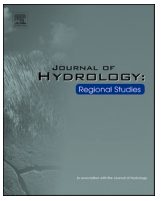
Journal of Hydrology: Regional Studies 47 (2023) 101407

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

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



Regional flood frequency analysis based on peaks-over-threshold approach: A case study for South-Eastern Australia

Xiao Pan [a], Ataur Rahman [a,*], Khaled Haddad [a], Taha B.M.J. Ouarda [b], Ashish Sharma [c]

[a] School of Engineering, Design and Built Environment, Western Sydney University, Australia
[b] National Institute of Scientific Research (INRS), Quebec, Canada
[c] School of Civil and Environmental Engineering, University of New South Wales, Australia

ARTICLE INFO

Keywords:
Floods
Peaks over threshold
Annual maximum flood
Regional flood frequency analysis
Australia
Very frequent floods

ABSTRACT

Study region: Southeast Australia
Study focus: Regional flood frequency analysis (RFFA) is a widely adopted statistical method to estimate design floods in ungauged catchments. Annual maximum flood (AMF) model is the most popular method in developing RFFA techniques. However, the AMF-based approaches are criticised for its limitations in the range of very frequent to frequent flood estimation. As an alternative, the peaks-over-threshold (POT) based approach has shown theoretical advantages in this flood range. POT based RFFA is currently underemployed internationally due to its complexity in modelling. This study develops POT-based RFFA techniques for south-eastern Australia using data from 151 catchments. A comparison is made between ordinary least squares (OLS) and weighted least squares (WLS) methods in developing POT-based RFFA techniques.
New hydrological insights for the region: The OLS based method is found to perform better than the WLS. The median relative error values of the developed prediction equations range 31–38%. The new POT-based RFFA technique overcomes the limitations of the current Australian Rainfall and Runoff, which does not have any RFFA technique for very frequent floods. It is expected that these new POT-based RFFA technique will be used in practice in south-east Australia.

1. Introduction

Flooding is one of the worst natural disasters across the globe. It leads to economic burdens due to direct and indirect consequences of flooding (Acosta et al., 2016). Accurate flood risk assessment is needed to reduce flood damage. Flood frequency analysis (FFA) is one of the most preferred methods for design flood estimation. Extreme value distributions are widely used in flood and rainfall frequency analyses (e.g., Yilmaz et al., 2017; Hossain et al., 2021). Traditionally, FFA adopts annual maximum flood (AMF) model, which extracts the single highest flow value per year from the streamflow record to construct the AMF series. Adopting AMF model in

Abbreviations: AMF, Annual Maximum Flood; FFA, Flood Frequency Analysis; GPA, Generalised Pareto; LOO, Leave-one-out; MLR, Multiple Linear Regression; OLS, Ordinary Least Squares; POT, Peaks-over-threshold; QRT, Quantile Regression Technique; RFFA, Regional Flood Frequency Analysis; WLS, Weighted Least Squares.

* Correspondence to: School of Engineering, Design and Built Environment, Western Sydney University, Building XB, Second Avenue, Kingswood, NSW 2751, Australia.
E-mail address: a.rahman@westernsydney.edu.au (A. Rahman).

https://doi.org/10.1016/j.ejrh.2023.101407
Received 31 January 2023; Received in revised form 29 April 2023; Accepted 30 April 2023
Available online 4 May 2023

FFA is straightforward and this is the most commonly applied practice in Australia and internationally. However, it has been criticised for its inaccurate and biased estimates in very frequent to frequent flood ranges (Metzger et al., 2020; Zaman et al., 2012). A common practice of applying AMF involves censoring of the potential low flow values to achieve better fitting in the higher flood ranges (Ball et al., 2016).

Peaks-over-threshold (POT) approach is an alternative to AMF and has been widely employed in environmental modelling (Bernardara et al., 2011; Liang et al., 2019; Northrop and Jonathan, 2011; Thompson et al., 2009). It is flexible in the data extraction process; and theoretically advantageous in estimating very frequent to frequent floods (Karim et al., 2017). Moreover, the additional flood information through retaining values above a selected threshold could be a particular interest to hydrologists (Kumar et al., 2020; Madsen et al., 1997a). One can adjust the threshold by altering average events per year using the POT modelling approach. Cunnane (1973) recommended POT1.63 (average 1.63 events over the threshold per year) to reduce sample variance compared with the AMF approach. Later, a practical guideline was proposed by Lang et al. (1999), which recommended an average of one to three events above the threshold per year for the POT modelling (i.e., between POT1 and POT3). Durocher et al. (2018) further investigated and proposed a hybrid method, which suggested a lower boundary of POT1 and a higher boundary of POT5.

Regional flood frequency analysis (RFFA) is the commonly adopted technique to estimate flood quantiles in ungauged catchments. The rationale of RFFA is to transfer flood characteristics from gauged to ungauged sites based on hydrological similarity (Haddad and Rahman, 2012; Haddad et al., 2012). Generally, the RFFA involves two steps: first, forming regions and then applying a parametric distribution or index flood or quantile regression technique (QRT) (Rahman, 2005) for design flood estimation. Index flood approach is a popular RFFA technique, which depends on the assumption of homogeneity (Hosking and Wallis, 1993). In contrast, the QRT relaxes the homogeneity assumption. In QRT, a multiple linear regression (MLR) technique is applied to develop a relationship between a flood quantile (dependent variable) and selected catchment characteristics (independent variables). In MLR, ordinary least square (OLS) and weighted least square (WLS) approaches are generally used to estimate the coefficients of the regression equations. Stedinger and Tasker (1985) compared the performance of OLS and WLS in RFFA and concluded that improvement in using WLS can be significant over the OLS.

Many studies in Australia and internationally focused on flood frequency analysis for both at-site and regional cases, and the majority of the studies were based on the AMF modelling approach (Ahammed et al., 2014; Burn et al., 2007; Haddad and Rahman, 2011, 2012; Haddad et al., 2011; Ishak and Rahman, 2015; Ishak et al., 2013; Masan and Hewa, 2020; Rahman et al., 2018; Shu and Ouarda, 2007; Walpita-Gamage et al., 2020: Rahman et al., 2020). However, there is a rising interest in using the POT modelling approach due to its suitability for ecological and environmental studies as they need frequent flow estimation (Ammar et al., 2020; Jarajapu et al., 2022; Kiran and Srinivas, 2021a; Langousis et al., 2016; Pan and Rahman, 2022; Pan et al., 2022; Sccarrott and Macdonald, 2012).

Madsen and Rosbjerg (1997) proposed an index flood method with a Bayesian approach based on Generalised Pareto (GPA) distribution using POT flood data from 48 New Zealand catchments. A simple linear estimator was found to be adequate in the adopted generalized least squares regression framework. Durocher et al. (2019) applied a nonstationary index-flood model for POT data from 425 catchments in Canada, which enabled extending stationary method to both at-site and regional cases. They compared four different estimators, two were based on regression and L moments and two others applied likelihood-based techniques. The independent likelihood method was found to provide most accurate estimates of regional floods for 10 and 100 years return periods. Mostofi Zadeh et al. (2019) proposed a general framework to undertake RFFA for both the annual maximum and POT data series using data from 684 Canadian stations.

The POT-based RFFA technique is underutilised in Australia (Rahman et al., 2019). This study aims to fill this knowledge gap by developing a POT-based RFFA technique focusing on very frequent to frequent flood ranges using data from south-eastern Australia. In this context, the main purposes of this study are: (i) Development of prediction equations for selected return periods (very frequent to frequent flood ranges) based on POT modelling approach; (ii) Assessment of the performance of the developed prediction equations based on different regions in the study area; and (iii) Recommendation of the best performing prediction equations for the selected return periods. It is expected that the outcome of this study will promote the wider application of the POT-based modelling approach in flood frequency analysis.

2. Methods

2.1. Selection of events

In selecting the POT series at a given gauging station, the threshold value is selected based on an automated method adopting Pearson normality test as described in Pan et al. (2021). The average events per year (k = 1,2,3…k) is selected to determine the size of the POTk series. Based on series iterative processes, the threshold is determined once the critical p-value is met.

2.2. Multiple linear regression

The multiple linear regression (MLR) technique develops a relationship between the dependent variable, Q and u independent variables $X_1, X_2, \ldots, X_u$. Eq. 1 below is constructed for expressing the i_{th} observation ($i = 1,2,3,\ldots,n$), where b_0 and b_j ($j = 1,2,3, \ldots, u$) are unknown parameters associated with the error term, E_i, based on n observations. It is assumed that the term E_i is normally distributed based on a mean equal to zero. The model parameters (b_0, b_1, …, b_j) are generally estimated using an ordinary least square (OLS) approach. It is a common practice that the regional regression model is constructed based on log-transformed dependent and

independent variables (Durocher et al., 2016).

$$Q_i = b_0 + b_1X_{i1} + b_2X_{i2} + b_jX_{ij} + \ldots + + b_uX_{iu} + E_i \quad (1)$$

OLS is the most popular method to estimate parameters of Eq. (1) based on minimising the residual sum of squares. OLS is based on the assumption of homoscedasticity, which is not generally satisfied with hydrological data. In other words, the standard deviation of error terms is not constant across all values of the independent variables. To overcome this problem, weighted least square (WLS) is adopted, which treats parameters differently by assigning a weight in the parameter estimation process. Based on traditional WLS, the weights should ideally be equal to the reciprocal of the variance of the observations. Hydrologists also apply sampling error (Eq. 2) as weights in the WLS approach. This study uses statistical weight (reciprocal of variance, WS) and hydrological weight (sampling error, WH) for comparison.

$$Sampling \quad Error = \frac{Variance \quad of \quad POT \quad flood \quad series}{Size \quad of \quad POT \quad flood \quad series} \quad (2)$$

2.3. Model construction

A total of 1512 POT-based regression models are constructed and evaluated in this study for nine return periods between very frequent and frequent range of flows, which vary from 12 events per year (EY) to 10 average recurrence interval (ARI) or 10% annual exceedance probability (AEP) (12EY, 6EY, 4EY, 3EY, 2EY, 1EY, 0.5EY, 0.2EY and 10ARI). To specify probability of very frequent to frequent flows, EY is used following Australian Rainfall and Runoff (Book 2, Chapter 3) (Ball et al., 2019). Fig. 1 illustrates the adopted procedures for constructing models in this study.

The model construction starts with selecting POTk flood series for each of the selected stations based on k average events per year ($k = 1, 1.5, 2, 2.5, 3, 4, 5$). Flood quantiles are estimated for the selected nine return periods for each POTk series assuming a Generalised Pareto distribution. The next step is forming a group/region from the selected catchments for developing the prediction equations. A total of three regions are formed based on the geographical factors of the selected catchments. Region 1 includes all the selected 151 catchments in the study area (south-eastern Australia), while Regions 2 and 3 are formed based on Drainage Division 2 and Drainage Division 4 in Australia. Details of these regions are discussed in later sections.

Once regions are formed, two regression techniques are applied for each POTk flood series: OLS and WLS. In the regression, flood quantiles are dependent variables and climatic and catchment characteristics (as shown in Table 1) are independent variables. OLS technique adopts stepwise elimination based on p-value (5% significance level) (OP) and AIC value (OA), while WLS adopts statistical (WS) and hydrological (WH) weights. Baseflow factors (volume factor (VF) and peak factor (PF)) are additional independent variables when constructing models (M1 includes VF and PF; M2 excludes VF and PF). Finally, each model is evaluated based on Eq. 3 to Eq. 7

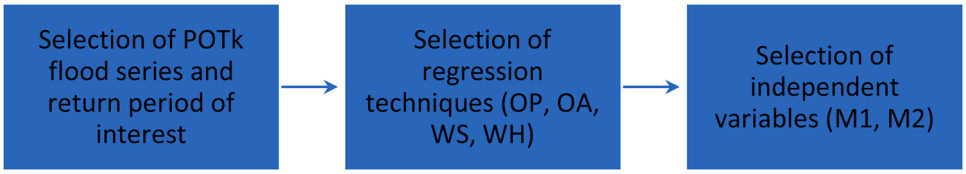


Fig. 1. Illustration of the adopted methodology for model construction process.

Table 1
Descriptive statistics of hydrological and physio-meteorological variables of selected 151 catchments in New South Wales and Victoria, Australia.

	Independent Variables	Minimum	Max	Mean	Median	Standard Deviation
Catchment characteristics	A (km^2)	11.00	1010.00	357.85	309.00	256.69
	MAR (mm)	485.32	1953.23	997.54	926.74	326.19
	SF (fraction)	0.26	1.43	0.77	0.77	0.21
	MAE (mm)	932.70	1543.30	1110.01	1066.20	129.12
	SDEN (km^{-1})	0.52	5.47	1.94	1.56	1.00
	S1085 (m/km)	0.80	69.90	12.71	9.59	10.81
	FST (fraction)	0.01	1.00	0.59	0.65	0.33
Climatic characteristics	I_{12} (mm/hr)	1.52	4.37	2.47	2.48	0.54
	I_6(mm/hr)	1.76	5.18	2.81	2.80	0.63
	I_4 (mm/hr)	2.20	6.67	3.41	3.37	0.82
	I_3 (mm/hr)	2.53	7.79	3.85	3.73	0.96
	I_2 (mm/hr)	2.92	9.45	4.50	4.31	1.18
	I_1 (mm/hr)	3.62	12.50	5.68	5.36	1.59
	$I_{0.5}$ (mm/hr)	4.62	15.90	7.09	6.66	2.06
	$I_{0.2}$ (mm/hr)	5.78	21.60	8.90	8.13	2.80
	I_{10} (mm/hr)	6.69	26.30	10.40	9.41	3.41
Baseflow index	VF (fraction)	0.00	1.82	0.25	0.21	0.26
	PF (fraction)	0.00	0.82	0.08	0.05	0.12

(using both split sample (SS) and leave-one-out (LOO) validation techniques). A typical model, 12EY_POT3_R1_OP_M1 is constructed based on 12 event per year (EY) frequency, POT3 model, data from Region 1, baseflow factors as independent variables and regressed using OLS stepwise elimination based on p-value (5% significance level).

This study first selects the best performing POTk flood series based on various statistical measures. Then the selected POTk flood series is adopted for further modelling and analysis. Finally, the performance of each region is compared and discussed.

2.4. Model evaluation

A range of statistical measures are applied to evaluate the performance of POTk RFFA models: median relative error (RE_m, Eq. 3), relative error (RE_r, Eq. 4), root mean square error ($RMSE_r$, Eq.5), the coefficient of determination (R^2, Eq. 6), and the ratio between predicted and observed flow quantiles ($Ratio_r$, Eq. 7). Q_{obs} is the at-site flood quantiles estimated using Generalised Pareto distribution (Pan and Rahman, 2022) based on POTk flood series. Q_{pred} is the estimated flood quantile based on various POT-based RFFE models. Here, a smaller RE_m value is desirable for the selected POTk RFFE model, and $Ratio_r$ closer to 1 indicates a better fit between the predicted and observed flood quantiles. R^2 is also applied for each constructed model and its values range from 0 to 1 (a value closer to 1 is preferable).

Each developed POTk RFFA model is evaluated based on split-sample (SS) (with 80%−20% split) and leave-one-out (LOO) validation techniques. The SS validation is repeated 100 times randomly for each of the developed models and is evaluated using Eq. 4.

$$RE_m(\%) = median\left|\frac{Q_{Pred} - Q_{Obs}}{Q_{Obs}}\right| * 100\% \tag{3}$$

$$RE_r(\%) = \frac{Q_{Pred} - Q_{Obs}}{Q_{Obs}} * 100\% \tag{4}$$

$$RMSE_r(\%) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{Q_{Obs} - Q_{Pred}}{Q_{Obs}}\right)^2} * 100\% \tag{5}$$

$$R^2 = 1 - \frac{Sum\ \ of\ \ squares\ \ of\ \ residuals}{Total\ \ sum\ \ of\ \ squares} \tag{6}$$

$$Ratio_r = \frac{Q_{Pred}}{Q_{Obs}} * 100\% \tag{7}$$

3. Study area and data

This study focuses on the coastal region of south-eastern Australia as shown in Fig. 2 since this part of Australia has high quality streamflow data. A total of 151 gauged catchments are selected. The area of selected catchments ranges from 11 km^2 to 1010 km^2, with

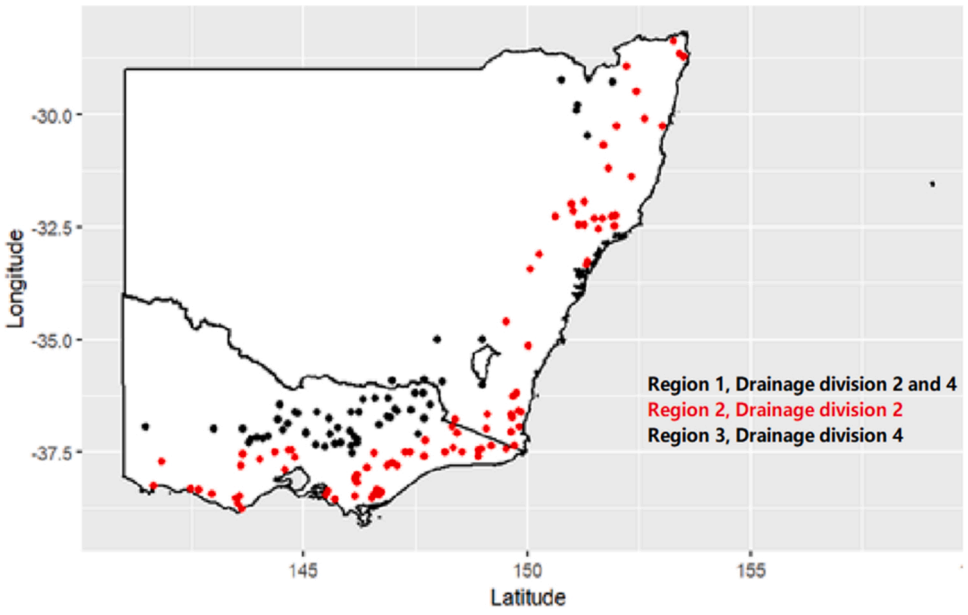


Fig. 2. Geographical locations of selected 151 catchments in New South Wales and Victoria, Australia.

an average of 357 km^2 and a median of 309 km^2. Records of streamflow data range from 27 to 82 years, with an average of 42 years and a median of 39 years. It should be noted here that the smaller the record length the higher the sampling variability in estimated quantiles. The adopted WLS method considers the impacts of record length in modelling. The selected catchments have no major change in land use and are unregulated. Among the selected 151 catchments, 57 are from New South Wales (NSW), and 94 are from Victoria (VIC) states of Australia.

This study considers three different regions: Region 1 consists of all the selected 151 catchments. Region 2 and Region 3 are formed based on Drainage Division 2 (coastal, 89 catchments) and Drainage Division 4 (inland, 62 catchments), respectively. Drainage Division 4 contains longer rivers than Drainage Division 2, which flow into the Murray Darling River system. The rationale behind forming three regions is to improve the accuracy of the developed prediction equations. Application of Hosking and Wallis (1993) test to these regions did not deliver any homogeneous regions since heterogeneity statistics (H) were much higher than threshold value of one (H1 >11). This finding is similar to previous Australian studies where homogeneous regions could not be identified (e.g., Bates et al., 1998; Rahman and Rahman, 2020).

A total of eighteen hydrological and physio-meteorological variables are adopted in this study as independent variables to develop the prediction equations based on the POT-based modelling approach. These variables are catchment areas (A), mean annual rainfall (MAR), shape factor (SF), mean annual evapotranspiration (MAE), stream density (SDEN), mainstream slope (S1085), fraction forest (FST), design rainfall intensity of 6-hr duration based on varied return period (I_i) (e.g., I_{12}, design rainfall intensity of 12EY), baseflow volume factor (VF) and baseflow peak factor (PF). The shape factor (SF) is defined as the shortest distance between a catchment's centroid and outlet divided by the square root of catchment area (A). Baseflow volume factor (VF) is defined as the ratio of baseflow volume and streamflow volume. Baseflow peak factor (PF) is defined as the ratio of peak of baseflow and peak of streamflow hydrograph. Australian Rainfall and Runoff (ARR) 2019 (Book 5, Chapter 4) has developed regional method to estimate baseflow for ungauged catchments for any location in Australia for catchment size of 7–7800 km^2 (Ball et al., 2019).

Table 1 summarises the adopted independent variables in this study. The dependent variables in regression analysis are Q_T (flood discharge for T-year return period) values, which are estimated by fitting a Generalised Pareto (GPA) distribution to each of the selected POTk series.

4. Results and discussion

4.1. Selection of best-performing POTk flood series

In this study, the best performing POTk flood series is firstly selected based on the adopted statistical measures using catchments from Region 1 (i.e., all of the selected 151 catchments) based on SS validation (80–20% split). Three evaluation statistics RE_m, $RMSE_r$ and R^2 (Eq. 3, 5 and 6) are applied for each POTk model (OP_M1, OP_M2, OA_M1, OA_M2, WS_M1, WS_M2, WH_M1, WH_M2). The return period of 12EY (smallest return period of very frequent flood) and 0.5EY (smallest return period of frequent flood) are selected and assessed and it is expected that other selected return periods will provide similar results.

Fig. 3 plots the R^2 values of very frequent (12EY) and frequent (0.5EY) POTk models. Overall, it is observed that the 0.5EY POTk models have a higher R^2 value (0.75–0.79) than the 12EY POTk models (0.62–0.69). A similar R^2 values of 0.68 for POT3, POT4 and POT5 models and a slightly smaller R^2 value for other POTk models (where k is less than 3) are observed. For a return period of 0.5EY, POT2 and POT3 models provide the highest R^2 value of 0.78 compared to all other POTk models. Among all the selected POTk RFFA models, POT3 provides the best performance in relation to R^2 value, considering both return periods of 12EY and 0.5EY. It is worth noting that the addition of the baseflow factor (M1) provides comparable results to models without considering the baseflow factor (M2).

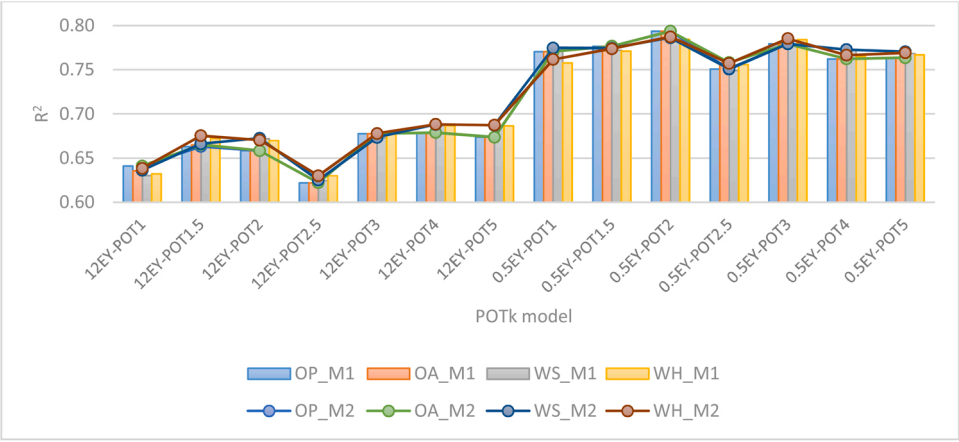


Fig. 3. R^2 plot for very frequent (12EY) and frequent (0.5EY) POTk RFFA models (Region 1).

RE_m (Fig. 4(a)(b)) and $RMSE_r$ (Fig. 4(c)(d)) are two other statistical measures applied in assessing the best performing POTk model. It is desirable for a model to have low RE_m and $RMSE_r$ values. For the return period of 12EY, POT4 and POT5 models provide the lowest RE_m and $RMSE_r$, 33% and 6.3%, respectively. In contrast, POT1.5 and POT2 models provide the lowest RE_m and $RMSE_r$ (32% and 6.5%). It is found that for very frequent floods, the smaller k of the POT flood series provides a higher RE_m and $RMSR_r$. Considering all

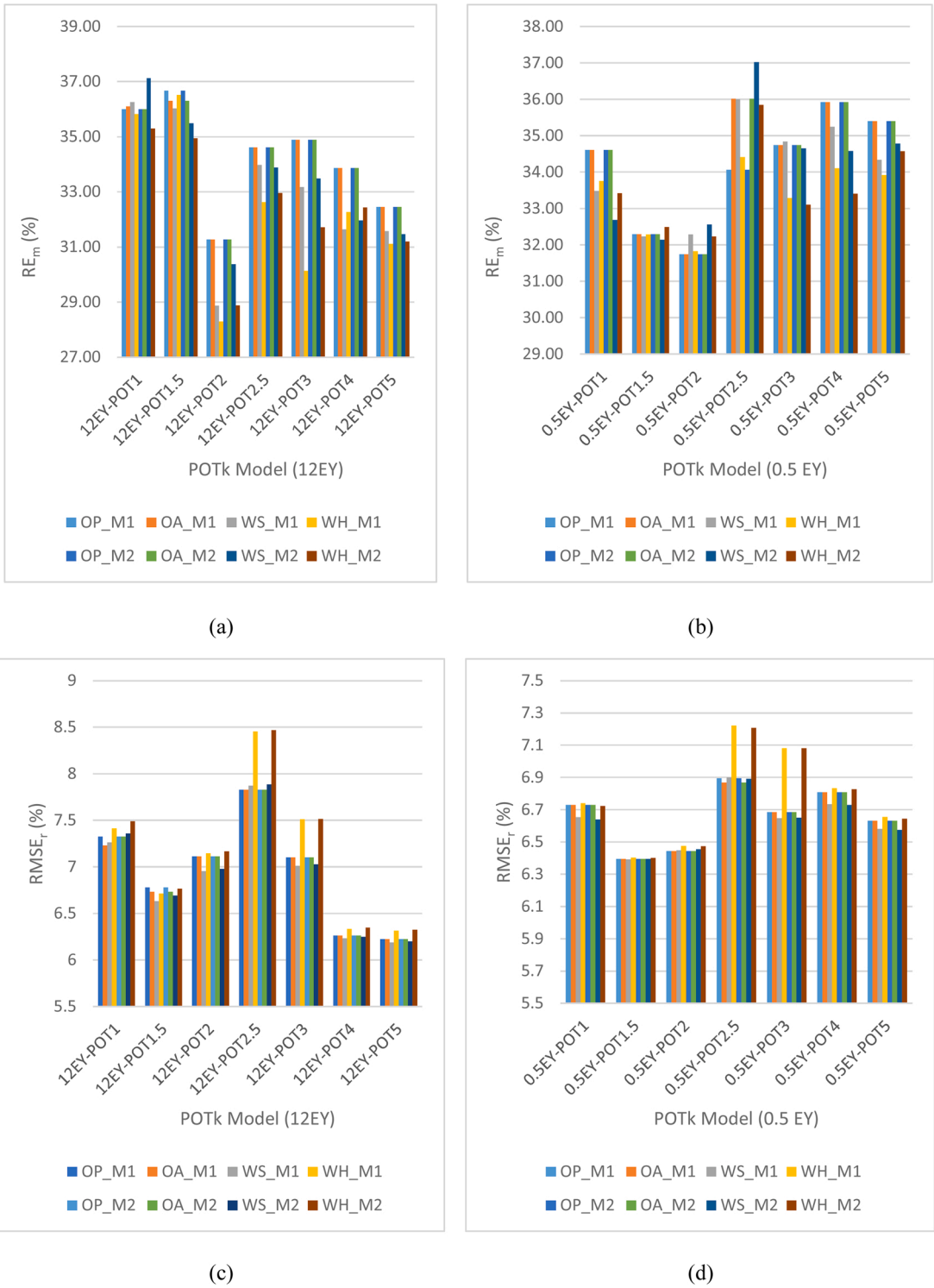


Fig. 4. (a) (b): Plot of median relative error (RE_m); (c) (d): Plot of root mean square error ($RMSE_r$) based on various POTk models. (Region 1).

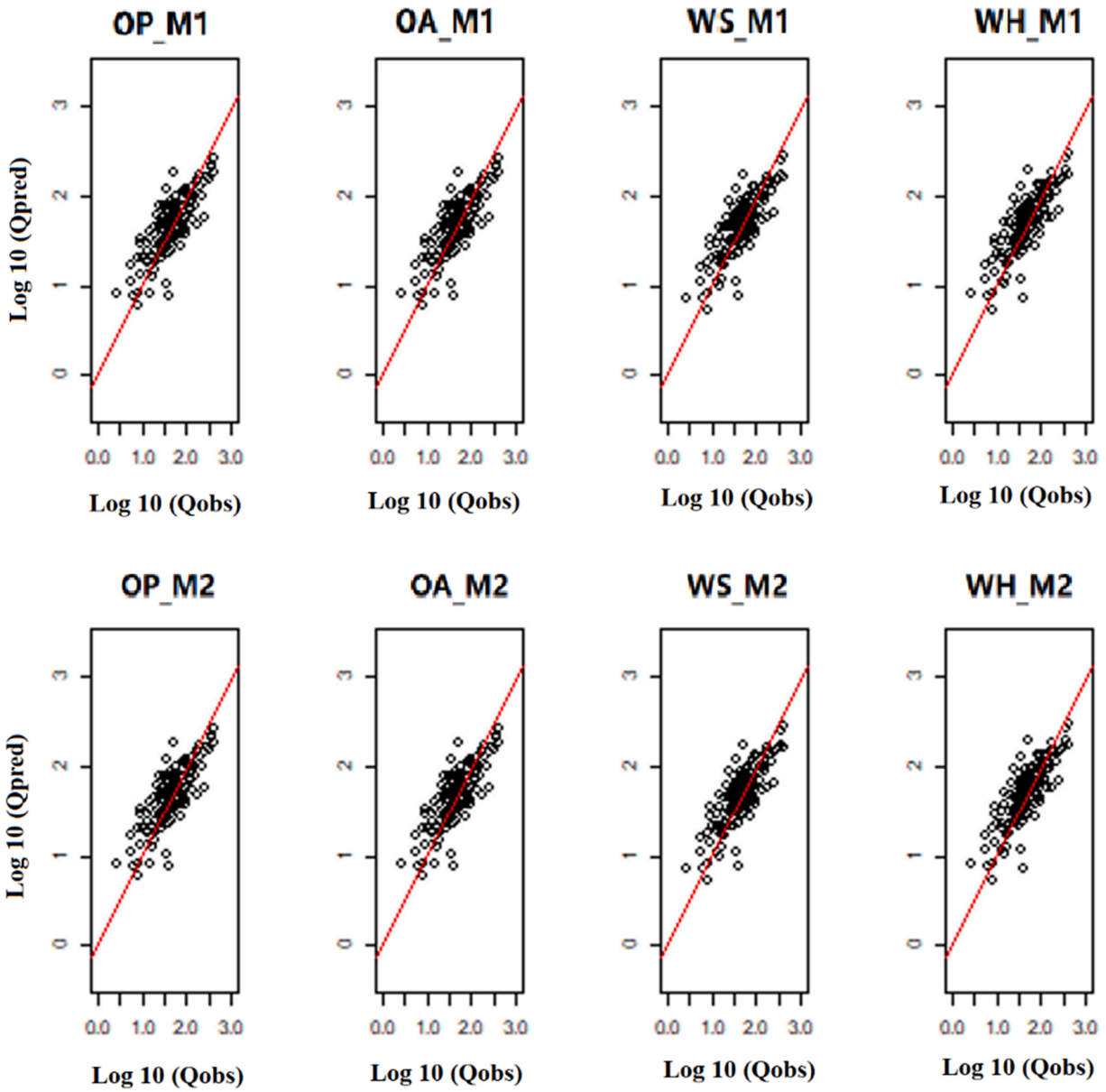


Fig. 5. Observed vs predicted flood quantiles (very frequent, 12EY) based on POT3 flood series in m3/s (Region 1).

the statistical measures applied, the POT3 model best balances the very frequent and frequent flood categories and is selected for further analysis.

4.2. Region 1

4.2.1. Very frequent flow, 12EY

The logarithm of predicted and observed flood quantiles (adopting POT3 flood series based on eight approaches using catchments in Region 1) are plotted in Fig. 5. It measures how well the very frequent flood can be estimated in ungauged catchments using MLR technique. Overall, it is shown a good agreement between the observed and predicted flood quantiles. Most of the catchments in Fig. 5 are within the narrow range along the 45-degree reference line. However, a slightly higher scatter is located above the reference line indicating some over-estimation.

Fig. 6 illustrates the standardised residual quantile-quantile (QQ) plot of very frequent (12EY) for all POT3 models in Region 1. These plots demonstrate a strong alignment to the normality assumption, although a small light-tailed behaviour is observed, highlighting some overestimation. OLS-based regression models have a slightly more over-estimation than WLS-based models. Overall, it is

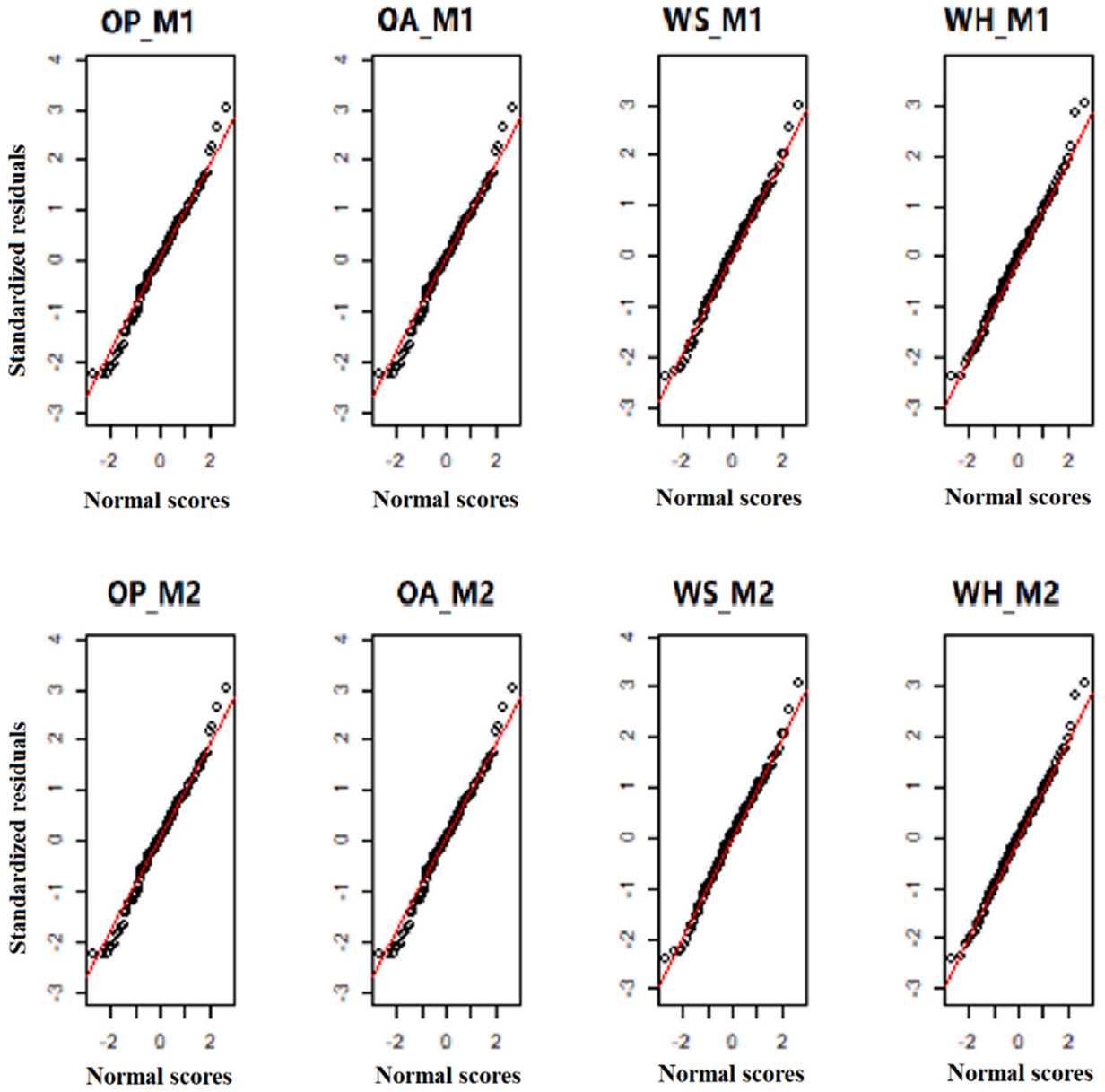


Fig. 6. Residual QQ-plot based on POT3 model (very frequent, 12EY, Region 1).

observed that more than 90% of the residuals fall within ± 2 normal scores, which indicates that the residuals closely follow a normal distribution.

The relative error (RE_r, Eq. 4) measures the difference between the predicted and observed flood quantile and is plotted in Fig. 7. A very similar boxplot between M1 and M2 models is observed using various regression techniques (i.e., OP_M1 and OP_M2). On the other hand, a slight difference is observed between OLS and WLS techniques (i.e., OP_M1 and WS_M1). Overall, the mean of RE_r ranges between 30% and 40%. This finding holds the preliminary finding discussed in the previous section, in which the overall fitting shows a slight tendency of over-estimation. Fig. 7 also illustrates a slight difference in the Inter Quantile Range (IQR) for different models.

$Ratio_r$ is another measure that indicates how well a predictive model performs to the observed data. If the $Ratio_r$ is close to 100%, it indicates a good agreement between estimated and observed flood quantiles. Fig. 7 indicates a similar tendency of over-estimation based on a mean of 120–135%, with an average of 100%. Combining the findings from Fig. 7 (Left and Right), it is clear that the OLS-based regression technique provides more accurate quantile estimates than the WLS for the return period of 12EY.

4.2.2. Frequent flow, 0.5EY

Fig. 8(a) plots the logarithm of the observed and predicted flood quantile for the return period of 0.5EY. It can be seen a narrower concentration along the 45-degree reference line, indicating a closer fit compared to 12EY. Although the scatter is smaller for the

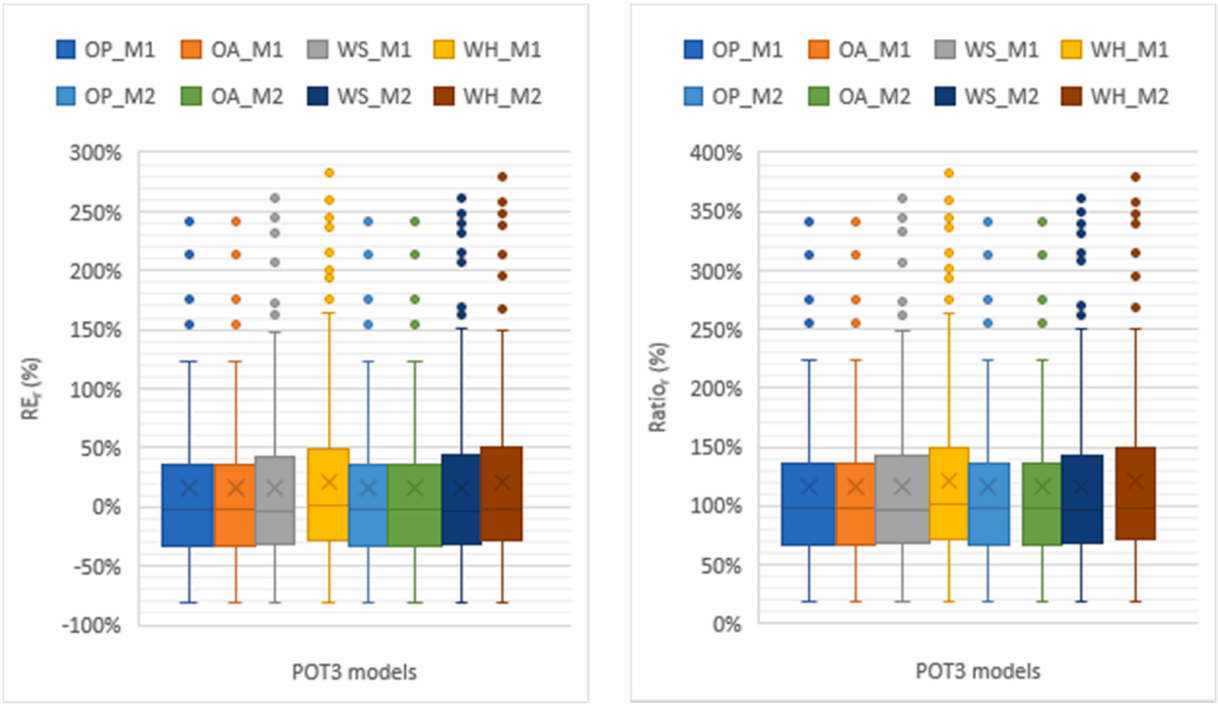


Fig. 7. Left: Boxplot of RE_r based on POT3 models (very frequent, 12EY); Right: Boxplot of $Ratio_r$ based on POT3 models ((very frequent, 12EY) (Region 1).

return period of 0.5EY (compared to 12EY), the over-estimation is still evident in the plot. Comparing various regression techniques, the difference in scatter is minor compared to the return period of 12EY. Fig. 8(b) plots the QQ plots for various regression models of 0.5EY return period. A light-tailed behaviour in the upper part is also observed. In general, it is reassuring that majority of the QQ plots fall in the range of $\pm$ 2.

Fig. 9 (Left) plots the RE_r of selected models for the return period of 0.5EY. The IQR is similar across all selected models, with an

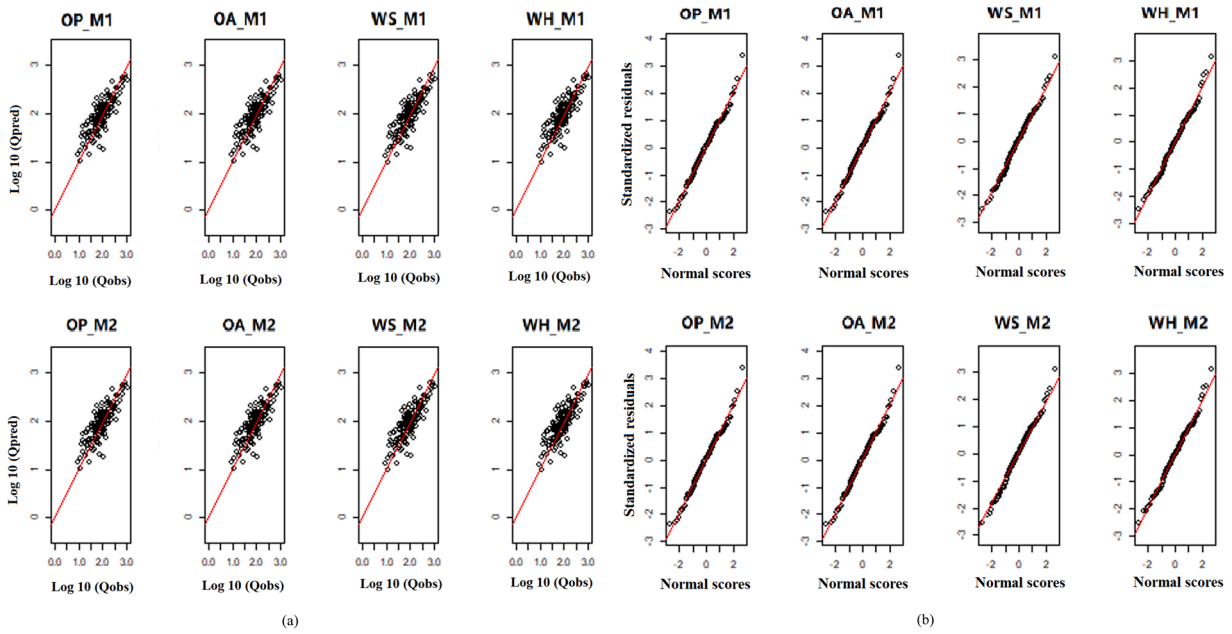


Fig. 8. (a): Observed vs predicted flood quantiles (frequent, 0.5EY) based on POT3 flood series in m3/s; (b): Residual QQ-plot based on POT3 model (frequent, 0.5EY. (Region 1).

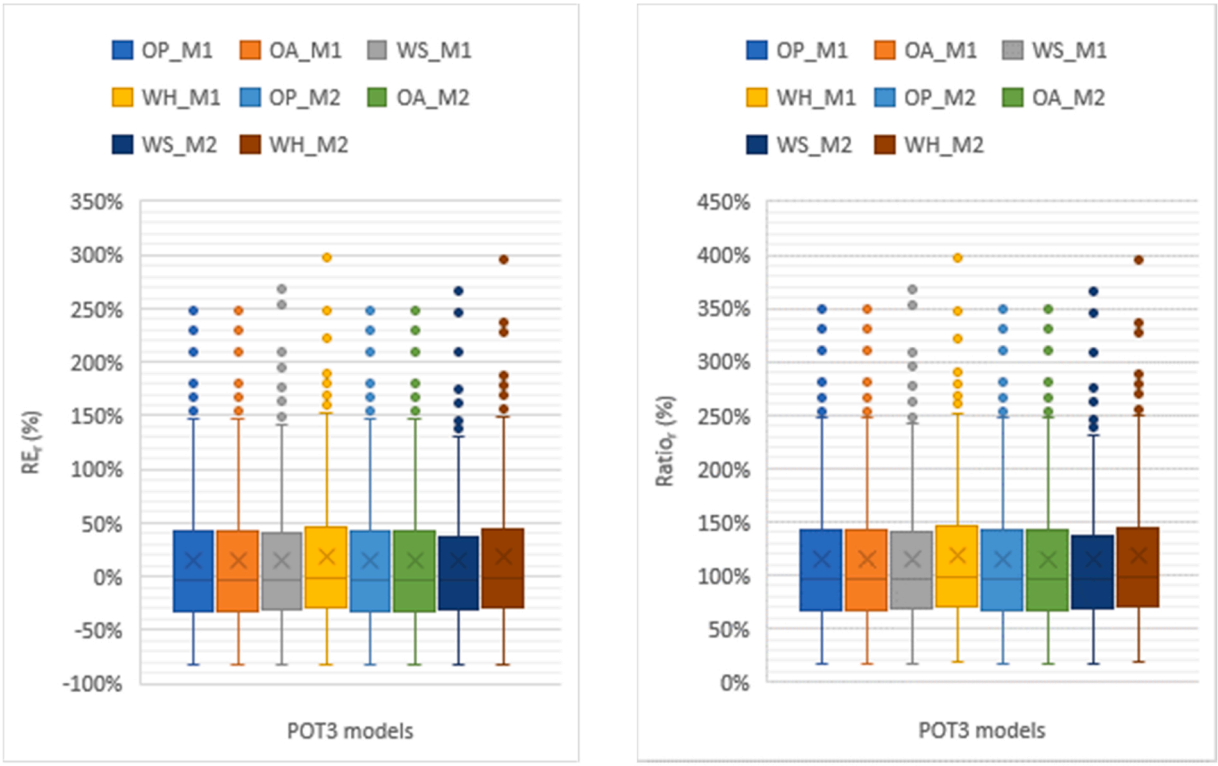


Fig. 9. Left: Boxplot of RE_r based on POT3 models (frequent, 0.5EY); Right: Boxplot of $Ratio_r$ based on POT3 models (frequent, 0.5EY) (Region 1).

average of 20–30% and a median just above 0%. The WLS-based model with the addition of baseflow factors (M1) presents a slightly better agreement (smaller IQR) among others. Fig. 9 (Right) plots the $Ratio_r$ with a mean of 120–130% and a median of 98–102%.

4.2.3. Cross-validation results

Split sampling with 100 repeats is firstly applied (80% training and 20% testing) for each selected model. Fig. 10 plots RE_m (Eq. 3) for 12EY (Left) and 0.5EY (Right). For the return period of 12EY, it can be seen that the median RE_m values are similar for the six selected approaches, which range from 35% to 37%. A similar result is also observed for 0.5EY. Among different regression techniques, the boxplots of OLS-based techniques provide a smaller IQR for both the 12EY and 0.5 EY return periods. Comparing models M1 and M2, a small reduction in IQR is found for M2.

Table 2 tabulates the median of RE_m using the LOO technique for all the selected return periods in this study. It can be seen that the variation in median RE_m is minor among all the selected regression models. Overall, the RE_m ranges from a minimum of 31.7% to a maximum of 38%. Table 3 presents the median of $Ratio_r$. It is observed that the $Ratio_r$ between various models is within the range of $\pm$ 4%, with a minimum of 94% and a maximum of 102%.

Considering the results and plots from various statistical indices, OLS based regression model (based on p-value of 0.1 and exclusion of baseflow factor, OP_M2) is selected for construction of prediction equations. Table 4 presents prediction equations and associated coefficients for selected return periods. The independent variable design rainfall intensity (I), shown in Table 4, is corresponded to the associated return period and 6 h duration (e.g., for 12EY, I corresponds to 12EY return period and 6 h duration and for 6EY, I corresponds to 6EY return period and 6 h duration). For the very frequent floods, A, SDEN, and I are found to be significant independent variables. For frequent floods, mean annual rainfall (MAR) is added with A, SDEN and I as significant independent variables.

Fig. 11 illustrate the spatial distribution of the absolute RE_r values of the individual catchments for return period of 12EY (Left) and 0.5EY (Right). A certain degree of spatial coherence can be seen in the lower region of the coastline of south-eastern Australia. With the increased distance from the coastline, the number of catchments showing larger absolute RE_r (>50%) increases. Overall, a good predictive performance based on absolute RE_r is found for 12EY and 0.5EY models (POT3_R1_OP_M2).

It is found that 38.4% of catchments (out of 151 in Region 1) have absolute RE_r values less than 25% for the return period of 12EY (this is 37.1% for the return period of 0.5EY). Table 5 provides the individual and cumulative percentages of catchments with a range of absolute RE_r values. Overall, the selected model (POT3_R1_OP_M2) provides a relatively accurate estimate for very frequent to frequent floods, with over 70.9% and 72.8% of catchments for 12 EY and 0.5EY models respectively have absolute RE_r values less than or equal to 50%.

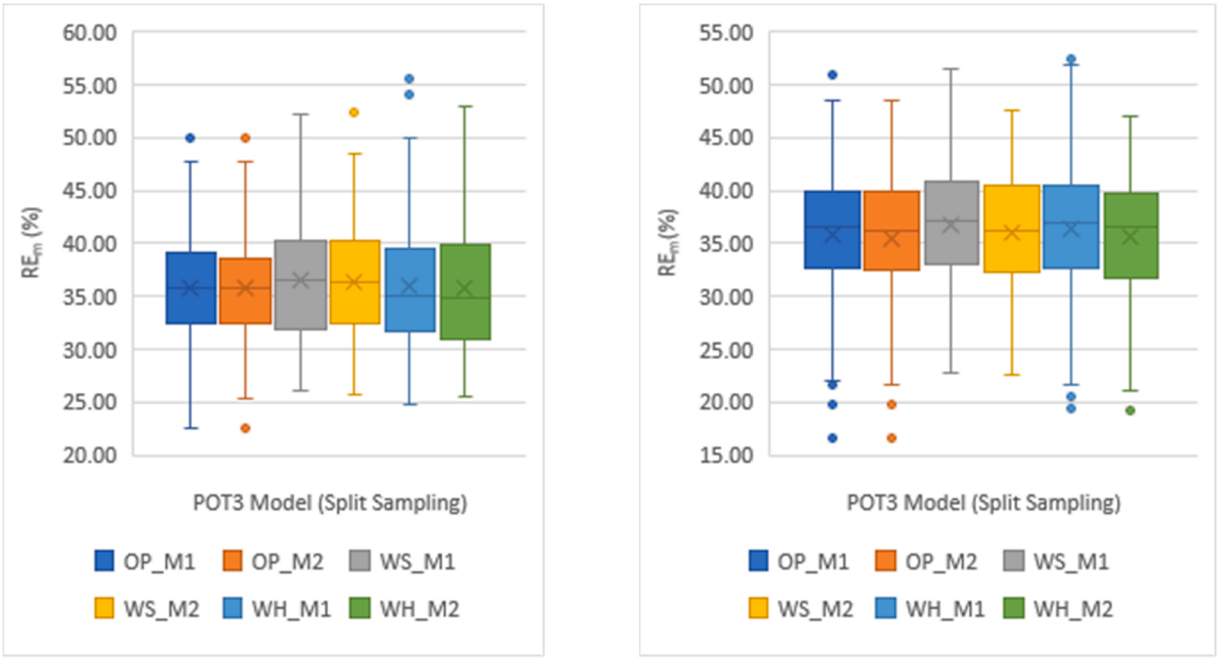


Fig. 10. Left: Boxplot of RE_m for POT3 models (very frequent, 12EY); Right: Boxplot of RE_m for POT3 models (frequent, 0.5EY). (Region 1).

Table 2
Median of RE_m based on LOO cross-validation (Region 1).

Median of RE_m (%)									
	12EY	6EY	4EY	3EY	2EY	1EY	0.5EY	0.2EY	10ARI
OP_M1	35.1	35.6	34.4	33.9	34.6	33.8	35.8	32.6	34.8
OP _M2	35.1	35.6	34.4	33.9	34.6	33.8	35.8	32.6	34.8
WS_M1	35.5	35.9	37.3	38.0	36.4	35.9	36.2	32.8	34.3
WS_M2	35.5	35.8	37.2	37.9	36.4	36.4	36.4	31.7	35.4
WH_M1	33.0	32.9	35.3	35.7	33.6	34.8	34.7	33.0	34.5
WH_M2	33.4	33.3	35.3	35.1	33.7	36.1	35.8	32.5	34.7

4.3. Region 2

R^2 values for Region 2 are plotted in Fig. S1 (supplementary section), which illustrates similar values of R^2 for return periods of 12EY and 0.5EY. Overall, R^2 (0.7–0.8) values for Region 2 are higher than for Region 1. Considering RE_m and $RMSE_r$, the POT3 model is the best performing POT time series based on balancing the score between very frequent (12EY) and frequent (0.5EY) return periods as can be seen in Fig. S2.

Considering the scatter plot between the logarithm of Q_{obs} and Q_{pred}, POT3 models show a slightly larger range of variation along the 45-degree line in Fig. S3 Left. QQ plot (Fig. S3 Right) shows that more than 90% of the scatter falls between the range of ± 2. RE_r and $Ratio_r$ for the return period of 12EY are plotted in Fig. S4 Left and S4 Right, respectively. There is some notable overestimation in Region 2 similar to Region 1. It is also observed that the median of each box plot in Fig. S4 is close to 0% and 100%, respectively, which indicates an unbiased estimate based on the POT3 model. Similar results for frequent flood estimates are found based on Fig. S5 and

Table 3
Median of $Ratio_r$ based on LOO cross-validation (Region 1).

Median of $Ratio_r$ (%)									
	12EY	6EY	4EY	3EY	2EY	1EY	0.5EY	0.2EY	10ARI
OP_M1	98	96	97	96	95	96	97	99	100
OP _M2	98	96	97	96	95	96	97	99	100
WS_M1	96	96	96	94	95	97	96	99	100
WS_M2	96	97	98	97	96	97	97	98	101
WH_M1	101	100	101	100	100	100	99	101	99
WH_M2	98	100	101	100	100	100	98	100	102

Table 4
Prediction equations for Region 1 (POT3_R1_OP_M2): $\log_{10} Q_i = b_0 + b_1 * \log_{10} A + b_2 * \log_{10} SDEN + b_3 * \log_{10} I + b_4 * \log_{10} MAR$.

Flood Categories	Q_i	b_0	b_1	b_2	b_3	b_4
Very Frequent	12EY	-0.6	0.65	0.73	1.36	-
	6EY	-0.68	0.65	0.69	1.42	-
	4EY	-0.79	0.64	0.63	1.48	-
	3EY	-0.87	0.64	0.6	1.52	-
	2EY	-0.95	0.64	0.57	1.58	-
Frequent	1EY	-1.09	0.63	0.53	1.73	-
	0.5EY	-0.34	0.62	0.5	2.11	-0.36
	0.2EY	-0.29	0.61	0.48	2.17	-0.41
	10ARI	-0.28	0.61	0.48	2.13	-0.4

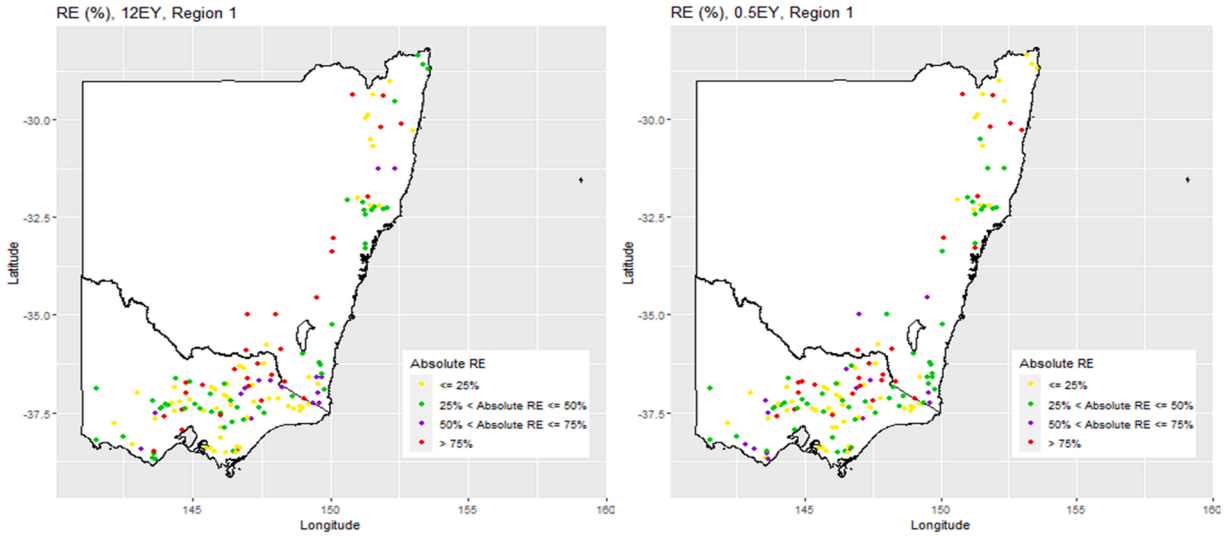


Fig. 11. Left: Spatial distribution of absolute RE_r values for 12EY model. Right: Spatial distribution of RE_r values for 0.5EY model (POT3_R1_OP_M2) (Region 1).

Fig. S6. It is found that OLS-based models excluding the baseflow factor provide the smallest IQR and smallest mean for both RE_r and $Ratio_r$.

The POT3 models are cross validated using split sample validation technique, and the results are illustrated in Fig. 12. Smallest IQR is observed for the OLS based model (OP_M2) for both return periods of 12EY and 0.5EY. The median of RE_m is 45% for the return period of 12EY and 48% for 0.5EY. Overall, OLS-based models outperform the WLS-based models, which have a wider IQR and a higher median of RE_m. LOO is also conducted for Region 2 and results are tabulated in Table S1 (median of RE_m) and Table S2 (median of $Ratio_r$).

The developed OLS based (POT3_R2_OP_M2) prediction equations for Region 2 are provided in Table 6. It is observed that the group of independent variables (based on 5% significance level) are similar to Region 1, except that MAR is not selected for Region 2.

The spatial distribution of the absolute RE_r values of the selected model, POT3_R2_OP_M2, is shown in Fig. 13 Left (12EY) and Fig. 13 Right (0.5EY). It is noted that few catchments have larger absolute RE_r values (>75%), which are located along the upper part of the south-eastern coastline for the return period of 12EY. In contrast, the total number of catchments having larger absolute RE_r values (>75%) is smaller for the return period of 0.5EY. The distribution of absolute RE_r values in Fig. 13 (for Region 2) and in Fig. 11 (Region 1) shows a broad agreement.

Table 5
Percentage of catchments having a range of absolute RE_r values (Region 1).

	12EY_ POT3_R1_OP_M2			0.5EY_ POT3_R1_OP_M2		
Range of REm (%)	Count of catchments	Individual percentage of catchments	Cumulative percentage of catchments	Count of catchments	Individual percentage of catchments	Cumulative percentage of catchments
< 25%	58	38.4%	38.4%	56	37.1%	37.1%
25–50%	49	32.5%	70.9%	54	35.8%	72.8%
50–75%	15	9.9%	80.8%	14	9.3%	82.1%
> 75%	29	19.2%	NA	27	17.9%	NA

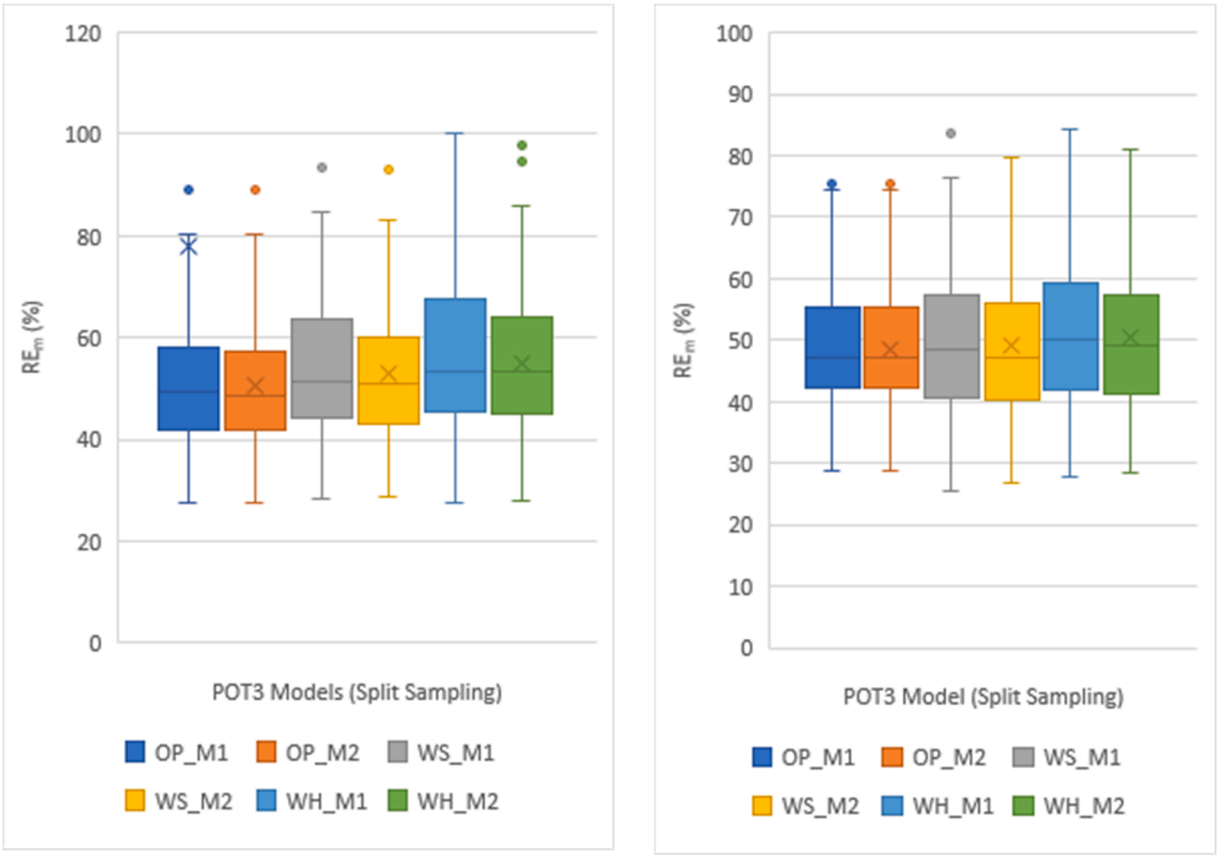


Fig. 12. Left: Boxplot of RE_m for POT3 models (very frequent, 12EY); Right: Boxplot of RE_m for POT3 models (frequent, 0.5EY) (Region 2).

4.4. Region 3

Fig. S7 plots R^2 values of POTk models for Region 3. Overall, R^2 values of very frequent flood (0.45–0.65) is significantly lower than frequent flood (0.6–0.68) among selected POTk models. Selected POTk models are also evaluated based on RE_m and $RMSE_r$, as illustrated in Fig. S8 Left and Right. Considering all the statistics from various POTk models in Region 3, the POT3 model performs the best while balancing the categories of the very frequent and frequent floods.

POT3 models for very frequent floods are further assessed through the scatter plot (Fig. S9 Left) and the QQ plot (Fig. S9 Right). A narrower concentration along the reference line can be seen for OA_M1 and OA_M2 models as compared to OP_M1 and OP_M2 models. Statistics of RE_m and $Ratio_r$ are evaluated and plotted in Fig. S10 Left and Right, respectively. It is observed that the smallest IQR is found for OA_M2, which is different to the findings of Region 1 and Region 2. Similar approaches are applied to frequent floods (0.5EY) as illustrated in Fig. S11 and Fig. S12. OA_M2 are the second-best performing models for the frequent flood. Considering both categories of very frequent and frequent floods, OA-based models are selected for further investigation while OP models are disregarded.

The split sample validation method is applied for Region 3 and results are presented in Fig. S12. It is observed that the OA_M2 model provides the smallest IQR for both very frequent and frequent floods (22%−35% and 28%−38%, respectively). It is also

Table 6
Prediction equations for Region 2 (POT3_R2_OP_M2): $\log_{10} Q_i = b_0 + b_1 * \log_{10} A + b_2 * \log_{10} SDEN + b_3 * \log_{10} I$.

Flood Categories	Q_i	b_0	b_1	b_2	b_3
Very Frequent	12EY	-0.90	0.65	0.73	1.36
	6EY	-0.98	0.65	0.69	1.42
	4EY	-1.09	0.64	0.63	1.48
	3EY	-1.15	0.64	0.6	1.52
	2EY	-1.23	0.64	0.57	1.58
Frequent	1EY	-1.35	0.64	0.29	2.20
	0.5EY	-1.47	0.63	0.29	2.26
	0.2EY	-1.38	0.62	0.35	2.09
	10ARI	-1.24	0.62	0.41	1.92

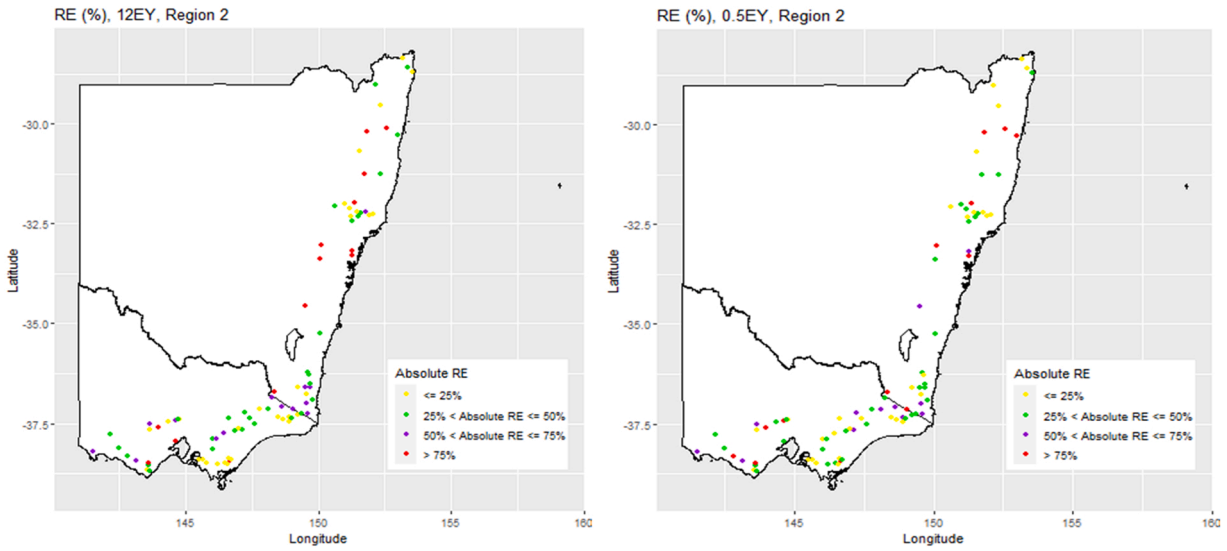


Fig. 13. Left: Spatial distribution of absolute REr values for the return period of 12EY. Right: Spatial distribution of absolute REr values for the return period of 0.5EY. (Region 2).

Table 7
Prediction equations for Region 3 (POT3_R3_OA_M2): $\log_{10} Q_i = b_0 + b_1 * \log_{10} A + b_2 * \log_{10} SF + b_3 * \log_{10} SDEN + b_4 * \log_{10} S1085 + b_5 * \log_{10} I$.

Flood Categories	Q_i	b_0	b_1	b_2	b_3	b_4	b_5
Very Frequent	12EY	-0.73	0.84	0.29	0.64	0.16	-
	6EY	-0.73	0.84	0.29	0.64	0.16	-
	4EY	-0.72	0.84	0.29	0.64	0.16	-
	3EY	-0.70	0.84	0.30	0.65	0.15	-
	2EY	-0.65	0.84	0.30	0.67	0.13	-
Frequent	1EY	-0.37	0.81	0.26	0.72	-	-
	0.5EY	-0.27	0.81	0.26	0.73	-	-
	0.2EY	0.64	0.80	-0.26	0.74	-	-
	10ARI	0.40	0.73	-0.41	0.59	-	1.05

observed that the median of RE_m based on OA_M2 is the smallest among all the selected models for very frequent floods. LOO cross-validation is also applied, and the results are shown in Table S3 and Table S4. Considering the evaluation statistics, POT3_R3_OA_M2 model performs the best. Table 7 presents the developed prediction equations (Fig. 14).

Spatial distribution of the absolute REr values for POT3_R3_OA_M2 model are plotted in Fig. 15 Left (12EY) and Fig. 15 Right (0.5EY). It can be seen that the majority of the catchments in the lower part of Region 3 are performing quite well (absolute RE_r <50%) for 12EY, and catchments showing significant variation (absolute RE_r >75%) are located far away from the coastline. Fig. 15 Right illustrates a similar result.

4.5. Comparison of three regional models (Region 1, Region 2 and Region 3)

In this section, a comparison is made to evaluate the performance of POT3-based RFFA models across three regions. The count of catchments falling in different ranges of absolute REr values are shown in Table 8. It can be seen that Region 3 performs the best (having the highest percentage of catchments having a smaller absolute REr values) followed by Region 1 and Region 2. Region 3 has smaller degree of heterogeneity compared to Region 1 and Region 2, which could be the reason for smaller prediction error for Region 3. Further study is needed to understand the differences in the performances of the prediction models across the three regions.

Table 9 shows the significant (at 5% level) independent variables for the RFFA models for three different regions. The variables A and SDEN are found to be significant for all the three regions for both very frequent and frequent floods. It is found to be significant for Region 1 and Region 2 and for frequent floods in Region 3. SF is significant for only Region 3. MAR is significant for Region 1 (for frequent floods) and S1085 is significant for Region 3 (for very frequent floods).

4.6. Recommended regional prediction equations for application

Considering different regional models (Region 1, 2 and 3), it is clear that Region 2 model has the lowest prediction accuracy and hence these models are not recommended for practical application. For simplicity, Region 1 models (shown in Table 4) should be applied to the whole study area since these provide quite accurate flood predictions (with absolute median relative error (RE_m) in the

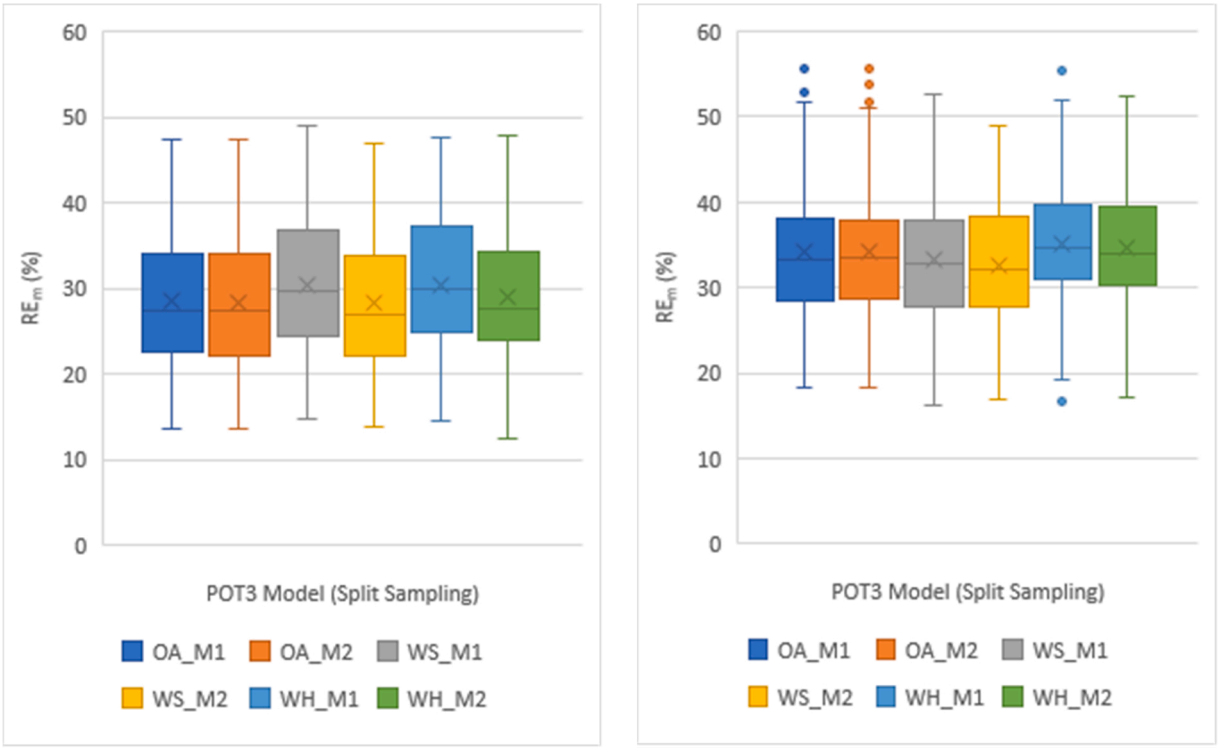


Fig. 14. Left: Boxplot of RE_m values for POT3 models (very frequent, 12EY); Right: Boxplot of RE_m values for POT3 models (frequent, 0.5EY) (Region 3).

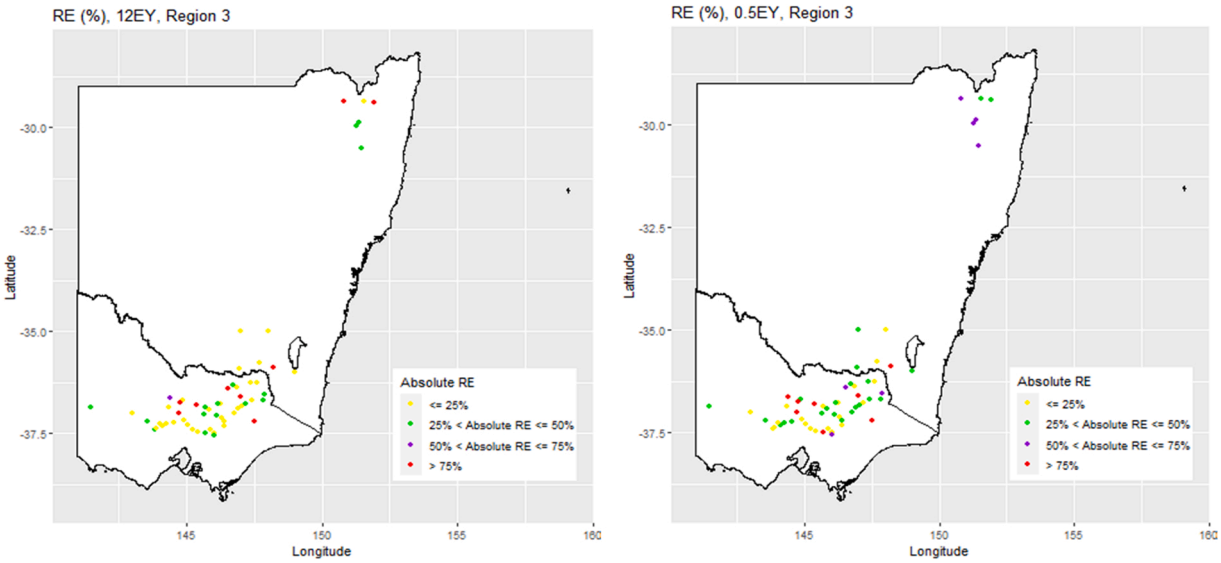


Fig. 15. Left: Spatial distribution of absolute REr values for the return period of 12EY. Right: Spatial distribution of absolute REr values for the return period of 0.5EY. (POT3_R3_OA_M2) (Region 3).

range of 33–36%). These results are comparable to Ali and Rahman (2022) who reported a RE_m value of 31–34% in south-east Australia for 2 and 5 years return periods based on kriging based RFFA techniques using annual maximum flood model. In another study for south-east Australia, Zalnezhad et al. (2022) reported RE_m values of 42% and 33% for 2 and 5 years return periods, respectively using artificial neural networks (ANN) based RFFA technique. We developed prediction equations for 100-year ARI for both the AM and POT cases and we found that the POT-based models provided more accurate results than AM based models. This

Table 8
Count of catchments based of range of absolute RE_r values.

	12EY_POT3			0.5EY_POT3		
Range of Absolute RE_r	Region 1, count of catchments	Region 2, count of catchments	Region 3, count of catchments	Region 1, count of catchments	Region 2, count of catchments	Region 3, count of catchments
Up to 25%	38%	35%	58%	37%	34%	37%
Up to 50%	71%	69%	85%	73%	72%	77%
Up to 75%	81%	84%	88%	82%	85%	90%

Table 9
Comparison of significant independent variables.

Flood categories	Regions	Significant independent variables			
Very Frequent	Region 1	A	SDEN	I_i	
	Region 2 (Coastal)	A	SDEN	I_i	
	Region 3 (Inland)	A	SDEN	SF	S1085
Frequent	Region 1	A	SDEN	I_i	MAR
	Region 2 (Coastal)	A	SDEN	I_i	
	Region 3 (Inland)	A	SDEN	SF	I_i

indicates that in future POT data can be used to develop regional prediction equations for very frequent to 100-year ARI ranges for Australia.

5. Conclusion

The method adopted and developed in this study focuses on the flood quantile estimation in ungauged catchments in south-east Australia in the range of very frequent to frequent floods (12, 6, 4, 3, 2, 1, 0.5, 0.2EY and 10-year ARI). Considering the advantages of POT-based approach, various POTk at-site estimates (average events per year, k = 1, 1.5, 2, 2.5, 3, 4, 5) are modelled and evaluated. Based on 151 selected catchments from south-east Australia, it has been found that POT3 based model provides the best performance while balancing the categories of very frequent to frequent floods. This study also compares the performance of OLS and WLS-based regression analyses. It has been found that the OLS-based regression technique provides the least relative error than the WLS method. Three different regions are considered in this study. In this regard, no significant improvement is found if the coastal region alone (Region 2) is considered. However, a considerable improvement in prediction performance is found by adopting the inland region alone (Region 3). For simplicity, the prediction equations developed for the whole NSW (Region 1) are recommended for practical application. These prediction equations require four predictor variables (catchment area, design rainfall intensity, mean annual rainfall and stream density), which are readily available. Based on independent testing, these predictions show a median relative error values in the range of 32–36%, which are smaller than Australian Rainfall and Runoff recommended RFFA models (49–59%) (Rahman et al., 2019). This study bridges the gap in the current edition of Australian Rainfall and Runoff, which does not have RFFA method in the range of very frequent floods. The findings of this study will promote adopting the POT-based RFFA approach in practice in south-east Australia. Further study should focus on the impacts of climate change on POT-based RFFA approach and uncertainty analysis similar to Durocher, Burn and Ashkar (2019). Also, to reduce parameter uncertainty in regression equations, generalised least squares regression approach should be adopted similar to Madsen and Rosbjerg (1997) and Haddad and Rahman (2012). Further study should consider developing RFFA models from very frequent to 100-year flood quantiles based on POT modelling for whole Australia.

Interactive map

Boxplot of relative error (RE) of the peaks over threshold based regional flood frequency analysis model for south-east Australia: very frequent 12 events per year floods (left) and frequent 0.5 events per year floods (right).

CRediT authorship contribution statement

X Pan: Data analysis, drafting of manuscript; **A Rahman:** Conceptualization, Supervision, Editing; **K Haddad:** Supervision, Editing; **T Ouarda:** Supervision, Editing; **A Sharma:** Supervision, editing.

Declaration of Competing Interest

Authors declare that there is no conflict of interest, and no funding was received for this study.

Data Availability

The authors do not have permission to share data.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejrh.2023.101407.

References

Acosta, L.A., Eugenio, E.A., Macandog, P.B.M., Magcale-Macandog, D.B., Lin, E.K.H., Abucay, E.R., Cura, A.L., Primavera, M.G., 2016. Loss and damage from typhoon-induced floods and landslides in the Philippines: community perceptions on climate impacts and adaptation options. Int. J. Glob. Warm. 9 (1), 33–65.

Ahammed, F., Hewa, G.A., Argue, J.R., 2014. Variability of annual daily maximum rainfall of Dhaka, Bangladesh. Atmos. Res. 137, 176–182.

Ali, S., Rahman, A., 2022. Development of a kriging based regional flood frequency analysis technique for south-east Australia. Nat. Hazards. ⟨https://link.springer.com/article/10.1007/s11069-022-05488-4⟩.

Ammar, M.E., Gharib, A., Islam, Z., Davies, E.G.R., Seneka, M., Faramarzi, M., 2020. Future floods using hydroclimatic simulations and peaks over threshold: An alternative to nonstationary analysis inferred from trend tests. Adv. Water Resour. 136, 103463 (Article).

Ball, J.E., Babister, M.K., Nathan, R., Weinmann, P.E., Weeks, W., Retallick, M., Testoni, I. , 2019. Australian Rainfall and Runoff - A Guide to Flood Estimation. Commonwealth of Australia.

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), 3369–3381.

Bernardara, P., Andreewsky, M., Benoit, M., 2011. Application of regional frequency analysis to the estimation of extreme storm surges. J. Geophys. Res.: Oceans 116 (2), C02008 (Article).

Burn, D.H., Ouarda, T.B.M.J., Shu, C., 2007. Estimation of Extreme Flow Quantiles and Quantile Uncertainty for Ungauged Catchments. IAHS-AISH Publication,.

Cunnane, C., 1973. A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. J. Hydrol. 18 (3), 257–271.

Durocher, M., Chebana, F., Ouarda, T.B.M.J., 2016. On the prediction of extreme flood quantiles at ungauged locations with spatial copula. J. Hydrol. 533, 523–532.

Durocher, M., Mostofi Zadeh, S., Burn, D.H., Ashkar, F., 2018. Comparison of automatic procedures for selecting flood peaks over threshold based on goodness-of-fit tests. Hydrol. Process. 32 (18), 2874–2887.

Durocher, M., Burn, D.H., Ashkar, F., 2019. Comparison of estimation methods for a nonstationary index-flood model in flood frequency analysis using peaks over threshold. Water Resour. Res. 55 (11), 9398–9416.

Haddad, K., Rahman, A., 2011. Selection of the best fit flood frequency distribution and parameter estimation procedure: a case study for Tasmania in Australia. Stoch. Environ. Res. Risk Assess. 25 (3), 415–428.

Haddad, K., Rahman, A., 2012. Regional flood frequency analysis in eastern Australia: Bayesian GLS regression-based methods within fixed region and ROI framework – quantile regression vs. parameter regression technique. J. Hydrol. 430–431, 142–161.

Haddad, K., Rahman, A., Green, J., 2011. Design rainfall estimation in Australia: a case study using L moments and Generalized Least Squares Regression. Stoch. Environ. Res. Risk Assess. 25 (6), 815–825.

Haddad, K., Rahman, A., Stedinger, J.R., 2012. Regional flood frequency analysis using Bayesian generalized least squares: a comparison between quantile and parameter regression techniques ([). Hydrol. Process. 26 (7), 1008–1021.

Hosking, J.R.M., Wallis, J.R., 1993. Some statistics useful in regional frequency analysis. Water Resour. Res. 29 (2), 271–281.

Hossain, I., Imteaz, M.A., Khastagir, A., 2021. Effects of estimation techniques on generalised extreme value distribution (GEVD) parameters and their spatio-temporal variations. Stoch. Environ. Res. Risk Assess. 35, 2303–2312.

Ishak, E.H., Rahman, A., 2015. Detection of changes in flood data in Victoria, Australia from 1975 to 2011. Hydrol. Res. 46 (5), 763–776.

Ishak, E.H., Rahman, A., Westra, S., Sharma, A., Kuczera, G., 2013. Evaluating the non-stationarity of Australian annual maximum flood. J. Hydrol. 494, 134–145.

Jarajapu, D.C., Rathinasamy, M., Agarwal, A., Bronstert, A., 2022. Design flood estimation using extreme gradient boosting-based on Bayesian optimization. J. Hydrol. 613, 128341.

Karim, F., Hasan, M., Marvanek, S., 2017. Evaluating annual maximum and partial duration series for estimating frequency of small magnitude floods. Water (Switz.) 9 (7), 481 (Article).

Kiran, K.G., Srinivas, V.V., 2021a. A mahalanobis distance-based automatic threshold selection method for peaks over threshold model. Water Resour. Res. 57 (1) e2020WR027534.

Kumar, M., Sharif, M., Ahmed, S., 2020. Flood estimation at Hathnikund Barrage, River Yamuna, India using the peak-over-threshold method. ISH J. Hydraul. Eng. 26 (3), 291–300.

Lang, M., Ouarda, T.B.M.J., Bobée, B., 1999. Towards operational guidelines for over-threshold modeling. J. Hydrol. 225 (3), 103–117.

Langousis, A., Mamalakis, A., Puliga, M., Deidda, R., 2016. Threshold detection for the generalized Pareto distribution: review of representative methods and application to the NOAA NCDC daily rainfall database. Water Resour. Res. 52 (4), 2659–2681.

Liang, B., Shao, Z., Li, H., Shao, M., Lee, D., 2019. An automated threshold selection method based on the characteristic of extrapolated significant wave heights. Coast. Eng. 144, 22–32.

Madsen, H., Rosbjerg, D., 1997. Generalized least squares and empirical Bayes estimation in regional partial duration series index-flood modeling. Water Resour. Res. 33 (4), 771–781.

Madsen, H., Rasmussen, P.F., Rosbjerg, D., 1997a. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 1. At-site modeling. Water Resour. Res. 33 (4), 747–757.

Masan, M.N., Hewa, G.A. , 2020. Developing regional flood frequency models for Victoria, Australia by using Index Flood (IF) approach. 18th World IMACS Congress and MODSIM 2009 - International Congress on Modelling and Simulation: Interfacing Modelling and Simulation with Mathematical and Computational Sciences, Proceedings.

Metzger, A., Marra, F., Smith, J.A., Morin, E., 2020. Flood frequency estimation and uncertainty in arid/semi-arid regions. J. Hydrol. 590, 125254.

Mostofi Zadeh, S., Durocher, M., Burn, D.H., Ashkar, F., 2019. Pooled flood frequency analysis: a comparison based on peaks-over-threshold and annual maximum series. Hydrol. Sci. J. 64 (2), 121–136.

Northrop, P.J., Jonathan, P., 2011. Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. Environmetrics 22 (7), 799–809.

Pan, X., Rahman, A., 2022. Comparison of annual maximum and peaks-over-threshold methods with automated threshold selection in flood frequency analysis: a case study for Australia. Nat. Hazards 111 (2), 1219–1244.

Pan, X., Rahman, A., Haddad, K., Ouarda, T., 2022. Peaks-over-threshold model in flood frequency analysis: a scoping review. Stoch. Environ. Res. Risk Assess. 36 (9), 2419–2435.

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., Charron, C., Ouarda, T.B.M.J., Chebana, F., 2018. Development of regional flood frequency analysis techniques using generalized additive models for Australia. Stoch. Environ. Res. Risk Assess. 32 (1), 123–139.
Rahman, A., Haddad, K., Kuczera, G., Weinmann, P.E., 2019. Regional flood methods. In: Ball, et al. (Eds.), Australian Rainfall & Runoff, Chapter 3, Book 3. Commonwealth of Australia.
Rahman, A.S., Rahman, A., 2020. Application of principal component analysis and cluster analysis in regional flood frequency analysis: a case study in New South Wales, Australia. Water 12 (3), 781.
Rahman, A.S., Khan, Z., Rahman, A., 2020. Application of independent component analysis in regional flood frequency analysis: comparison between quantile regression and parameter regression techniques. J. Hydrol. 581, 124372.
Sccarrott, C., Macdonald, A., 2012. A review of extreme value threshold estimation and uncertainty quantification. Stat. J. 10, 33–60.
Shu, C., Ouarda, T.B.M.J., 2007. Flood frequency analysis at ungauged sites using artificial neural networks in canonical correlation analysis physiographic space. Water Resour. Res. 43 (7), W07438 (Article).
Stedinger, J.R., Tasker, G.D., 1985. Regional hydrologic analysis: 1. ordinary, weighted, and generalized least squares compared. Water Resour. Res. 21 (9), 1421–1432.
Thompson, P., Cai, Y., Reeve, D., Stander, J., 2009. Automated threshold selection methods for extreme wave analysis. Coast. Eng. 56 (10), 1013–1021.
Walpita-Gamage, S.H.P., Hewa, G.A., Subhashini, W.H.C., Daniell, T.M., Kemp, D. , 2020. Modelling the extreme floods of South Australian catchments. 18th World IMACS Congress and MODSIM 2009 - International Congress on Modelling and Simulation: Interfacing Modelling and Simulation with Mathematical and Computational Sciences, Proceedings.
Yilmaz, A.G., Imteaz, M.A., Perera, B.J.C., 2017. Investigation of non-stationarity of extreme rainfalls and spatial variability of rainfall intensity–frequency–duration relationships: a case study of Victoria, Australia. Int. J. Climatol. 37 (1), 430–442.
Zalnezhad, A., Rahman, A., Nasiri, N., Vafakhah, M., Samali, B., Ahamed, F., 2022. Comparing Performance of ANN and SVM Methods for Regional Flood Frequency Analysis in South-East Australia. Water 14, 3323.
Zaman, M., Rahman, A., Haddad, K., 2012. Regional flood frequency analysis in arid regions: a case study for Australia. J. Hydrol. 475, 74–83.