

## Towards a new regional flood frequency analysis method for Western Australia

K. Haddad<sup>a</sup>, A. Rahman<sup>a</sup>, W. Weeks<sup>b</sup>, G. Kuczera<sup>c</sup>, P.E. Weinmann<sup>d</sup>

<sup>a</sup>*School of Engineering, University of Western Sydney, NSW, Australia*

<sup>b</sup>*Department of Transport and Main Roads, Queensland, Australia*

<sup>c</sup>*School of Engineering, The University of Newcastle, Australia*

<sup>d</sup>*Department of Civil Engineering, Monash University, Victoria, Australia*

*Email: a.rahman@uws.edu.au*

**Abstract:** The development of a regional flood frequency analysis (RFFA) method for Western Australia (WA) poses a number of challenges: the region is very large and sparsely gauged, except for the South-west of the State, and in this more densely gauged part there is a significant variation of hydrological characteristics from coastal to inland areas. This paper describes how these challenges can be overcome by dividing the state into three different regions and by selecting the most appropriate RFFA method from a range of candidate methods. The database adopted for WA comprises data from 146 catchments. The State was divided into three distinct regions since there are concentrations of stream gauging stations in these three parts, which are separated by long distances. These three regions are (i) Kimberley region, with 14 stations (top part of WA from Drainage Division VIII) (ii) Pilbara region, with 12 stations (middle western part of WA, Drainage Division VII) and (iii) South-west WA, with 120 stations (Drainage Division VI). The analysis adopted in this study includes a Bayesian Generalised Least Squares (BGLS) regression approach in developing prediction equations for selected flood quantiles (Quantile Regression Technique – QRT) or for the first 3 moments of the LP3 distribution (Parameter Regression Technique – PRT). The BGLS regression model explicitly accounts for the sampling variability in the dependent variable data e.g. inter-station correlation and variation in record lengths from site to site, and offers a better alternative to the most commonly adopted Ordinary Least Squares regression.

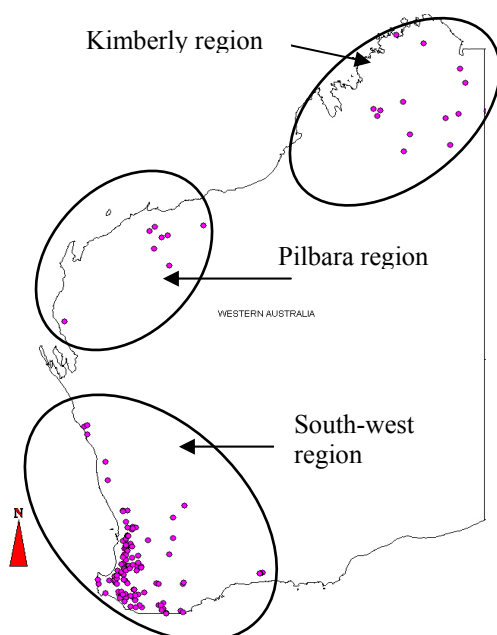
For the South-west of WA, two methods for forming regions are considered. Firstly, in the fixed region approach, all the 120 sites are assumed to form one region. Secondly, in the Region of Influence (ROI) approach, a region is formed around each of the 120 sites, with an appropriate number of stations, based on the criterion of minimum model error variance. To make the best use of the data, the developed prediction equations are validated using one-at-a-time cross validation. For South-west WA it has been found that area and design rainfall intensity are the best predictor variables for use with the QRT and PRT based on a number of statistical diagnostic criteria. It has been found that ROI outperforms the fixed regions and that ROI consistently delivers better regions with reduced level of heterogeneity on average. This is evidenced by the reduction in model error variance, standard error of prediction and average variance of prediction with the ROI approach in comparison to the fixed region. Based on independent testing it was found that the QRT-ROI and PRT-ROI methods perform very similarly with relative root mean square errors (RMSEr) ranging between 13% to 19% for PRT and 8% to 15% for the QRT, which compare very well with the validation results of RFFA studies from other Australian states. For the Kimberly and Pilbara regions both QRT and PRT in the BGLS framework are tested using a fixed region approach. The results for both the regions show that area and design rainfall intensity are the best predictor variables for use with the QRT and PRT. Independent testing from one-at-a-time validation for the Kimberly region reveals that QRT and PRT are very similar in predictive performance in that the RMSEr are in the range of 3% to 4% for the QRT and 4% to 5% for the PRT. Similarly the validation for the Pilbara region reveals satisfactory results, with the RMSEr ranging from 5% to 8% for the QRT and 4% to 7% for the PRT. This suggests that both QRT and PRT are viable options for RFFA in WA and that ROI is the best option to deal with the spatial heterogeneity of flood data in South-west WA. The findings from this study and other on-going RFFA studies will form the basis of recommendations for new RFFA methods in the new Australian Rainfall and Runoff for WA.

**Keywords:** *Regional flood frequency analysis, Western Australia, Bayesian Generalised Least Squares Regression, Quantile Regression Technique, Parameter Regression Technique*

## 1. INTRODUCTION

Australian Rainfall and Runoff 1987 (ARR87) recommended various regional flood frequency analysis (RFFA) methods for the different states of Australia depending on the data availability in respective states. In ARR87, the recommended methods for Western Australia (WA) were based on both a rational method and an index type RFFA approach. To apply these RFFA methods, five different regions based on climate and rainfall regimes were identified being the “South-west”, “Wheat Belt”, “Kimberley” and “Arid Interior”; these regions were further sub-divided on the basis of vegetation and soil type (see Figure 1.7 in ARR87, page 14, Book IV). It should be noted that the quality and quantity of the streamflow data used for deriving these RFFA methods varied quite considerably. For the South-west region the quality of the data was considered to be good, however the data quality deteriorated with the distance from the South-west region. Since 1987, the RFFA methods in ARR for WA have not been upgraded similar to other Australian states even though there have been notable developments in both at-site and RFFA techniques in Australia and internationally (e.g. Tasker and Stedinger 1989; Reis *et al.*, 2005 and Micevski and Kuczera, 2009). Also, there has been an additional 20 years of streamflow data available at many gauging locations in WA. As a part of the current upgrade of ARR, new RFFA methods for WA are being investigated similar to other Australian states.

The main objective of this paper is to develop a RFFA method for WA using Bayesian Generalised Least Squares (BGLS) regression. The advantages of GLS regression are that this technique accounts for the inter-station correlation and varying record lengths from site to site. In this paper, the BGLS regression technique is used to regionalise the first three moments of the log Pearson Type 3 (LP3) distribution, which is referred to as Parameter Regression Technique (PRT) here. Also, a Quantile Regression Technique (QRT) is developed in this paper where prediction equations are developed for selected flood quantiles for comparison with the PRT. Two approaches are used for the formation of regions: (a) fixed regions where all the available stations are considered to form one region; and (b) a Region of Influence (ROI) approach (Burn, 1990) where each of the stations form its own region by selecting appropriate number of nearby stations. For South-west WA both the fixed region and ROI approaches were tested for two reasons (i) this is a vast area with a good number of stations being available and (ii) the South-west region is likely to have different sub-regions due to different climate and catchment regimes. With the ROI a region is formed around each of the study catchments, where the size of the ROI is guided by the criterion of ‘minimum model error variance’; this is likely to reflect any natural variability in regions. For both the Kimberly and Pilbara regions there was not many stations available (only 14 and 12 respectively) and given that they were situated at great distances apart a fixed region regression approach was adopted for each of the Kimberly and Pilbara regions. Model performance for all the three regions is examined using a one-at-a-time (OAT) cross validation approach where one site is left out before developing the model, and then the developed model is applied to this catchment and the procedure is repeated until each of the catchments in the data set has been independently tested. A number of statistics are used to select the best set of predictor variables in the BGLS regression. The relative root mean square error was used to identify the best performing model between the fixed regions and ROI with both the QRT and PRT for South-west WA and fixed region QRT and PRT for the Kimberly and Pilbara regions.



**Figure 1.** Geographical locations of the streamflow gauging stations used in this study.

## 2. DATA PREPARATION

A total of 146 catchments were selected from Western Australia (WA) for this study. The spatial distributions of these catchments are shown in Figure 1. These sites are mostly unregulated and have not been affected by major land use changes (catchments with more than 10% of the area affected by urbanisation were excluded). The streamflow data of these sites were prepared following a stringent procedure as described in Haddad *et al.* (2010). The lengths of the annual maximum flood series in the catchments vary between 20 and 57 years (mean: 31 years). The catchment areas in the database are in the range of 0.1 to 7406 km<sup>2</sup> (mean: 323 km<sup>2</sup>). Based on the findings from previous studies

(e.g. Rahman *et al.*, 2009), a total of seven explanatory variables were used: (i) catchment area expressed in  $\text{km}^2$  (*area*); (ii) design rainfall intensities (mm/h) for the 2-year average recurrence interval (ARI) with 1 and 12-hour durations ( ${}^2I_1$ ,  ${}^2I_{12}$ ), 50-year ARI with 1-hour duration ( ${}^{50}I_1$ ) and 50-year ARI with 12-hour duration ( ${}^{50}I_{12}$ ), design rainfall intensity values in mm/h  $I_{AR,t_c}$  (where ARI = 2, 5, 10, 20, 50 and 100 years and  $t_c$  = time of concentration (hour), estimated from  $t_c = 0.76(\text{area})^{0.38}$ ); (iii) mean annual rainfall expressed in mm/y (*rain*); (iv) mean annual evapo-transpiration expressed in mm/y (*evap*); (v) stream density expressed in  $\text{km}/\text{km}^2$  (*sden*); (vi) main stream slope expressed in m/km (*SI085*) and (vii) forest cover expressed as a percentage (%) of catchment area (*forest*). The explanatory variables were transformed using natural logarithms. Each log-transformed explanatory variable was centered by subtracting its mean value (obtained from the available data set), so that the intercept term in the regression equation represents the log-mean of the observed dependent variable data.

### 3. METHODS

#### 3.1. At-site Flood Frequency Analysis

The at-site flood quantiles for each of the selected stations were estimated by fitting the LP3 distribution (I. E. Aust, 1987) to the observed annual maximum flood series using the Bayesian parameter fitting procedure in FLIKE (Kuczera, 1999). No prior information was used in fitting the data, low outliers were censored and the effects of rating curve errors were incorporated in the flood quantile estimates using the inbuilt facility in FLIKE. The average recurrence intervals (ARIs) considered in this study was 2, 5, 10, 20, 50 and 100 years. The mean, standard deviation and skew values of  $\ln(Q)$  for each site  $i$ , were extracted from FLIKE for use with the PRT (where  $Q$  represents annual maximum flood series in  $\text{m}^3/\text{s}$ ).

#### 3.2. Generalised Least Squares Regression and Bayesian Approach

The GLS regression framework (e.g. Tasker and Stedinger, 1989) can be used to develop empirical relationships between the desired hydrologic variable, such as the flood quantile or mean, standard deviation and skew of the annual maximum flood series, and catchment and climatic characteristics data. The GLS analysis assumes that the hydrological variable of interest denoted by  $y_i$  for a given site  $i$  can be described by a linear function of catchment and climatic characteristics ( $X$ ) (e.g. design rainfall intensity or catchment area) with an additive error. Following Reis *et al.* (2005):

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

where ( $j = 1, \dots, k$ ) are predictor variables,  $n$  = number of sites in the region,  $\beta$  are the regression coefficients,  $\delta_i$  is the model error which is assumed to be normally and independently distributed with model error variance  $\sigma_\delta^2$ . In most cases only an at-site estimate of  $y_i$  is available. To describe the error in the data, a sampling error  $\eta_i$  must be introduced into the model. Thus the observed regression model error  $\varepsilon_i$  is the sum of the model error and the sampling error. The sampling error in the sample estimators of the flood quantiles and the parameters of the LP3 distribution may in some situations explain all the variability observed in the data, which means that  $\hat{\sigma}_\delta^2$  may become negative. A negative model error variance is unrealistic as noted by Reis *et al.* (2005). In this situation, a Bayesian estimator (BGLS) of the model error variance may be used to safeguard against this happening. Further details on this 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 and quantifying associated uncertainty about it.

#### 3.3. Region of Influence Approach for South-west Western Australia

The Region-of-Influence (ROI) approach (Burn, 1990) is based on the concept of pooling of appropriate ‘nearby’ sites into groups. Each site of interest (i.e. site for which flood quantiles or moments are to be estimated) has its own ‘unique’ group. In any ROI approach the selection of a distance measure for defining the ‘unique’ group is needed. For this study the ROI is based on the geographical distance from the site of interest. Initially we carry out a fixed region BGLS regression to identify the best set of predictor variables in the region that minimises the model error variance or heterogeneity in the model, while also satisfying a number of statistical diagnostic measures. In the second step, the ROI is implemented to reduce the heterogeneity that is not accounted for by the catchment variables. In this step, the first 15 sites nearest to the site in question are selected and the BGLS regression is performed and the model error variance is noted. The

ROI then goes on to select the next five closest sites which are then added to the previously selected 15 sites and the BGLS regression is repeated. The procedure continues until all the sites (here 120) in the database are included in the analysis. The optimum ROI for the ungauged site of interest is selected as the one which exhibits the lowest model error variance.

### 3.4. Selecting Predictor Variables in the GLS regression

This section describes the approach adopted for selecting explanatory variables that should be included in the prediction equations. We use a procedure similar to stepwise regression utilising all the 120 sites for South-west Western Australia (WA) and the 12 and 14 sites for the Pilbara and Kimberly regions respectively and initially only adopting a constant term in the regression equation. The model error variance and its standard error values are noted. We then add predictor variables starting with *area* followed by different combinations of other predictor variables. In all, 16 different combinations of predictor variables were considered for the mean, standard deviation and skew models, while 25 combinations were trialed for the flood quantile models for South-west WA and Pilbara regions, while 11 combinations for the mean, standard deviation, skew and flood quantile models were tested for the Kimberly region. The choice for the preferred regional BGLS model was the combination that best satisfied all the following statistical measures, including: (i) minimum model error variance; (ii) both the minimum average variance of prediction for a new and old station (AVPN) and (AVPO) – 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 parameters; (iii) the minimum Akaike and Bayesian information criteria (AIC and BIC); these two statistics penalise more heavily models with greater number of predictors, i.e. the inclusion of a predictor variable must significantly improve the model if it is to be included; and (iv) the highest Pseudo  $R^2$  value ( $\bar{R}_{GLS}^2$ ). In all the cases the simplest model was always preferred.

### 3.5. Regression Diagnostics

The assessment of the regional regression model was made by using statistical diagnostics such as the standard error of prediction in percent (SEP). An analysis of variance for the BGLS models was undertaken to examine the sampling and model errors. The analysis of residuals provides a means of assessing the model fit and identifying potential outliers. In this study, the standardised residual was used, which is the raw residual divided by the square root of its variance, which also takes into account the sampling error variance. To assess the adequacy of the model in estimating flood quantiles, a Z score was used, here the numerator is the difference between the at-site flood quantile and regional flood quantile (estimated from the developed regression equation, QRT and PRT) and the denominator is the square root of the sum of the variances of the at-site and regional flood quantiles in natural logarithmic space.

### 3.6. Evaluation Statistics

A one-at-a-time (OAT) cross validation procedure was applied to assess the performance of both the BGLS-QRT and PRT fixed and ROI 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 in turn as ungauged for ROI, this automatically satisfies the OAT validation approach. The following performance statistic was calculated from the fixed and ROI analysis: the relative root mean square error ( $RMSE_r$ ) as described below.

$$RMSE_r(\%) = 100 \times \sqrt{\frac{1}{n} \sum_{i=1}^n \left( \frac{Q_{pred} - Q_{obs}}{Q_{obs}} \right)^2} \quad (2)$$

where  $Q_{obs}$  is the observed flood quantile obtained from at-site flood frequency analysis estimated using FLIKE (Kuczera, 1999);  $Q_{pred}$  is the predicted flood quantile from the regional prediction equation and  $n$  is the sample size. The  $RMSE_r$  (%) statistic provides an indication of the overall accuracy of the regional model. The model with the smallest  $RMSE_r$  between the two competing models with the same number of parameters is generally preferred.

## 4. RESULTS

Figure 2 shows example plots of the statistics used in selecting the best set of predictor variables for the skew coefficient of the LP3 distribution for the Kimberly region. For the skew model, combination 1 with no explanatory variable had the lowest model error variance ( $\sigma_\delta^2$ ) of 0.0858. The AVPO and AVPN values were

also the lowest with combination 1 (see Figure 2). There is enough evidence to stay with combination 1 (it may be argued that a regional average skew is applicable), as the increase in  $\bar{R}_{GLS}^2$  value for combination 8 was insignificant. Also the variables with combination 7 are not highly significant as the regression coefficients with both *area* and *rain* were less than two standard deviations away from zero. In this case, combination 1 was adopted. For  $Q_{10}$  the model error variance showed that combinations 3, 4, 5, 6, 7 and 8 were the top potential sets of predictor variables. All these combinations did not differ greatly in model error variance and  $\bar{R}_{GLS}^2$  values. The AVPO, AVPN, AIC and BIC were also very similar. In this case, combination 6 with variables *area* and  $^2I_{12}$  was adopted. It was also found that the variable  $^2I_{12}$  was significant in the regression analysis for all the flood quantile models with the regression coefficient being greater than three times the posterior standard deviation away from zero. The same procedure as discussed above was adopted in selecting the best set of predictor variables for other models with the QRT. The set of predictor variables selected as above were used in the OAT validation approach. The same analysis as presented above was carried out for the Pilbara and South-west regions of WA. For the Pilbara region both *area* and  $^2I_{12}$  were significant variables for the mean flood and flood quantile models. For the standard deviation both *area* and *forest* were not significant, while for the skew coefficient both *area* and *sden* were not significant. For South-west WA, both *area* and  $I_{ARb1c}$  were significant predictors for the flood quantile models. For the parameters of the LP3 distribution, *rain* and *evap* were relatively good predictors for the skew and *area* and  $^2I_{12}$  were the best predictors for the mean flood model. From this analysis the developed equations are taken to be the best prediction equations for the fixed region and ROI methods.

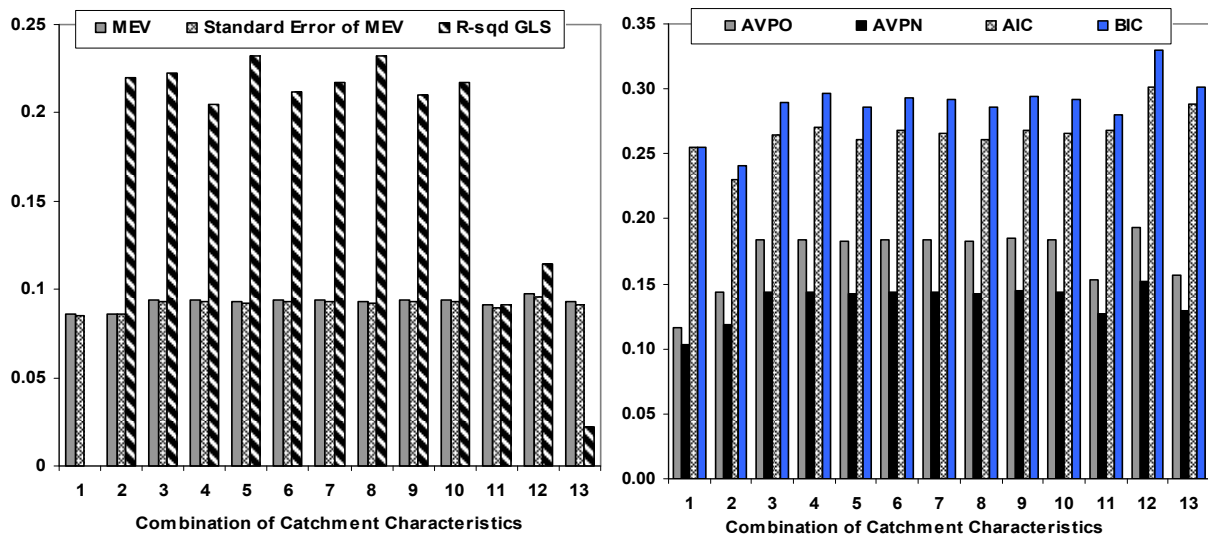


Figure 2. Selection of predictor variables for the BGLS regression model for skew

The ROI results for South-west WA are summarised as follows, the average number of sites selected in the ROI for the moments and flood quantiles of the LP3 distribution ranged between 42 (mean flood) and 71 ( $Q_{100}$ ) stations. The results revealed that the ROI mean flood typically has fewer sites than the ROIs for the standard deviation and skew. The results also showed that the skew ROI model has the highest number of sites which includes nearly all the sites in South-west WA. Interestingly, it was found that the ROI of the ARI = 2 years model has a slightly higher model error variance ( $\sigma_{\delta}^2 = 0.61$ ) than that of the ROI ARI = 100 years ( $\sigma_{\delta}^2 = 0.60$ ). This was in line with the ANOVA results (discussed below), which also show that in this study the ARI = 2 years estimates exhibit greater heterogeneity. In terms of heterogeneity, it was found overall that the predictive variance for a fixed region regression for South-west WA can be inflated by up to 32% for the flood quantile estimates. An important point to note is that there could be an inherent weakness in the fixed region models, as they tend to include more sites, which can inflate  $\sigma_{\delta}^2$ , where this is not related to the catchment variables alone. Another important point noted is that physical distance may become a useful surrogate for the unknown processes in RFFA. The spatial variation for the grouped values of the minimum model error variances for the mean flood model were plotted. It was observed that distinct spatial patterns were formed, which resembled the ARR1987 regions. The results of this analysis reveal that South-west WA is highly heterogeneous and that BGLS-ROI might be the most appropriate RFFA technique to deal with this heterogeneity.

### 4.1. Regression Diagnostics

Below we present a Pseudo Analysis of Variance (ANOVA) for the fixed region (OAT cross validation) and ROI approaches. The ANOVA table describes how the total variation among the  $\hat{y}_i$  values can be portioned among that explained by the model error and the sampling error. This is an extension of the ANOVA table in OLS regression which does not recognise and correct for the expected sampling variance. The results for the flood quantile model for  $Q_{20}$  are presented in Table 1 for South-west WA. The  $Q_{20}$  experienced the lowest error variance ratio (EVR, sampling error variance to model error variance), for all the regions especially the Kimberly region. This reflects the much greater spatial variability of the mean which is dominated by local factors (as compared to the higher moments) which is manifested in  $Q_{20}$  as it is mostly dominated by the mean. The  $Q_{20}$  shows an EVR of 10% (see Table 1) for the ROI which suggests that the BGLS and 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. The EVRs for the Pilbara and Kimberly regions were 0.39 and 0.75 respectively which show that the sampling error has had some affect on the analysis; however with 12 and 14 stations respectively the sampling error may be slightly overestimated. For all the three regions the EVR for the LP3 moment predictors, the sampling error increases as the order of moment increases, therefore the EVR increases as well. In the case of South-west WA, for the fixed region and ROI for the mean flood and standard deviation models, the model errors dominate the regional analysis. This was more pronounced for the mean flood. The EVR for the skew model for South-west WA was 7 and 7.2 for the fixed region and ROI respectively, indicating the variation due to sampling error is seven times the variation due to the model error. As far as the ROI is concerned there is not a drastic change in the EVR as the skew tends to include more stations in the regional analysis given the significantly lower model error variance as compared to the sampling error.

Quantile-quantile (QQ)-plots of the standardised residuals vs. normal score are used to check if the regression assumptions are satisfied. The QQ-plot for South-west WA, ROI and ARI = 20 years is illustrated in Figure 3. Figure 3 shows that the assumption of normality and the homogeneity of variance for the standardised residuals are largely satisfied for both the QRT and PRT, with all the points closely following a straight line. The assumption of the normality of the residuals could not be rejected at the 10% level of significance using the Anderson-Darling and Kolmogorov-Smirnov tests for normality. It was found that the ROI approach approximates the normality of the residuals slightly better than the fixed region approach, as found with the standardised residual plots (not shown). Similar results were also found for the mean, standard deviation and other flood quantile models.

The analysis of results in terms of average standard error of prediction in percent (SEP) values for the flood quantiles and the parameters of the LP3 distribution for fixed region and ROI analyses for South-west WA, Pilbara and Kimberly regions is summarised below. For the parameters of the LP3 distribution over all the sites in the region, the fixed region SEPs were larger than those of the ROI for South-west WA. For the mean flood a difference of 26% was observed. The SEP values for the Pilbara and Kimberly regions for the mean flood were also reasonably high. A negligible difference in SEP was observed for the standard deviation and

Table 1. Pseudo ANOVA table for flood quantile model (ARI = 20 years), South-west WA

Source	Degrees of Freedom		Sum of Squares	EVR	
	Fixed	ROI		Fixed	ROI
Model	$k=3$	$k=3$	$n(\sigma_{\hat{y}_0}^2 - \sigma_{\hat{y}}^2) =$	23	23
Model error $\delta$	$n-k-1=116$	$n-k-1=61$	$n(\sigma_{\delta}^2) =$	108	67
Sampling error	$N = 120$	$N = 65$	$tr[\Sigma(\hat{y})] =$	7	6
Total	$2n-1 = 239$	$2n-1 = 130$	<b>Sum of the above</b>	<b>138</b>	<b>95</b>
			<b>EVR</b>	<b>0.06</b>	<b>0.10</b>

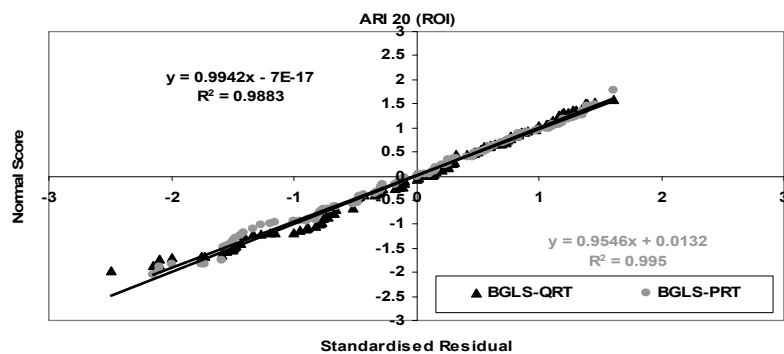


Figure 3. QQ-plot of the standardised residual vs. Z score of the fitted quantiles, 20 years ARI

skew models for ROI in South-west WA, the SEPs values for the standard deviation and skew models were modestly high because of the large sampling errors. In all the cases (fixed and ROI) the mean flood has the higher average variance of prediction (AVP) (i.e. shows greater heterogeneity) than both the standard deviation and skew model. What was striking for South-west WA is the ROI values are consistently better suggesting that ROI would provide parameter estimates of the LP3 distribution with less heterogeneity on average.

#### 4.2. Evaluation Statistics

Based on independent testing it was found that the QRT-ROI and PRT-ROI methods perform very similarly with relative root mean square errors (RMSEr) ranging between 13% to 19% for PRT and 8% to 15% for the QRT, which compare very well with the validation results from other Australian states (see Rahman *et al.*, 2011). For the Kimberly and Pilbara regions, both the QRT and PRT are tested using a fixed region approach. Independent testing from one-at-a-time validation for the Kimberly region reveals that QRT and PRT are very similar in predictive performances in that the RMSEr are in the range of 3% to 4% for the QRT and 4% to 5% for the PRT. Similarly the validation for the Pilbara region reveals satisfactory results with the RMSEr ranging from 5% to 8% for the QRT and 4% to 7% for the PRT.

### 5. DISCUSSION AND CONCLUSIONS

This study compared two regional flood frequency analysis methods, the fixed region and region-of-influence (ROI) approaches for South-west WA, the Pilbara and Kimberly regions. A BGLS regression was used to develop prediction equations for flood quantiles of ARIs of 2 to 100 years (QRT) and the first three moments of the LP3 distribution (PRT). It has been found that area and design rainfall intensity are the most significant for the estimation of the flood quantiles and parameters of the LP3 distribution. When compared to the fixed region approach, the ROI approach with both QRT and PRT shows improvements by reducing the influence of regional heterogeneity, with a decrease in the model error variance and the average standard error of prediction for South-west WA. The diagnostic plots of the ROI approach for South-west WA and fixed regions for the Pilbara and Kimberly satisfy the underlying model assumptions very well. Based on independent testing it was found that the QRT-ROI and PRT-ROI methods perform very similarly which compare very well with the validation results from other Australian states. Independent testing from one-at-a-time validation for the Kimberly and Pilbara region reveals that QRT and PRT are very similar in predictive performances as well with similar RMSEr. This suggests that both QRT and PRT are viable options for RFFA in WA and that ROI is the best option to deal with the spatial heterogeneity of flood data in South-west WA. The findings from this study and other on-going RFFA studies will form the basis of recommendations for the new RFFA methods in the upcoming Australian Rainfall and Runoff for WA.

### ACKNOWLEDGMENTS

The authors would like to acknowledge the financial contribution of the Federal Department of Climate Change and Engineers Australia to undertake the study, the Bureau of Meteorology for providing the climatic data CDs, Department of Water (WA) for providing the streamflow data, A/Prof James Ball, Professor Ashish Sharma, Mr Mark Babister and all the WA team for their suggestions and input to the project. This study is a part of revision of Australian Rainfall and Runoff (Project 5) which is aimed at the development of more precise regional flood estimation methods for Australia, which is being undertaken in association with Engineers Australia's National Committee on Water Engineering.

### REFERENCES

- Burn, D.H. (1990). An appraisal of the "region of influence" approach to flood frequency analysis. *Journal of Hydrological Sciences*, 35(2), 149-165.
- Haddad, K., Rahman, A., Weinmann, P.E., Kuczera, G., and Ball, J.E. (2010). Streamflow data preparation for regional flood frequency analysis: Lessons from south-east Australia. *Australian Journal of Water Resources*. 14(1), 17-32.
- Institution of Engineers Australia (I. E. Aust.) (1987). Australian Rainfall and Runoff: A Guide to Flood Estimation.
- Kuczera, G. (1999). Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian Inference. *Water Resources Research*, 35(5), 1551-1557.
- Micevski, T and Kuczera, G. (2009). Combining site and regional flood information using a Bayesian Monte Carlo approach. *Water Resources Research*, W04405, doi:10.1029/2008WR007173.
- Rahman, A., Haddad, K., Kuczera, G. and Weinmann, P.E. (2009). Regional flood methods for Australia: data preparation and exploratory analysis, Engineers Australia Report No. P5/S1/003, Project 5, Stage I report.
- Rahman, A., Haddad, K., Zaman, M., Ishak, E., Kuczera, G. and Weinmann, P.E. (2011). Regional flood methods for Australia, Project 5 Stage II Report (Under review).
- Reis Jr., D.S., Stedinger, J.R., and Martins, E.S. (2005). Bayesian GLS regression with application to LP3 regional skew estimation. *Water Resources Research*, 41, W10419, (1) 1029.

Haddad *et al.*, Towards a new regional flood frequency analysis method

Tasker, G.D. and Stedinger, J.R. (1989). An operational GLS model for hydrologic regression. *Journal of Hydrology*, 111, 361-375.