 3 distributions. Water Resour. Res. 22 (10), 1487–1499.
Stedinger, J.R., Tasker, G.D., 1986b. Correction to regional hydrologic analysis, 1. Ordinary, weighted, and generalised least squares compared. Water Resour. Res. 22 (5), 844.
Stedinger, J.R., Vogel, R.M., Foufoula-Georgiou, E., 1993. Frequency analysis of extreme events. In: Handbook of Hydrology, McGraw Hill Book Co., NY, pp. 18.1–18.66 (chapter 18).
Tasker, G.D., Stedinger, J.R., 1989. An operational GLS model for hydrologic regression. J. Hydrol. 111, 361–375.
Tasker, G.D., Eychaner, J.H., Stedinger, J.R., 1986. Application of generalised least squares in regional hydrologic regression analysis. US Geol. Surv. Water Supply Paper 2310, 107–115.
Tasker, G.D., Hodge, S.A., Barks, C.S., 1996. Region of Influence regression for estimating the 50-year flood at ungauged sites. Water Resour. Bull. 32 (1), 163–170.
Thomas, D.M., Benson, M.A., 1970. Generalization of streamflow characteristics from drainage basin characteristics. US Geol. Surv. Water Supply Paper 1975, 55.
Wiltshire, S.E., 1986a. Identification of homogeneous regions for flood frequency analysis. J. Hydrol. 84 (3–4), 287–302.
Wiltshire, S.E., 1986b. Regional flood frequency analysis I: homogeneity statistics. Hydrol. Sci. J. 31 (3), 321–333.
Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. John Wiley and Sons, Inc., New York.
Zrinji, Z., Burn, D.H., 1994. Flood Frequency analysis for ungauged sites using a region of influence approach. J. Hydrol. 153 (1–4), 1–21.
Zrinji, Z., Burn, D.H., 1996. Regional flood frequency with hierarchical region of influence. J. Water Resour. Plan. Manage. 122 (4), 245–252.