Design Flood Estimation in Ungauged Catchments: Quantile Regression Technique And Probabilistic Rational Method Compared

N Rijal and A Rahman

School of Engineering and Industrial Design, University of Western Sydney, Australia
E-Mail: **a.rahman@uws.edu.au**

Keywords: *Quantile Regression Technique; Probabilistic Rational Method; Ungauged catchments; Design floods; Floods; Rainfall runoff modeling*

EXTENDED ABSTRACT

Estimation of design floods in ungauged catchments is frequently required in hydrological practice and is of great economic significance. The most commonly adopted methods for this task include the Probabilistic Rational Method, the U.S. Soil Conservation Service Method, the Index Flood Method and the U. S. Geological Survey Quantile Regression Technique. The Probabilistic Rational Method has been recommended in the Australian Rainfall and Runoff for general use in south-east Australia (I. E. Aust., 1997). The central component of this technique is a dimensionless runoff coefficient which in the ARR is assumed to vary smoothly over geographical space, an assumption that may not be satisfied in many cases because two nearby catchments though are likely to share similar climatic characteristics but may exhibit quite different physical characteristics.

There has been limited study on the assessment of the Probabilistic Rational Method on independent test catchments. Recently, a Quantile Regression Technique has been proposed for south-east Australia (Rahman, 2005). This paper compares the performances of the Probabilistic Rational Method and Quantile Regression Technique for south-east Australian catchments.

The study uses streamflow and catchment characteristics data from 98 catchments in south-east Australia. A total of 20 catchments were selected randomly from the 98 catchments and put aside for independent testing of the Quantile Regression Technique and the Probabilistic Rational Method. The 20 test catchments and the 78 catchments used for the model development were found to have very similar catchment characteristics.

It has been found that the Quantile Regression Technique in general provides more accurate design flood estimates than the Probabilistic Rational Method. The 75th percentile values of the relative errors in design flood estimates for the average recurrence intervals of 2, 5, 10, 20, 50 and 100 years were in the range of 45 to 62% for the Quantile Regression Technique as compared to 61% to 80% for the Probabilistic Rational Method. It has also been found that there is a chance of about 10% that the error in design flood estimates will exceed 100% with both the Quantile Regression Technique and the Probabilistic Rational Method. Hence, the users of these techniques should be aware of this large error and provision should be made accordingly.

1. INTRODUCTION

Flood estimation in ungauged catchments is a common problem in hydrologic practice. There are several methods that are generally adopted for this task including the Probabilistic Rational Method (PRM), the U. S. Soil Conservation Service Method, the Index Flood Method and the Quantile Regression Technique (QRT). In South-east Australia, the PRM has been recommended for general use by the Australian Rainfall and Runoff (ARR) mainly due to its simplicity (I. E. Aust., 1997). The central component of this technique is a dimensionless runoff coefficient which in the ARR is assumed to vary smoothly over geographical space, an assumption that may not be satisfied in many cases because two nearby catchments though are likely to share similar climatic characteristics but may exhibit quite different physical characteristics.

Rahman (2005) presented a QRT for south-east Australia which provides reasonably accurate design flood estimates for ungauged catchments in this region. The main focus of this paper is to compare the performances of the QRT and the PRM to a set of independent test catchments in south-east Australia.

2. METHODS

2.1 Quantile Regression Technique

The United States Geological Survey (USGS) proposed Quantile Regression Technique (QRT) in that a large number of gauged catchments are selected from a region and flood quantiles are estimated from recorded streamflow data, which are then regressed against relevant climatic and catchment characteristics variables that govern the flood generation process (Benson, 1962; Cruff and Rantz, 1965; Riggs 1973). The quantile regression method is expressed as follows:

$$Q_Y = aB^bC^cD^d \ldots \quad (1)$$

where B, C, D, … are climatic and catchment characteristics variables (predictors) and Q_Y is the flood magnitude with Y year average recurrence interval (ARI) (flood quantile), and a, b, c, d, … are regression coefficients.

2.2 Probabilistic Rational Method

In Probabilistic Rational Method (PRM), the design flood for an average recurrence interval (ARI) of Y years, Q_Y (m^3/s) is given by:

$$Q_Y = 0.278 C_Y I_{t_c,Y} A \quad (2)$$

where C_Y is the dimensionless runoff coefficient for ARI of Y years, $I_{t_c,Y}$ is the average rainfall intensity (mm/h) for a design duration equal to the time of concentration t_c (h) and ARI of Y years, and A is the catchment area (km^2). For south-east Australia, t_c is obtained from:

$$t_c = 0.76 A^{0.38} \quad (3)$$

The contour maps of C_{10} in the ARR were developed based on Equation (2) and using partial series of flood peak and rainfall data from 325 gauged catchments in New South Wales and Victoria. The runoff coefficients for other Y values are computed using frequency factors provided in ARR (I. E. Aust., 1997). The preparation and use of the contour maps of C_{10} in the ARR assumes a smooth variation of C_{10} values over geographical space. Study by Rahman and Hollerbach (2003) on 104 small to medium-sized catchments in south-east Australia showed that the C_{10} values exhibit little spatial coherence and many nearby catchments showed quite different C_{10} values. Also their attempt to develop a regression equation between C_{10} and catchment characteristics were proved to be unsuccessful. The final regression equation included 6 independent variables showing an R^2 value of only 50%. An application of this regression equation to 25 test catchments provided unsatisfactory results in that over 20% of the estimated C_{10} values were found to be negative.

3. DATA

A total of 98 gauged catchments from south-east Australia were selected for this study. These catchments are mainly rural with no major regulations and land use changes over the periods of records. The catchments are small to medium sized having areas in the range of 3 to 950 km^2; the first, second and third quartiles are 128, 308 and 509 km^2, respectively. The sites have streamflow record lengths in the range of 24 to 59 years, with a mean value of 34 years and 75th percentile of 37 years. The data for these catchments were assembled in the CRC for Catchment Hydrology (Rahman et al., 1999).

An empirical distribution was fitted to each station's annual flood data using Cunnane's unbiased plotting position formula (Cunnane, 1978). The calculated ARIs were then plotted against the observed floods on normal probability paper and a best-fit line was drawn by eye, and flood quantiles Q_Y (for Y = 2, 5, 10, 20, 50 and 100 years ARIs) were read from the graph. Here, Q_Y values were obtained from the annual maximum flood series and were not converted into partial duration series. It may be noted here that given the record lengths, the estimated 50 and 100 years floods are likely to be subjected to a high degree of extrapolation and measurement error, and hence the prediction equations for these ARIs should be used with caution.

A total of 12 explanatory (predictor) variables were included in the analyses: rainfall intensity of 12-hour duration and 2-year average recurrence interval (*I*12, mm/h), mean annual rainfall (*rain*, mm); mean annual rain days (*rdays*), mean annual class A pan evaporation (*evap*, mm); catchment area (*area*, km^2); lemniscate shape, a measure of the rotundity of a catchment (*shape*); slope of the central 75% of the mainstream (*slope*, m/km); river bed elevation at the gauging station (*elev*, m); maximum elevation difference in the basin (*relief*, m); stream density (*sden*, km/km^2); fraction of basin covered by medium to dense forest (*forest*); and fraction quaternary sediment area (*qsa*). The *qsa* is a measure of the extent of alluvial deposits and is an indicator of floodplain extent in the study area. The explanatory variables *rain*, *rdays*, *evap*, and *I*12 were determined at the catchment centroid.

From the 98 catchments, twenty were selected at random and put aside for independent testing of the QRT and PRM. The remaining 78 catchments were used to develop prediction equations in the QRT and derive runoff coefficients in the PRM. The number of test catchments (20 out of 98 i.e. 20%) appears to be adequate. The statistics of the catchment characteristics of the test catchments and the 78 catchments used in the model development (model catchments) are compared in Table 1, which shows that the test and model catchments have 'similar' range of characteristics.

Table 1. Comparison of catchment characteristics of the 78 model and 20 test catchments

Character-istics	Mean		Median		Standard deviation	
	Model	Test	Model	Test	Model	Test
area (km^2)	325	334	260	296	251	265
I12 (mm/h)	4.77	4.21	4.47	4.15	1.31	0.61
sden (km/km2)	1.36	1.42	1.39	1.42	0.43	0.41
evap (mm)	1294	1235	1255	1200	174	144
qsa (fraction)	0.17	0.18	0.03	0.14	0.25	0.20

4. RESULTS

4.1 Development of Prediction Equations by Quantile Regression Technique

Each of the flood quantiles (e.g. Q_2) was regressed against the 12 predictor variables using the statistical package SPSS. A number of alternative models were developed for each of the quantiles and the one showing the highest co-efficient of determination (R^2) and lowest standard error of estimate (SEE) and satisfying the model assumptions most closely (as discussed below) were selected.

The regression coefficients in the prediction equations 4 to 9 were found to be significantly different from zero (at a significance level of 0.05 or less). The values of R^2 are reasonably high (range: 0.74-0.79) and SEEs are generally small (less than 7% of the mean observed flood quantile in log domain) for all the six quantiles. The selected regression equations were checked against the least squares assumptions (Norusis, 2000).

The normal cumulative probability plots did not show significant departure from a straight line, indicating that residuals were near-normally distributed (typical plots are shown in Figure 1). Plots of standardised residuals against the standardised predicted values did not show any systematic patterns between the predicted values and the residuals (typical plots are shown in Figure 2). Also, no pattern was detected on the plots of predicted and observed quantiles. These indicate that the assumptions of linear model and homogeneity of variance have largely been satisfied for the prediction equations.

The value of Durbin-Watson statistic was found to be in the range 1.55-1.65, which is close to 2, thus the residuals are not highly correlated. The values of Durbin-Watson statistic range from 0 to 4 and a value of 2 indicates absence of any correlation. No outlier and influential data point was found.

The selected prediction equations are given below:

$$\begin{aligned}\log(Q_2) = -3.958 + 0.682\log(area) + 1.558\log(I12) \\ + 0.741\log(sden) + 1.535\log(evap)\end{aligned} \quad (4)$$

R^2 = 0.75, Adjusted R^2 = 0.74, SEE = 0.22 (6.16% of the mean log Q_2)

$$\log Q_5 = -5.611 + 0.727\log(area) + 1.584\log(I1$$
$$+ 0.714\log(sden) + 2.124\log(evap)$$
(5)

R^2 = 0.76, Adjusted R^2 = 0.75, SEE = 0.23 (5.94% of the mean log Q_5)

$$\log Q_{10} = -5.789 + 0.674\log(area) + 1.435\log(I1$$
$$+ 2.296\log(evap) + 0.861\log(sden)$$
(6)

R^2 = 0.74, Adjusted R^2 = 0.73, SEE = 0.23 (5.81% of the mean log Q_{10})

$$\log Q_{20} = -5.464 + 0.733\log(area) + 1.610\log(I1$$
$$+ 2.141\log(evap) + 0.880\log(sden)$$
(7)

R^2 = 0.76, Adjusted R^2 = 0.74, SEE = 0.23 (5.77% of the mean log Q_{20})

$$\log Q_{50} = -6.025 + 0.710\log(area) + 1.100\log(I12)$$
$$+ 2.430\log(evap) + 0.848\log(sden)$$
$$- 0.127\log(qsa)$$
(8)

R^2 = 0.79, Adjusted R^2 = 0.77, SEE = 0.22 (5.24% of the mean log Q_{50})

$$\log Q_{100} = -6.270 + 0.714\log(area) + 1.097\log(I12)$$
$$+ 2.529\log(evap) + 0.921\log(sden)$$
$$- 0.128\log(qsa)$$
(9)

R^2 = 0.79, Adjusted R^2 = 0.77, SEE = 0.23 (5.28% of the mean log Q_{100})

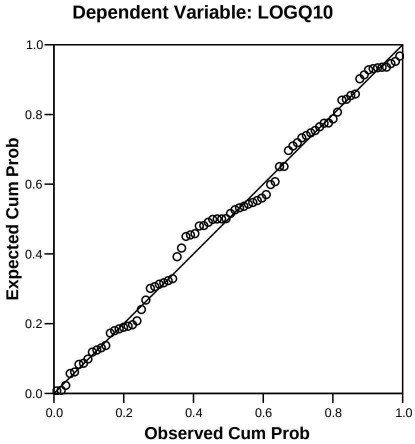


Figure 1. Normal quantile-quantile plot of residuals for logQ_{10}

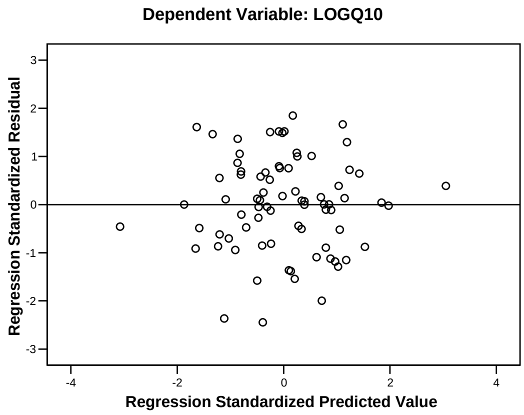


Figure 2. Plots of standardised residual and standardised predicted values for logQ_{10}

4.2 Estimation of Runoff Coefficients for the PRM

The runoff coefficient (C_Y) for a catchment was computed using Equation 2 for Y = 2, 5, 10, 20, 50 and 100 years. This requires estimates of Q_Y and $I_{t_c,Y}$. The values of Q_Y were estimated by a non-parametric method as discussed in Section 3. In obtaining the design rainfall intensity ($I_{t_c,Y}$), the time of concentration (t_c) was estimated using Equation 3. Given the Y and duration of design rainfall (taken as t_c), $I_{t_c,Y}$ value was computed at catchment centroid using the ARR method (I. E.

Aust., 1997). The values of C_{10} were estimated and plotted in a map of the area. The values of other C_Y were obtained using the frequency factors as shown in Table 2.

Table 2. Frequency factors for the PRM

Y(ARI)	Frequency Factor
2	0.56
5	0.82
10	1
20	1.09
50	1.21
100	1.35

4.3 Comparison Between QRT and PRM

For the 20 test catchments, Q_Y were estimated for Y = 2, 5, 10, 20, 50 and 100 years using the developed prediction Equations 4 to 9. These are referred to as QRT estimates. The QRT estimates for the test catchments are presented in Figure 3, which shows that flood quantiles generally increase with the ARIs for most of the test catchments but some smoothing may be required for some catchments so that flood quantiles increase consistently with ARIs.

To obtain the estimates by the PRM, the values of the runoff coefficients (C_{10}) for the 20 test catchments were estimated assuming a smooth variation over geographical space (as per the ARR method) on the plot of C_{10}. Based on these values, Q_2, Q_5, Q_{10}, Q_{20}, Q_{50} and Q_{100} were estimated using Equation 2, which are referred to as 'PRM estimates'.

Using historical streamflow data, the values of Q_2, Q_5, Q_{10}, Q_{20}, Q_{50} and Q_{100} were estimated for the test catchments using a non-parametric method, as mentioned in Section 3. These estimates are referred to as 'observed flood quantiles'. The difference between the QRT/PRM estimates and observed flood quantiles may be taken as a measure of uncertainty in design flood estimates by the QRT/PRM and are referred to as 'relative error' here.

The median values of the relative errors (ignoring the sign of the relative errors) associated with the QRT and PRM based on the 20 test catchments are presented in Table 3, which shows that QRT has remarkably smaller median relative error for 2 years ARI, for 5, 10, 20 and 50 years ARIs, both the methods have similar median relative error values, and for 100 years ARI, PRM has remarkably smaller relative error as compared to that of the QRT. In the case of the 75th percentiles of the relative errors (Table 4), QRT shows much smaller values as compared to the PRM. The relative error values are greater than 100% for 15% and 10% cases with the QRT and PRM, respectively.

Table 3. Median relative errors (%) associated with the QRT and PRM (ignoring the sign of the relative errors)

Method	Q_2	Q_5	Q_{10}	Q_{20}	Q_{50}	Q_{100}
QRT	28	41	38	36	34	47
PRM	52	43	35	38	31	36

Table 4. 75th percentile values of relative errors (%) associated with the QRT and PRM (ignoring the sign of the relative errors)

Method	Q_2	Q_5	Q_{10}	Q_{20}	Q_{50}	Q_{100}
QRT	45	55	50	57	57	62
PRM	80	70	72	61	70	73

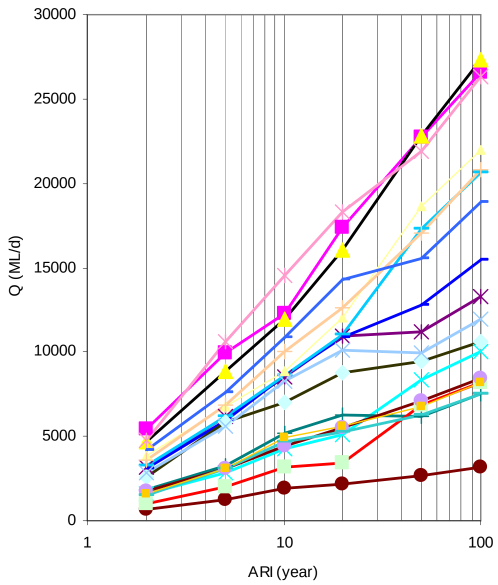


Figure 3. Flood quantiles for the test catchments using the QRT

The box plots of relative errors (considering their sign) for the QRT and PRM are shown in Figures 4 and 5 respectively, which show that median values for the QRT are closer to zero line than that of the PRM. This indicates that on average the PRM estimates will show greater bias than that of the QRT. The box plots also show that the PRM has wider error band than that of the QRT. Table 5 presents the proportion of cases that QRT/PRM underestimates/overestimates the observed flood quantiles. Here the QRT provides better results in which proportions with under and over estimation are closer to 0.5 as compared to the PRM.

Table 5. Proportion of cases with underestimation and overestimation

ARI (years)	QRT (Proportion of cases)		PRM (Proportion of cases)	
	Under-estimation	Over-estimation	Under-estimation	Over-estimation
2	0.55	0.45	0.35	0.65
5	0.45	0.55	0.35	0.65
10	0.45	0.55	0.40	0.60
20	0.45	0.55	0.55	0.45
50	0.50	0.50	0.65	0.35
100	0.40	0.60	0.65	0.35

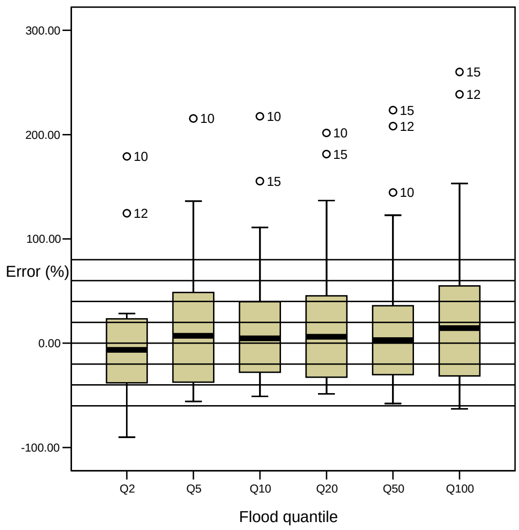


Figure 4. Box plot of the relative errors (QRT)

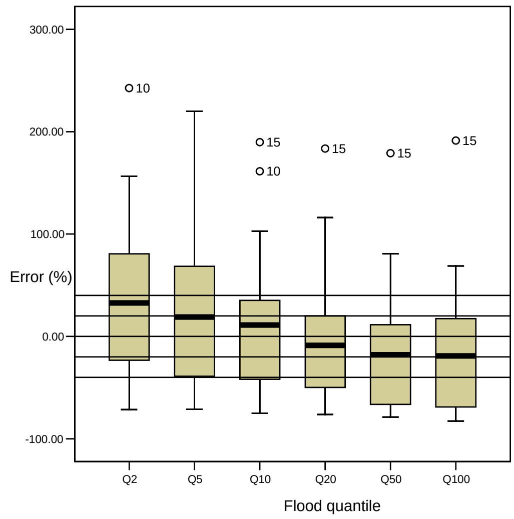
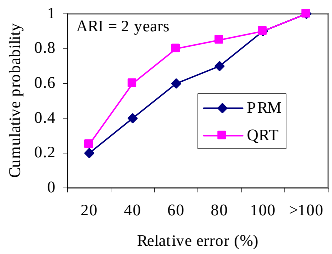
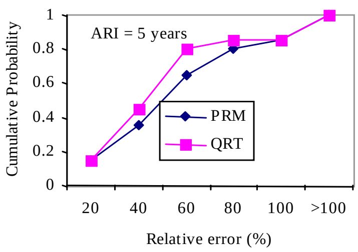
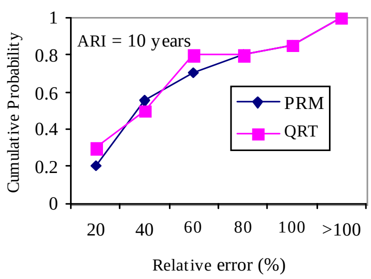


Figure 5. Box plot of the relative errors (PRM)

The cumulative distributions of relative errors for each of the ARIs are plotted in Figure 6, which shows that QRT has smaller relative error values for greater proportion of cases for 2 and 5 years ARIs. For the other ARIs, two methods show very similar distributions of relative errors.







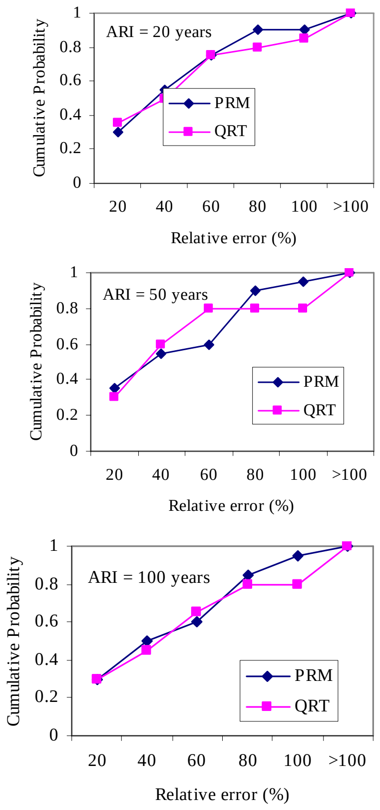


Figure 6. Distribution of relative errors

5. CONCLUSIONS

This paper compares Quantile Regression Technique and Probabilistic Rational Method for design flood estimation in south-east Australian catchments. Following conclusions can be drawn from this study:

- The Quantile Regression Technique in general provides more accurate design flood estimates than the Probabilistic Rational Method.
- The Quantile Regression Technique in general shows smaller bias in flood estimates than the Probabilistic Rational Method.
- There is a chance of about 10% that the error in design flood estimates will exceed 100% with both the Quantile Regression Technique and the Probabilistic Rational Method. The users of these techniques should be aware of this large error and provision should be made accordingly.

6. REFERENCES

Benson, M.A. (1962). Evolution of methods for evaluating the occurrence of floods. U.S. Geol. Surv. Water Supply Pap., 1580-A, 30pp.

Cruff, R.W. and Rantz, S.E. (1965) A comparison of methods used in flood frequency studies for coastal basins in California. U.S. Geol. Surv. Water Supply Pap., 1580-E, 56pp.

Cunnane, C. (1978). Unbiased plotting positions – a review. J. Hydrol., 37, 205-222.

Institution of Engineers, Australia (I. E. Aust.). (1997). Australian Rainfall and Runoff: A Guide to Flood Estimation. Book IV, I. E. Aust., Canberra, 1998.

Norusis, M.J. (2000). SPSS for Windows. SPSS Inc. Chicago, Illinois.

Rahman, A., Mein, R.G., Bates, B.C. and Weinmann, P.E. (1999). An Integrated Dataset of Climate, Geomorphological and Flood Characteristics for 104 Catchments in South-east Australia. Working Document 99/2. CRC for Catchment Hydrology, 20 pp.

Rahman, A. and Hollerbach, D. (2003). Study of Runoff Coefficients Associated with the Probabilistic Rational Method for Flood Estimation in South-east Australia In Proc. 28th Intl. Hydrology and Water Resources Symp., I. E. Aust., Wollongong, Australia, 10-13 Nov. 2003, Vol. 1, 199-203.

Rahman, A. (2005). A quantile regression technique to estimate design floods for ungauged catchments in south-east Australia. Australian Journal of Water Resources. Vol 9, No. 1. (In press).

Riggs, H.C. (1973). Regional analysis of streamflow characteristics. U.S. Geol. Surv. Water Resour. Invest. Tech., Book 4, 15pp