REVIEW PAPER

Peaks-over-threshold model in flood frequency analysis: a scoping review

Xiao Pan[1] · Ataur Rahman[1] · Khaled Haddad[1] · Taha B. M. J. Ouarda[2]

Abstract

In flood frequency analysis (FFA), annual maximum (AM) model is widely adopted in practice due to its straightforward sampling process. However, AM model has been criticized for its limited flexibility. FFA using peaks-over-threshold (POT) model is an alternative to AM model, which offers several theoretical advantages; however, this model is currently underemployed internationally. This study aims to bridge the current knowledge gap by conducting a scoping review covering several aspects of the POT approach including model assumptions, independence criteria, threshold selection, parameter estimation, probability distribution, regionalization and stationarity. We have reviewed the previously published articles on POT model to investigate: (a) possible reasons for underemployment of the POT model in FFA; and (b) challenges in applying the POT model. It is highlighted that the POT model offers a greater flexibility compared to the AM model due to the nature of sampling process associated with the POT model. The POT is more capable of providing less biased flood estimates for frequent floods. The underemployment of POT model in FFA is mainly due to the complexity in selecting a threshold (e.g., physical threshold to satisfy independence criteria and statistical threshold for Generalized Pareto distribution – the most commonly applied distribution in POT modelling). It is also found that the uncertainty due to individual variable and combined effects of the variables are not well assessed in previous research, and there is a lack of established guideline to apply POT model in FFA.

Keywords GP distribution · Poison process · Peaks over threshold · Floods · Partial duration series

1 Introduction

Flooding is one of the worst natural disasters worldwide leading to significant economic losses (Acosta et al. 2016; Lathouris 2020). To effectively reduce flood damage, arising from the stochastic nature of extreme rainfall and runoff, assessments are usually undertaken by statistical methods. Flood frequency analysis (FFA) is one of the most preferred statistical methods, and widely used in infrastructure planning and design. FFA aims to estimate flood discharge with an associated frequency by fitting a probability distribution function to observed flood data. Hydrologists generally apply two modelling frameworks to perform FFA, annual maximum (AM) and peaks-over-threshold (POT). The AM model uses maximum discharge value from each year (i.e. one value from each year) at the location of interest. On the other hand, POT model extracts all the flow data above a threshold. The POT model is receiving a greater interest recently to understand the the nature of frequent floods, which are useful in characterising channel morphology and aquatic habitat, and helping river restoration efforts (Karim et al. 2017).

The AM model is the most popular method in practice given its straightforwardness in the sampling process. However, the sampling process in the AM model eliminates a large portion of the data from recorded streamflow time series. As an example, for a station with 50 years of streamflow record, AM model only considers 50 elements for modelling, each being the highest discharge data in a single year. Several studies note that AM model results in

✉ Ataur Rahman
a.rahman@westernsydney.edu.au

[1] School of Engineering, Design and Built Environment, Western Sydney University, Building XB.3.43, WSU, Kingswood, Locked Bag 1797, Penrith, NSW 2751, Australia

[2] INRS, National Institute for Scientific Research, INRS-ETE, 490, De la Couronne, Quebec City, QC G1K9A9, Canada

loss of useful information, e.g., the second highest flow data in a year (which could be higher than many data points in the AM series) is not selected (Bačová-Mitková and Onderka 2010; Bezak et al. 2014; Gottschalk and Krasovskaia 2002; Robson and Reed 1999). Traditional at-site FFA favors a 2 T rule, e.g. 100 years of streamflow record is needed for estimating 2% annual exceedance probability (AEP) or 50-year flood quantile estimate. The AM model has been criticized for its biased flood estimates in arid and semi-arid regions, in particular for smaller average recurrence intervals (ARIs) or frequent floods (Metzger et al. 2020; Zaman et al. 2012). The AM model has also been criticized for its considerable uncertainty in estimating frequent floods (Karim et al. 2017).

The POT model has been employed in extreme value analysis using Generalised Pareto (GP) distribution (Bernardara, Andreewsky and Benoit 2011; Coles 2001; Coles 2003; Coles et al. 2003; Liang et al. 2019; Northrop and Jonathan 2011; Pan and Rahman 2021; Thompson et al. 2009). The POT sampling process extracts a greater number of data points from the historical record compared to the AM model. The extracted POT series provides additional information by retaining all the data points above a selected threshold (Kumar et al. 2020; Madsen et al. 1997). POT model is also advantageous in terms of flexibility of sampling process, i.e. based on the purpose of the analysis, the POT model can extract desired numbers of data points by adjusting the level of threshold (Pan and Rahman 2018). However, the additional complexity associated with the POT model in relation to data independence is a negative aspect, which is one of the reasons for its under-employment (Lang et al. 1999; Pan and Rahman 2021). There is no unique procedure to select a threshold value in the POT modelling, and hence an iterative process is commonly adopted. As the threshold reduces, the number of selected flood peaks in the POT model increases; however, a very small threshold value can compromise with the independence criteria for some of the selected flood peaks. The threshold varies from catchment to catchment depending on catchment characteristics and flood generation mechanism. In contrast, in the AM series, the selected peak floods are most likely to be independent, as in this method, only one discharge data per year is selected.

As mentioned earlier, one of the complexities associated with POT model is the threshold selection (Beguería 2005; Gharib et al. 2017; Sccarrott and Macdonald 2012). Several methods have been proposed in selecting the threshold, which includes graphical methods (e.g. mean residual life plot) and shape stability plot of the GP distribution. These methods assume that for all the thresholds, above a well-chosen level, results in a stable shape parameter of the GP distribution (Durocher et al. 2018a, b). The graphical

methods are subjective and difficult to apply for a large number of stations. To select a threshold without human intervention, some studies (Irvine and Waylen 1986; Lang, Ouarda and Bobée, 1999) proposed selection of threshold associated with a given exceedance rate, which is mainly governed by site characteristics (with an acceptable range of 1.2–3 events per year on average). Selection of a threshold based on a given exceedance rate may not guarantee the fulfilment of the POT model assumption. To overcome this problem, Davison and Smith (1990) proposed Anderson–Darling (AD) test to identify a range of thresholds where the GP distribution hypothesis cannot be rejected.

Despite the rising interest of applying POT model, there is lack of commonly accepted guideline for its wider application, and only limited reviews on this approach have been conducted. For example, Lang et al. (1999) reviewed POT modelling and prepared a guide for application. Sccarrott and Macdonald (2012) reviewed advances in POT modelling based on statistical perspective. Later, Langousis et al. (2016) performed a critical review on threshold selection. Recently, the automated threshold detection techniques have been compared (Curceac et al. 2020; Durocher et al. 2019; Durocher et al. 2018a, b). To the best of our knowledge, scoping review on POT approach is limited and the literature on applying POT model remains sparse as compared to the AM model. To bridge the knowledge gap, this study aims to review and summaries the current status of POT model and to identify the difficulties in applying the POT model in FFA. It is expected that this scoping review will enhance the applicability of POT model and provide guidance on future research needs.

2 Review methodology

To undertake this scoping review, we followed the recommended framework by Sccarrott and Macdonald (2012) and Langousis et al. (2016). We firstly formulate the research questions, which is followed by identification of relevant keywords. To initialize the thought process, we asked: '*Why is POT modelling framework under-employed in FFA?*' If we find the answer to this question, we then ask: '*What are the conveniences in applying POT in FFA?*' and '*What challenges do we face in applying POT in FFA?*' Besides the questions mentioned above, we also formulated a range of additional questions to fully develop the framework of this scoping review:

i.   What are the current progresses in applying the POT model and at what scale (at-site or regional)?

ii. What are the major research gaps and future research needs in applying POT based FFA?

Based on the above research questions, we identified the relevant keywords for searching published articles on the POT model. The following keywords were applied to maximize the searching performance and to locate relevant articles from scientific database: 'extreme value' & limited to 'Pareto', 'partial duration' & limited to 'POT' and partial duration series ('PDS'), 'peaks over the threshold', 'Pareto', 'pooled analysis', 'threshold selection', 'annual maximum' & limited to 'POT'. The following scientific databases were used to locate relevant publications: Science Direct, Google Scholar and Scopus. We found more than 500 articles related to POT based FFA and only selected 135 publications that satisfied the above-mentioned criteria. The final step was to examine the selected articles and compile the review results in the form of this article. Figure 1 presents components and sub-components that were considered in this study.

The paper is organized as follows. Section 3 focuses on the common model assumptions applied to POT model, followed by reviewing the independence criteria and threshold selection in Sect. 4. Section 5 and Sect. 6 cover the parameter estimators and distribution functions for the POT model, respectively. Section 7 contains regional techniques, followed by discussion on stationarity in Sect. 8. Section 9 presents discussion, and Sect. 10 presents a summary of this review.

3 Model assumptions

Two common forms of model assumptions are associated with POT modelling, Poisson and Binomial (or negative binomial) processes coupled with either exponential or GP distributions. Poisson arrival assumes that the occurrence of the flood peaks above the selected threshold follows a Poisson process, provided the magnitude of flood peak is identically, independently distributed (i.i.d.). The most useful aspect of the Poisson arrival is that, with a given threshold, X, if the model follows Poisson process then other thresholds with values greater than X also follow Poisson process. Poisson assumption is the most applied sampling technique for construction of POT data series. However, the model assumption of Poisson arrival is not always valid. Cunnane (1979) used recorded flood data from 26 stations in the U.K to assess the validity of Poisson assumption, coupled with the exponential distribution and found that, the variance of the number of yearly flow peaks is significantly larger than the mean, which rejects the Poisson arrivals and suggests the fitting of binomial or negative binomial distributions. Ben-Zvi (1991) applied the Chi-square test to evaluate the model fitting performance of Poisson and negative binomial distributions using data from eight gauged stations in Israel. This study supported the negative binomial arrival in contrast to Poisson arrival. However, this study was inconclusive due to the limitations of the Chi-square test.
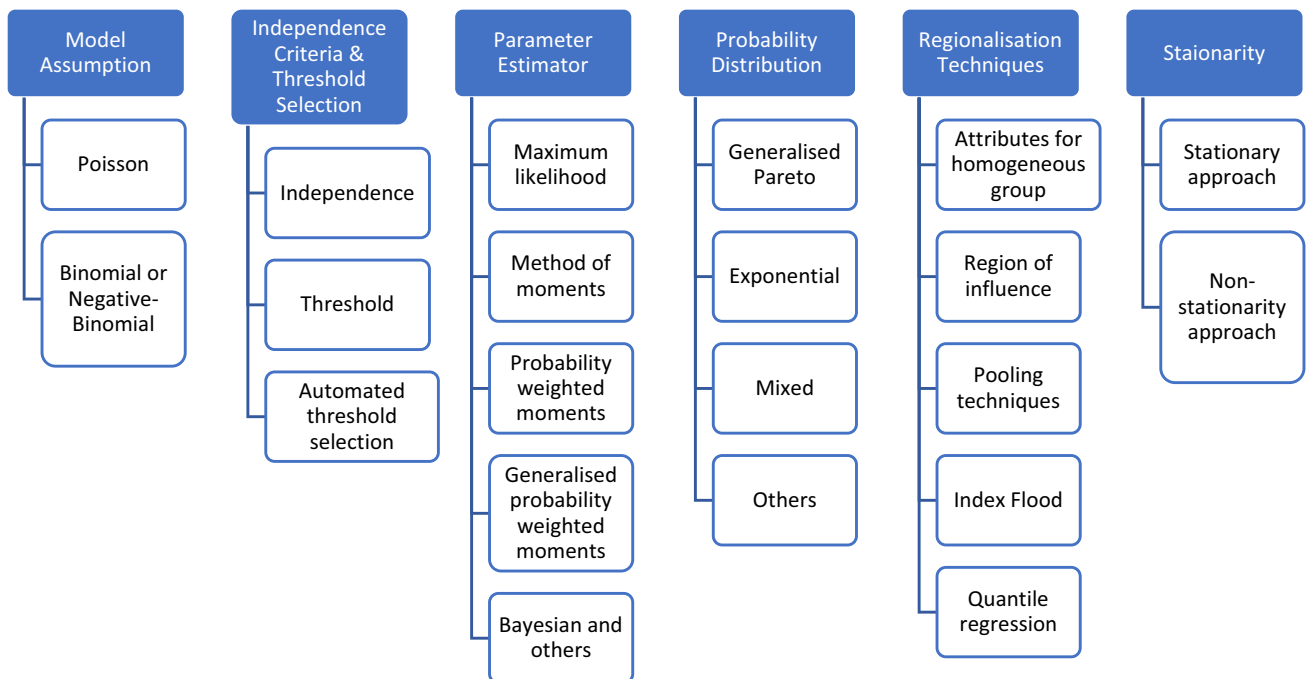

Fig. 1 Selected components and sub-components of POT based FFA

Lang et al. (1997) suggested using dispersion index test based on the given threshold for a choice of model assumption. They stated if the index test is greater or smaller than one, negative binomial or binomial assumption should constitute the Poisson process, respectively. Lang et al. (1999) proposed a practical guideline for POT based FFA, which only validate the Poisson assumption if the dispersion index is located within a 5% confidence interval with the selected threshold.

Önöz and Bayazit (2001) further evaluated the validity of applying binomial or negative binomial process in combination with the exponential probability distribution in POT based FFA. They found that the flood estimates based on binomial or negative binomial variates, are identical to the ones obtained from the Poisson process. The study concluded that Poisson arrival is the preferred process over binomial or negative binomial ones even if the Poisson process hypothesis is rejected. Furthermore, Eastoe and Tawn (2010) proposed mixed models to account for the overdispersion issue of annual peak count concerning the Poisson assumption. This study included the use of regression and mixed model to extend the homogeneous Poisson process.

To summarize the findings on the model assumption in POT based FFA, the Poisson arrival is the most preferred sampling assumption provided that the i.i.d. criterion is fulfilled. A dispersion index test could indicate if a constituted process to Poisson arrival should be used. However, only limited studies have evaluated the combined effects of model assumptions and sample size (Ben-Zvi 1991; Cunnane 1973, 1979). This potentially poses some degree of uncertainty in the design flood estimates based on POT modelling, which requires further investigation. Figure 2 illustrates POT modelling assumption.

4 Independence and threshold selection

POT sampling requires a dual-domain approach including time and magnitude. Coles (2001) proposed a de-clustering method, which filters the dependent elements from natural streamflow records (Solari and Losada 2012). Bernardara et al. (2014) later proposed a two-step framework for POT

modelling for estimating environmental extremes through defining physical and statistical thresholds for de-clustering and GP distribution, respectively. Two commonly applied independence criteria are discussed below.

(i) USWRC (1976) stated that independent flood peaks must have at least five days separation period (θ), plus the natural logarithm of selected basin area (A) measured in square mile. Besides, the intermediate flow value between two consecutive peaks must be dropped below three-quarter of the lowest of these two flow values. The second or any other flood peaks must be rejected if any of the criteria from Eq. 1 is met. The independence criteria specified by USWRC (1976) have been applied to POT based FFA by several studies (Bezak et al. 2014; Hu et al. 2020; Nagy et al. 2017).

$$\theta < 5\,days + \log(A)\,or\,Q_{min} > 75\%Min.\,[Q_1, Q_2] \qquad (1)$$

where A is basin area in mile ^{2}.

(ii) Another commonly applied independence criteria is recommended by Cunnane (1979), which is that two peaks must be separated by at least three times of the average time to peak. The average time to peak is obtained by assessing the hydrographs. Also, the minimum discharge between two consecutive peaks must be less than two-thirds of the discharge of the first of the two extremes. The second or any other flood peaks must be rejected if any of the criteria from Eq. 2 is met. Silva et al. (2012) and Chen et al. (2010) adopted the below independence criteria:

$$\theta < 3T_p \, or \, Q_{min} > \frac{2}{3}(Q_1) \qquad (2)$$

where T_p is the average time to peak.

Noteworthy, the criteria in Eqs. 1 and 2 for extraction of POT data series have been criticized due to the associated uncertainty. Ashkar and Rousselle (1983) reviewed the above two equations and concluded that the restriction for independence might render the Poisson process inapplicable.

Besides the criteria mentioned above, other methods are also proposed. For example, Lang et al. (1999) and Mostofi Zadeh et al. (2019) proposed: (i) Fixing the average number of exceedance per year for a predefined condition; (ii) Retaining flood peaks based on a predefined return

Fig. 2 Illustration of POT modelling assumptions
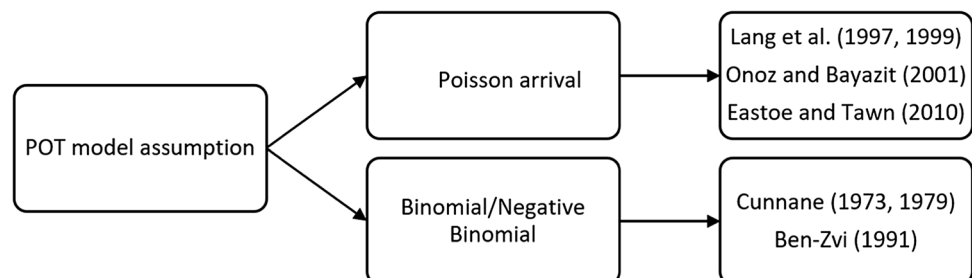
period; (iii) Retaining flood peaks based on a predefined frequency factor; and (iv) Selecting flood peaks that exceed 60% of bankfull discharge at a given station (Page et al., 2005). The bankfull discharge can be found examining the stage-discharge relationship and riverbank elevation at the gauging location, e.g. an inflection point on the stage-discharge relationship points to the bankfull discharge (Karim et al. 2017). Bankfull discharge is unique to each river and depends on factors like catchment area, geology and channel geometry and is generally taken as the AMF discharge with 1.5-year return period (Edwards et al. 2019). Solari and Losada (2012) proposed a unified statistical model for POT based FFA. Dupuis (1999) applied optimal bias-robust estimates (OBRE) to detect the threshold.

Another complexity associated with POT based FFA is determining the statistically sound threshold to fit GP distribution (POT-GP). Based on Pickands (1975), an extracted POT series with a sufficiently high threshold, the tail behavior follows a GP distribution. One of the most well-known properties of GP distribution is that the shape and modified scale parameter remain constant with the increased threshold. This property of GP distribution had been employed widely to verify the suitability of POT-GP approach in frequency analysis.

Traditionally, graphical diagnostics are employed to find a suitable threshold. Coles (2001) suggested three different types of graphical diagnostics including mean residual life plot (MRLP), threshold stability (TS) plot and other tools such as Q-Q, P-P and return level plot. Li, Cai and Campbell (2004) studied the extreme rainfall in Southwest Western Australia using POT-GP approach and identified the threshold based on the TS plot. Langousis et al. (2016) performed a critical review on representative methods for threshold selection for POT model based on GP distribution. This study suggested using MRLP over other methods as it is less sensitive to record length. However, graphical diagnostics have some drawbacks. For example, Sccarrott and Macdonald (2012) noted that graphical method required the practitioner to have substantial experience, and selecting a threshold could be subjective and associated uncertainty is difficult to quantify. There are several proposals to overcome this drawback by automating the threshold selection for POT-GP model through a computer program (Dupuis 1999; Liang et al. 2019; Solari and Losada 2012; Thompson et al. 2009). The main objectives of using such computer programs include quantifying the associated uncertainty in the flood estimates based on different sets of estimated parameters; and guiding the threshold selection based on goodness-of-fit (GOF) statistics, then to determine range of suitable thresholds based on a significance level while not rejecting the hypothesis of GP distribution.

The automation of threshold selection is developed based on the property of POT-GP model. Davison and Smith (1990) suggested using the Anderson–Darling (AD) GOF test to select a range of the threshold candidates for POT-GP model based on acceptable normality p-value (ND). This methodology is applied by Solari et al. (2017); in their study, the obtained thresholds based on POT-GP-ND approach mostly agreed with the ones obtained using the traditional graphical method. Durocher et al. (2018a, b) compared several automated threshold selection techniques based on POT-GP-ND approach and then proposed a hybrid method. The suggested method has a lower boundary of one peak per year (PPY) and an upper limit of 5 PPY, which accommodates the 1.6 PPY as recommended by Cunnane (1973) (at least of 1.63PPY for POT framework to have smaller sample variance compared with AM framework), and to accommodate the practical guideline between 1 and 3 PPY by Lang et al. (1999). This study found the shape parameter's consistency through a hybrid method in most of their selected sites. Zoglat et al. (2014) proposed another method using square error (SE), which was applied by Gharib et al. (2017). This method aims to find the optimum threshold by locating the minimum SE between stimulated and observed flood quantiles.

Curceac et al. (2020) evaluated the POT-GP-SE and POT-GP-ND approaches and proposed an empirical automated threshold selection based on cubic curve fitting to TS plot. They found that the proposed method based on TS plot had the greatest agreement of indices between empirical and theoretical quantiles at different time scales (15 min to daily). They addressed the need for further research on the combined effects of data scale, threshold selection and parameter estimator of the shape parameter of the GP distribution. Hu et al. (2020) applied POT model to USA and noted that when automatic threshold selection method was adopted with shorter data length, POT was unable to offer any additional benefit compared to AMF model.

Noteworthy, a single threshold for POT-GP approach might not be suitable for all situations. To overcome this, Deidda (2010) proposed a multiple threshold method (MTM) and found that MTM was better as compared to a single threshold through Monte Carlo simulation. Later, a quantitative assessment was performed by Emmanouil et al. (2020) for comparison of estimated quantiles using several approaches such as, AM, POT-MTM and multifractal approach.

Based on the above discussion, it is evident that the associated uncertainty with the POT-GP model does not arise only due to model assumption as noted in Sect. 3, but it is due to the combined effect of model assumption, threshold selection, data scale and parameter estimator. Future research could include the extent of the statistical

argument in an automated algorithm and explore the upper limit of applying MRLP. Figure 3 illustrates different threshold selection methods in POT modelling.

5 Parameter estimator

Parameter estimation is an essential step in POT based FFA like the AM model. Commonly applied estimators include maximum likelihood (ML), method of moments (MOM), probability weighted moments (biased/unbiased (PWMB/PWMU)), generalized probability weighted moments (GPWM) and other methods. This chapter briefly reviews the advances of parameter estimators in relation to FFA based on POT and GP distribution.

Pickands (1975) introduced GP distribution, and Hosking and Wallis (1987) first adopted this distribution in FFA. Hosking and Wallis (1987) compared the performance of the MOM, PWM and ML for estimation of GP parameters. They concluded that the PWM and MOM are the preferred estimator over ML, except for large sample size for quantile estimation, which aligns with Bobée et al. (1993) and Zhou et al. (2017a, b, c). MOM and PWM were preferred by Hu et al. (2020) and Metzger et al. (2020), respectively. ML has been widely adopted in many studies despite a limited sample size (Martins & Stedinger 2001; Mostofi Zadeh et al. 2019; Nagy et al. 2017; Ngongondo, Zhou and Xu 2020; Zhao et al. 2019a, b; Zhou et al. 2017a, b, c). Madsen et al. (1997) compared the performance of parameter estimator between AM and POT-GP models. Recently, Curceac et al. (2020) applied several commonly used estimators by Monte Carlo experiment and found that PWMU and PWMB are consistently least biased and less sensitive to the sample size, which is also in agreement with Hosking and Wallis (1987).

Other estimators were also evaluated in FFA using POT-GP model. For example, Ashkar and Ouarda (1996) evaluated the generalized method of moment (GMOM) for different shape parameters using observed and simulated data. Rasmussen (2001) developed a GPWM method and provided practical guidelines on this. In this study, GPWM was found to be outperforming the PWM, but only a relatively small difference was observed when compared with MOM. Martins and Stedinger (2001) examined the performance of generalized maximum likelihood estimator (GML) and compared its performance with MOM and L moments (LMOM). Kang and Song (2017) reviewed six estimators for GP distribution and found that the nonlinear least square-based method with modified POT series outperforming other estimators.

Selection of the best estimator for POT-GP model is an area that needs further research. The uncertainty due to parameter estimator is difficult to quantify. Ashkar and Tatsambon (2007) evaluated the upper bound of GP distribution applying different estimators, including MOM, ML, PWM and GPWM, through stimulation studies. They found that the upper bound of GP distribution is inconsistent between estimated and observed data. Later, Gharib et al. (2017) proposed a two-step framework for selecting both threshold selection techniques and the associated parameter estimators. However, the study was inconclusive (only POT-GP-SE method was assessed), and the study stressed the need of assessing the combined effects of estimator and threshold selection, similar to Curceac et al. (2020).

The Bayesian approach can be used to estimate distributional parameters in FFA for both AM and POT models. Here, the parameter of a distribution is treated as a random variable where the knowledge on a parameter is expressed by a prior distribution. In the context of PDS, Madsen et al.
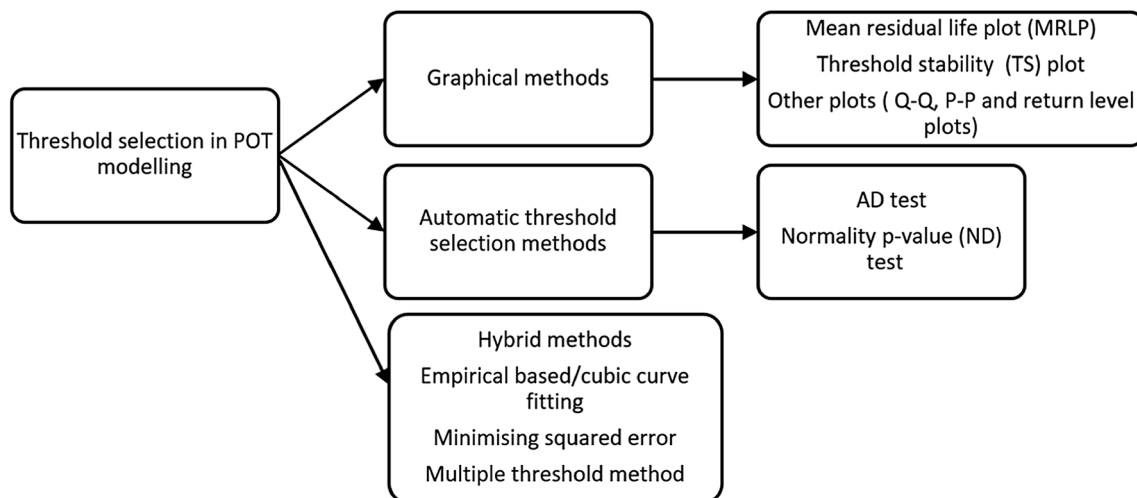

Fig. 3 Threshold selection methods in POT modelling

(1994) adopted a regional Bayesian approach for extreme rainfall modelling in Denmark where the empirical regional distributions of the parameters of the POT model were used as prior information for both exponential and GP distributions. Parent and Bariner (2003) also adopted a Bayesian approach to deal with the classical Poisson–Pareto-POT model for design flood estimation in the Garonne river in France. Bayesian POT approach was also adopted by Ribatet et al. (2007) and Silva et al. (2017).

Figure 4 illustrates MLRP plot for an Australian stream gauging station (223,204 Nicholson River at Deptford) having 56 years of streamflow data. This presents a comparison of the identified threshold based on MLE and PMWU in MRLP. It can be seen that there is no remarkable distinction in the identified thresholds using the two estimators having the same *p*-value (Pan and Rahman, 2021).

To summarize, it may be stated that the most applied parameter estimators for POT model include MOM, PWM and ML provided the sample size is large. The application of Bayesian approach in POT modelling is limited. We conclude that the sample size and range of the shape parameters are the primary considerations for selecting the parameter estimator. A choice of parameter estimator should be made based on the nature of a given data set. There is a lack of general practice guideline for selecting parameter estimator combined with threshold selection technique, and there is a lack in assessment of the upper bound of GP distribution.

6 Probability distributions

Fitting of the probability distribution to observed flood data is a primary step in any FFA exercise. GP distribution and its reduced form, exponential distribution, remains the most popular distribution in POT based FFA (Bobée et al. 1993; Davison and Smith 1990; Lang et al. 1999; Lang

et al. 1997; Lang et al. 1999; Madsen et al. 1997; Rasmussen & Rosbjerg 1991; Rosbjerg, Madsen and Rasmussen 1992; Silva et al. 2014; Yiou et al. 2006; Zhao, et al. 2019a, b). In this regard, use of extreme value theory by Pickands (1975) is justified. The GP distribution (two-parameter) is preferred over exponential (one-parameter) based on flexibility in modelling. However, use of GP distribution is restricted by its complexity of selecting a statistically sound threshold as discussed in Sect. 4.

GP distribution is well-known for its flexibility of modelling upper tail behavior, which is typical for observed flood records and other environmental extreme events. GP distribution is also widely employed in other disciplines for forecasting, trend analysis and risk assessment (Kiriliouk et al. 2019). Other probability distributions have also been applied in POT based FFA but remained unpopular. For example, Bačová-Mitková and Onderka (2010) applied Weibull distribution in POT based FFA and compared the obtained results with AM based FFA. This study concluded that the POT based FFA could produce comparable quantile estimates, especially for a shorter record length. Ashkar and Ba (2017) compared the Kappa distribution with GP due to their inherent similarity. Chen et al. (2010) proposed a bi-variate joint distribution for POT based FFA.

Figure 5 illustrates fitting of the GP distribution to POT series for stream gauging station 419,016 (Cockburn River at Mulla Crossing in New South Wales, Australia). It shows that POT 2-ND-MLE model provides the best fit to the observed flood data (POT 2 indicates 2 events per year being selected in the POT series). Also, POT3-ND-MLE model provides very good fit to the observed flood, in particular in frequent flood ranges (smaller ARIs).

Although the GP distribution remains the most popular distribution in POT based FFA, the associated uncertainty in higher return period is difficult to quantify. There are proposals to enhance the model fitting to POT series by introducing a mixture of models, which is GP based. For

Fig. 4 Mean residual life plot based on POT-ND method (*p*-value = 0.25) (223,204 Nicholson River at Deptford) (Pan and Rahman 2021)
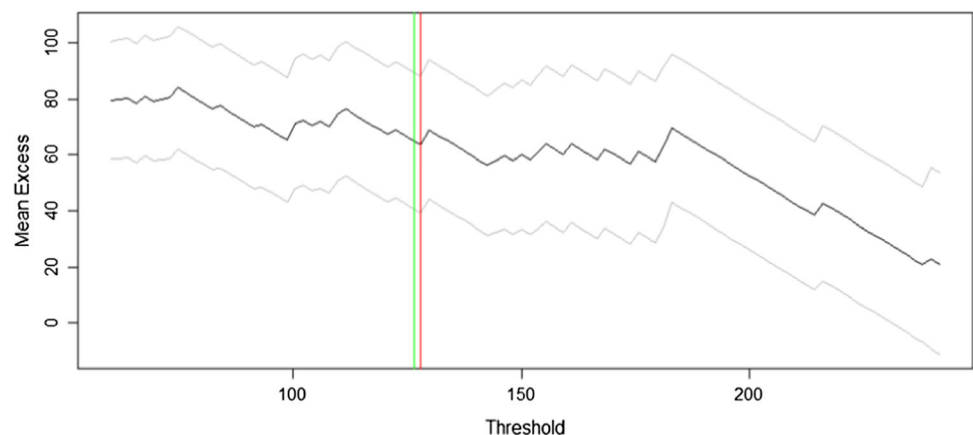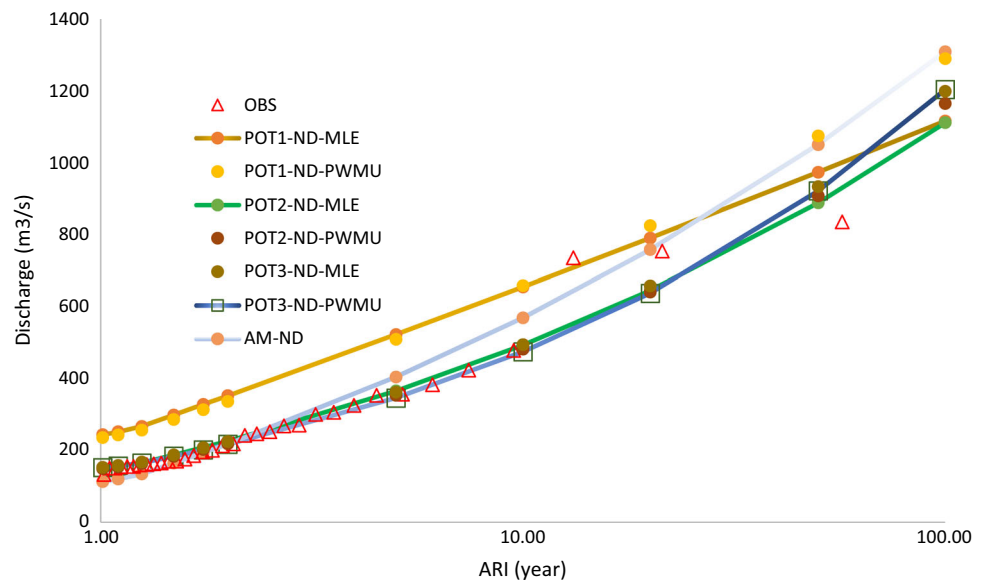
Fig. 5 Fitting different methods to the POT flood series (Station 419,016)

example, Solari and Losada (2012) proposed a unified statistical model called log-normal mixed with GP and quantified uncertainty associated with its tail behavior. However, the study was inconclusive and required further investigation to assess the combined effects (Curceac et al. 2020).

Another overlooked area is the bulk data below the threshold, as the GP distribution is only ideal for approximating the behavior of elements above the threshold. Several methods for the mixture of models (below and above threshold) are reviewed by Sccarrott and Macdonald (2012) and still, the associated uncertainty of these models are unquantifiable. The difficulties introduced by the mixture of models are the transitional point between two distribution functions and difficulties to accommodate the site specifics in modelling.

The goodness-of-fit (GOF) is commonly applied in selection of probability distribution(s) by comparing the empirical and theoretical distributions, such as Akaike information criterion (AIC), Bayesian information criterion (BIC), Kolmogorov–Smirnov (KS) and Anderson–Darling (AD) test. GOF is commonly applied in AM based FFA and can be used as a supplementary verification tool in POT based FFA. Choulakian and Stephens (2001) assessed the GP distribution fitting to 238 stream gauging stations in Canada by applying AD & Cramér–von Mises statistics and found that GP distribution providing an adequate fit to the observed POT data series. Gharib et al. (2017) assessed six parameter estimators using Relative Mean Square Error (RMSE) and AD test for the proposed framework based on the shape parameter. This study found that using the AD test, the proposed framework improved for 38% of the stations by an average of 65%. Laio, Di Baldassarre and Montanari (2009) reviewed AIC, BIC, and AD, and

concluded that the GOF tests produced satisfactory results. Haddad and Rahman (2011) also assessed several GOF tests and found from the Monte Carlo simulation that ADC was more successful in recognizing the parent distribution correctly than the AIC and BIC when the parent is a three-parameter distribution. On the other hand, AIC and BIC were better in recognizing the parent distribution correctly than the ADC when the parent was a two parameter distribution. Heo et al. (2013) proposed a modified AD test to assess the POT model. In regional POT framework, GOF is vital to assess the fit for the individual and group of sites. Silva et al. (2016) applied AIC, BIC, and likelihood ratio test for their study.

The parent probability distribution remains unknown at a given site. There are limited studies to quantify the uncertainty in fitting GP distribution to the POT data series, and current practice in comparing the flood estimates with either observed data or estimated flood quantiles by AM based FFA may not be adequate.

7 Regional flood frequency analysis

Regional flood frequency analysis (RFFA) is used to estimate design floods at ungauged catchments or at gauged catchments with limited data length or having data with poor quality(Komi et al. 2016; Haddad and Rahman 2012; Walega et al. 2016). In RFFA, AM flood data have widely been adopted (Haddad, Rahman and Stedinger 2012), and only minor attention has been given to the POT-based RFFA methods (Kiran and Srinivas 2021).

Identifying the hydrologically similar group or, the homogenous region is the first step in any RFFA. Three main categories of homogeneous regions are considered,

which are based on catchment attributes, flood data and geographical proximity (Ashkar 2017; Paixao et al. 2015; Shu & Burn 2004; Zhang et al. 2020). Traditionally, the geographical proximity is commonly adopted to form homogeneous groups. Other methods include catchment characteristics data to form homogeneous groups (Bates et al. 1998). Zhang and Stadnyk (2020) reviewed the popular attributes considered in RFFA in forming homogeneous groups in Canada, which included geographical proximity, flood seasonality, physiographic variables, monthly precipitation pattern, and monthly temperature pattern. A revision of RFFA algorithm was proposed by Zhang et al. (2020) based on AM based RFFA and to satisfy 5 T rule, which was initially suggested by Hosking and Wallis (1987) and Reed et al. (1999). Rahman et al. (2020) performed an independent component analysis using data from New South Wales, Australia; however, it considered AM flood data and homogeneity was not specially considered. All of the above methods to identify homogeneous regions used AM flood data. Research on homogeneous regions using POT data is limited to-date.

Cunderlik and Burn (2002) proposed a site-focused pooling technique based on flood seasonality, namely flood regime index, to increase the number of initial homogenous groups and found it to be superior to the mean of mean day (MDF) descriptor; however, the sampling variability was not considered. Cunderlik and Burn (2006) later proposed a new pooling group for flood seasonality based on non-parametric sampling, where the similarity between the target site and potential site was assessed by the minimum confidence interval of the intersection of Mahalanobis ellipses. Shu and Burn (2004) developed a method using fuzzy expert system to derive an objective similarity measure between catchments. There are also other methods for forming of the homogeneous groups (Burn and Goel 2000; Cord 2001). Carreau et al. (2017) proposed an alternative approach using hazard level to partition the region into sub-regions for POT model, which aims to formulate the approach as a mixture of GP distributions.

Index flood method was proposed by Dalrymple (1960) and remains one of the most popular methods in AM based RFFA (Hosking & Wallis 1993; O'Brien and Burn 2014; Robson and Reed 1999). This is due to its simplicity in developing a regional growth curve and weighing the sites by index-variables such as mean annual flood. Index flood approach was applied with POT data by Madsen and Rosbjerg (1997b) where GP shape parameter was regionalized. They examined the impacts of regional heterogeneity and inter-site dependence on the accuracy of quantile estimation. They found that POT-based RFFA was more accurate than the at-site FFA estimate even for extremely heterogenous regions. They also noted that modest inter-site dependence had only minor effects on the

accuracy of POT based index flood method. The POT based RFFA was further explored by Madsen and Rosbjerg (1997a) and Madsen et al. (2017). Roth et al. (2012) developed a nonstationary index flood method using POT data based on a composite likelihood test. Mostofi Zadeh et al. (2019) performed a pooled analysis based on both AM and POT data (using AM pooling technique and then applied to POT model and vice versa) and concluded that the POT model pooling group reduced uncertainty in design flood estimates. Quantile regression technique has been widely employed under AM model (Haddad and Rahman 2012). Durocher et al. (2019) compared four estimators based on index flood method and quantile regression technique including regression analysis, L-moments and likelihood method using POT data.

Gupta et al. (1994) noted that the coefficient of variation of AM flows should not vary with catchment area in a proposed region/group. This may not satisfy for many regions, which led to the Bayesian approaches, which was studied by Madsen, Rosbjerg and Harremoës (1994) by using an exponential distribution in RFFA using POT data. This study adopted total precipitation depth and maximum 10-min rainfall intensity of individual storms for Bayesian inferences. The proposed method was found to be preferable for estimating design floods of return period less than 20 years. Madsen and Rosbjerg (1997a) proposed an index flood method based on a Bayesian approach, which combined the concept of index flood with empirical Bayesian approximation so that the inference on regional information can be made with more accuracy. Ribatet et al. (2007) implemented Markov Chain Monte Carlo (MCMC) technique along with GP distribution to sample the posterior distribution. Silva et al. (2017) studied Bayesian inferential paradigm coupled with MCMC under POT framework. Some attention was drawn to using historical information in RFFA to enhance accuracy of flood estimates. Sabourin and Renard (2015) proposed a new model utilising historical information similar to Hamdi et al. (2019). Kiran and Srinivas (2021) used POT data from 1031 USA catchments to develop regression based RFFA technique. They noted that scale and shape parameters of the GP distribution fitted to PDS data were largely governed by catchment size and 24-h rainfall intensity corresponding to 2-year return period.

8 Stationarity

Stationarity is one of the most critical concepts in applying extreme value theory in hydrology, which implies that the estimated parameters of the given probability distributions do not change with time, i.e., the current parameters used for modelling remain constant for the future so that the

quantile estimates remain consistent over time. However, in assessing the AM and POT time series data, a trend or a jump may be found in many cases, which undermines the stationary assumption (Ishak and Rahman 2015; Ishak et al. 2013). The identified anomalies may be statistically significant or insignificant, which may be due to climate change or other reasons such as land use changes (Burn et al. 2010; Cunderlik et al. 2007; Cunderlik and Ouarda 2009; Ngongondo et al. 2013; Ngongondo, Zhou and Xu 2020; Silva et al. 2012; Zhang, Duan and Dong 2019). Recent application of POT with non-stationarity is proposed by Lee, Sim and Kim (2019) for extreme rainfall analysis.

To account for non-stationarity, parameters of the distribution requires adjustment as a function of time. El Adlouni et al. (2007) and Villarini et al. (2009) applied the function of temporal covariates for parameters of a probability distribution. Koutsoyiannis (2006) evaluated the two commonly applied approaches for nonstationary analysis and argued that the common FFA approaches are not consistent with the rationale of the stationary analysis.

In POT modelling, GP distribution is the most used distribution. In non-stationary approach, GP distribution commonly presents a constant shape parameter and a varied scale parameter (time-dependent or covariate with climate indices under nonstationary condition). In this regard, Coles (2001) argued that even under stationary condition, the shape parameter is difficult to estimate. Regression analysis is commonly applied to quantify the trend before applying the variation in the scale parameter (Vogel, Yaindl and Walter 2011). Moreover, for GP distribution, the threshold can also be treated as time-dependent (Kyselý et al. 2010). Recently, Vogel and Kroll (2020) compared several estimators for non-stationary frequency analysis.

In the context of POT based RFFA, Roth et al. (2012) examined non-stationarity by varying scale parameter using index flood method and suggested a time-dependent regional growth curve for temporal trends observed in the study data set. Silva et al. (2014) proposed a zero-inflated Poisson GP model for the non-stationarity condition, and proposed a non-stationarity RFFA technique based on Bayesian method (Le Vine 2016; Parent and Bernier 2003; Silva et al. 2017). Mailhot et al. (2013) proposed a POT based RFFA approach to a finer resolution using rainfall time series data. O'Brien and Burn (2014) studied nonstationary index flood method using AM flood data. They noted the challenge of forming a homogenous group, which was due to several sites presenting significant level of nonstationarity. Durocher et al. (2019) compared several estimators under nonstationary condition using index flood method for a data set of 425 Canadian stations and found that the L moments approach was more robust and less biased than ML estimator. This study also found that a

hybrid pooling group approach, which included sites with stationary and nonstationary conditions, improved the accuracy of the quantile estimates. The recent study by Agilan, Umamahesh and Mujumdar (2020) stated the uncertainty due to threshold under non-stationarity condition is 54% higher than the ones under stationary consideration. Reed and Vogel (2015) questioned the applicability of return period concept in FFA under nonstationary condition. They demonstrated how a parsimonious nonstationary lognormal distribution can be linked with nonstationary return periods, risk, and reliability to gain a deeper understanding of future flood risk. For the non-stationary POT models, the risk and reliability concept need to be further explored as suggested by Reed and Vogel (2015).

Iliopoulou and Koutsoyiannis (2019) developed a probabilistic index based on the probability of occurrence of POT events that can discover clustering linked to the persistence of the parent process. They found that rainfall extremes could exhibit notable departures from independence, which could have important implications on POT based FFA under both stationary and non-stationary regimes. Thiombiano et al. (2017) and Thiombiano et al. (2018) presented how climate change indices can be used as covariates in a non-stationary framework in the POT modelling.

9 Discussion

Traditional FFA based on AM model is the most popular FFA method given its straightforward sampling process and availability of a wide range of literature and guidelines. Even though there are theoretical advantages with POT based FFA, this is still under-employed. Table 1 presents a summary where POT based FFA have been examined. It should be mentioned that although many researchers have examined the suitability of the POT based FFA, its inclusion in flood estimation guide is limited. For example, Australian Rainfall and Runoff (ARR) 2019 stated that POT-based method can be adopted for FFA, but its FLIKE software does not include any POT-based analysis (Kuczera and Franks, 2019). They stated that POT is more appropriate in urban stormwater applications and for diversion works, coffer dams and other temporary structures; for most of these cases recorded flow data are unavailable. In design rainfall estimation for Australia in ARR 2019, POT-based methods have been adopted to estimate more frequent design rainfalls (Green et al., 2019).

Unlike AM model, which only extracts a single value per year, the POT is more complex in its sampling process. Cunnane (1973) argued to have at least 1.6 event per year on average in the POT model to provide less biased

Table 1 Summary of POT based flood frequency analysis studies in selected countries

United Kingdom	Cunnane (1973), (1979); Hosking and Wallis (1987); Acreman (1987); Davison and Smith (1990); Reed et al. (1999); Eastoe and Tawn (2010); Northrop and Jonathan (2011); Le Vine (2016); Curceac et al. (2020)
USA	Hu et al. (2020); Kiran and Srinivas (2021); Metzger et al. (2020); Phillips et al. (2018); Armstrong et al. (2014); Edwards et al. (2019); Armstrong et al. (2012);
Australia	Page and McElroy (1981); Keast and Ellison (2013); Rustomji (2009); Karim et al. (2017); Pan and Rahman (2018); Green et al. (2019); Pan and Rahman (2021)
Canada	Ashkar and Ba (2017); Ashkar et al. (1987); Ashkar and El Adlouni (2015); Ashkar and Ouarda (1996); Ashkar and Rousselle (1983); Ashkar and Tatsambon (2007); Irvine and Waylen (1986); Burn (1990a, b); Burn and Goel (2000); Burn and Whitfield (2016); Bobée et al. (1993); Adamowski (2000); Adamowski et al. (1998); Dupuis (1999); Shu and Burn (2004); Ribatet et al. (2007); O'Brien and Burn (2014); Gharib et al. (2017); (Mostofi Zadeh and Burn 2017; Mostofi Zadeh et al. 2019); (Durocher et al. 2019; Durocher et al. 2018a, b; Durocher et al. 2018a, b); (Zhang et al 2020); Lang et al. (1999)
France	Navratil et al. (2010); Bernardara et al. (2011); Bernardara et al. (2014), (2012); Evin et al. (2016); Laio et al. (2009); Lang et al. (1999); Lang et al. (1997); Weiss et al. (2012); Carreau et al. (2017); Tramblay et al. (2013); Thompson et al. (2009); Hamdi et al. (2015), (2019); Sabourin and Renard (2015);
Denmark	Madsen et al. (2002); Madsen et al. (1997); Madsen and Rosbjerg (1997a, b); Madsen et al. (1994); Madsen et al. (1995)
China	Chen et al. (2010); Liang et al. (2019); Zhou et al. (2017a, b)
Portugal	Silva et al. (2016); Silva et al. 2012, 2014; Silva et al. (2017); (Tavares and Da Silva 1983)
New Zealand	Mohssen (2009); Sccarrott and Macdonald (2012); Nagy et al. (2017)

Turkey	Önöz and Bayazit 2001)	Poland	Rutkowska et al. (2017a, b); Rutkowska et al. (2017a, b); Walega et al. (2016);
Morocco	Zoglat et al. (2014)	Spain	Solari et al. (2017); Solari and Losada (2012); Cord (2001)
Algeria	Renima et al. (2018)	Greece	Koutsoyiannis (2006)
Slovak Republic	Bačová-Mitková & Onderka (2010)	Norway	Ngongondo et al. (2013); Ngongondo et al. (2020)
Israel	Metzger et al. (2020); Ben-Zvi (1991)	Germany	Komi et al. (2016)
The Netherlands	Beguería (2005); Roth et al. (2012); Kiriliouk et al. (2019)	India	Kumar et al. (2020)
Slovenia	Bezak et al. (2014), (2016)	South Korea	Heo et al. (2013); Kang and Song (2017)
Czech Republic	Yiou et al. (2006); Kyselý et al. (2010)		

estimates than AM model. However, with the recent advances in computational modelling, Durocher et al. (2018a, b) applied an upper limit of 5 events per year in POT based FFA and obtained results which are comparable to the AM model. Besides, ensuring the independence of the extracted data is one of the other difficulties faced in applying the POT model. Two commonly applied criteria in constructing POT series as described in Sect. 4 have been criticized by Ashkar and Rousselle (1983) for possible violation of the model assumptions. POT based FFA is also constrained to the model assumptions, either with Poisson or binomial (or negative binomial) arrivals. Önöz and Bayazit (2001) reported a comparable result even when the assumptions are violated, although the associated uncertainty is not well studied.

Another well-known difficulty is to identify the threshold for GP distribution in POT modelling. Based on the critical review of current methods by Langousis et al. (2016), MRLP is found to be the most effective detection method, which leads to the least bias design flood estimates. However, this study suggested further research on automated procedure in POT data construction under stationary condition with additional statistical arguments. Nonstationary POT based FFA has attracted more attention recently (Durocher et al. 2019; Mostofi Zadeh et al. 2019); however, the findings of these studies are not conclusive, and further research is warranted on nonstationary POT based FFA.

Uncertainty in flood estimates is still a challenging topic with the recent floods in Europe and China, it is noted that traditional FFA approaches need an overhaul, and a comprehensive uncertainty analysis is warranted. POT model is flexible in data extraction as compared to AM model, but this brings additional levels of uncertainty such as sample

size variation (i.e., average events per year is not fixed) and effects of time scale of data extraction (e.g., 15 min, hourly, daily or monthly).

Besides, POT model requires a threshold to be determined for GP distribution. For this, no universal method has been proposed. To successfully extract the POT series, two independence criteria for retaining flood peaks are discussed in Sect. 4. However, the uncertainty associated with independence criteria is not fully understood despite studies have shown that the independence criteria need to be situation specific. On the other hand, the threshold determination to suit the assumption of the GP distribution is one of the other concerns. The associated uncertainty is more complex in homogeneous group formation for POT modelling. Beguería (2005) performed a sensitivity analysis on the threshold selection to the parameter and quantile estimates and stated no unique optimum threshold value could be detected. Durocher et al. (2018a, b) also stated that the threshold selection affects the trend detection significantly, and currently, no acceptable method has been found. Moreover, as discussed in Sects. 5 and 6, various estimators, distribution functions, and GOF tests aim to reduce the uncertainty in POT based FFA.

AM model certainly is the most popular and well-studied FFA approach but it has limitations too. POT model is an alternative FFA approach, which has been proven to be advantageous in many studies. Below is the list that summarizes key points on the conveniences of applying POT model in FFA.

- POT model is not that limited (compared to AM model) by smaller data length due to its sampling process as the overall data length is controllable. This provides additional flexibility with the POT model.
- POT model is proven to be efficient for the arid/semi-arid regions as streams here may have low/zero flows in many years over the gauging period.
- POT is proven to be efficient in estimating very frequent to frequent flood quantiles, which are needed in environmental and ecological studies.
- Due to the controllable resolution of the time series, POT is more suited to present the trend and perform the nonstationary FFA.
- POT can provide bigger data set in the context of RFFA due to its nature of data extraction process, which may be useful to regionalize very frequent floods.

10 Conclusion

Two main modelling approaches, annual maximum (AM) and peaks-over-threshold (POT), are adopted in FFA. The AM model is well employed due to its straightforward

sampling process, while the POT is under-employed internationally. In this scoping review, we found that POT model is more flexible than AM one due to the nature of the data extraction process. It is found that POT based FFA can provide less biased estimates for small to medium-sized flood quantiles (in the more frequent ranges). Furthermore, POT model is more suitable for design flood estimation in the arid and semi-arid regions as in these regions many years do not experience any runoff.

Despite the advantages with the POT model, it has several complexities. The physical threshold determination (to ensure the independence of the extracted data points and to satisfy the model assumptions) is the first obstacle that discourages the wider application of the POT model in FFA. The effects of independence criteria on the uncertainty in the final flood estimates are not examined thoroughly as well as the combined effects of the extracted sample size and independence criteria.

The statistical threshold selection in applying GP distribution in POT based FFA is the second obstacle. In this regard, a commonly accepted guideline has not been produced yet. In the critical review by Langousis et al. (2016), the MRLP is found to be a promising method of threshold detection; however, the scaling effect is not well examined (e.g., to what scale, the MRLP is efficient for detection of threshold). The suggested approach is to examine a more suitable statistical argument in the iterative/automated process. Additional complexity arises from the model assumption (Poisson or binomial/negative binomial), parameter estimators and distribution functions. The uncertainty of the individual component may not be significant; however, the combined effects of these aspects may increase the level of uncertainty in flood quantile estimates by the POT model, which requires further investigation.

We have found few recent studies involving the mixture of AM and POT modelling frameworks, but this needs further research as there are several unanswered questions with this combination. The POT based RFFA also requires further research as there are only handful of studies on this, which would be very useful to increase the accuracy of design flood estimates in ungauged catchments for smaller return periods, which are often needed in environmental and ecological studies. The non-stationary FFA based on POT model needs further research as the future of FFA lies in the non-stationary approaches. Since POT model has more parameters, the estimation of the effects of climate change on this model is more challenging, and this is an area that needs further research.

manuscript. We also acknowledge Australian Government for providing streamflow data used in this study.

Declarations

References

Acosta LA, Eugenio EA, Macandog PBM, Magcale-Macandog DB, Lin EKH, Abucay ER et al (2016) Loss and damage from typhoon-induced floods and landslides in the Philippines: community perceptions on climate impacts and adaptation options. Int J Global Warm 9(1):33–65

Acreman, M 1987, 'Regional flood frequency analysis in the UK: Recent research-new ideas', *Institute of Hydrology, Wallingford, UK.*

Adamowski K (2000) Regional analysis of annual maximum and partial duration flood data by nonparametric and L-moment methods. J Hydrol 229(3–4):219–231

Adamowski K, Liang GC, Patry GG (1998) Annual maxima and partial duration flood series analysis by parametric and non-parametric methods. Hydrol Process 12(10–11):1685–1699

Agilan V, Umamahesh NV, Mujumdar PP (2020) Influence of threshold selection in modeling peaks over threshold based nonstationary extreme rainfall series. J Hydrol 593:125625

Armstrong WH, Collins MJ, Snyder NP (2012) Increased frequency of low-magnitude floods in New England 1. JAWRA J Am Water Res Associat 48(2):306–320

Armstrong WH, Collins MJ, Snyder NP (2014) Hydroclimatic flood trends in the northeastern United States and linkages with large-scale atmospheric circulation patterns. Hydrol Sci J 59(9):1636–1655

Ashkar F, El-Jabi N, Bobee B (1987) 'On the choice between annual flood series and peaks over threshold series in flood frequency analysis', pp. 276–80, Scopus

Ashkar F (2017) 'Delineation of homogeneous regions based on the seasonal behavior of flood flows: an application to eastern Canada', pp. 390–7, Scopus

Ashkar F, Ba I (2017) Selection between the generalized Pareto and kappa distributions in peaks-over-threshold hydrological frequency modelling. Hydrol Sci J 62(7):1167–1180

Ashkar F, El Adlouni SE (2015) Adjusting for small-sample non-normality of design event estimators under a generalized Pareto distribution. J Hydrol 530:384–391

Ashkar F, Ouarda TBMJ (1996) On some methods of fitting the generalized Pareto distribution. J Hydrol 177(1–2):117–141

Ashkar F, Rousselle J (1983) The effect of certain restrictions imposed on the interarrival times of flood events on the Poisson distribution used for modeling flood counts. Water Resources Res 19(2):481–485

Ashkar F, Tatsambon CN (2007) Revisiting some estimation methods for the generalized Pareto distribution. J Hydrol 346(3):136–143

Bačová-Mitková V, Onderka M (2010) Analysis of extreme hydrological events on the sanube using the peak over threshold method. J Hydrol Hydromech 58(2):88–101

Bates BC, Rahman A, Mein RG, Weinmann PE (1998) Climatic and physical factors that influence the homogeneity of regional floods in south-eastern Australia. Water Resour Res 34(12):3369–3381

Beguería S (2005) Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value. J Hydrol 303(1):215–230

Ben-Zvi A (1991) Observed advantage for negative binomial over Poisson distribution in partial duration series. Stoch Hydrol Hydraul 5(2):135–146

Bernardara P, Mazas F, Weiss J, Andreewsky M, Kergadallan X, Benoît M et al. (2012) 'On the two step threshold selection for over-threshold modelling', Scopus

Bernardara P, Andreewsky M, Benoit M (2011) Application of regional frequency analysis to the estimation of extreme storm surges. J Geophy Res Oceans. https://doi.org/10.1029/2010JC006229

Bernardara P, Mazas F, Kergadallan X, Hamm L (2014) A two-step framework for over-threshold modelling of environmental extremes. Natural Hazard Earth Sys Sci 14(3):635–647

Bezak N, Brilly M, Šraj M (2014) Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. Hydrol Sci J 59(5):959–977

Bezak N, Brilly M, Šraj M (2016) Flood frequency analyses, statistical trends and seasonality analyses of discharge data: a case study of the Litija station on the Sava River. J Flood Risk Manage 9(2):154–168

Bobée B, Cavadias G, Ashkar F, Bernier J, Rasmussen PF (1993) Towards a systematic approach to comparing distributions used in flood frequency analysis. J Hydrol 142(1–4):121–136

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

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

Burn DH, Goel NK (2000) The formation of groups for regional flood frequency analysis. Hydrol Sci J 45(1):97–112

Burn DH, Sharif M, Zhang K (2010) Detection of trends in hydrological extremes for Canadian watersheds. Hydrol Process 24(13):1781–1790

Burn DH, Whitfield PH (2016) Changes in floods and flood regimes in Canada. Canadian Water Resource J 41(1–2):139–150

Carreau J, Naveau P, Neppel L (2017) Partitioning into hazard subregions for regional peaks-over-threshold modeling of heavy precipitation. Water Resour Res 53(5):4407–4426

Chen L, Guo S, Yan B, Liu P, Fang B (2010) A new seasonal design flood method based on bivariate joint distribution of flood magnitude and date of occurrence. Hydrol Sci J 55(8):1264–1280

Choulakian V, Stephens MA (2001) Goodness-of-fit tests for the generalized pareto distribution. Technometrics 43(4):478–484

Coles S (2003) 'The use and misuse of extreme value models in practice', In: Extreme values in finance, telecommunications, and the environment, pp. 79–100

Coles S (2001) An introduction to statistical modelling of extreme values. Springer, London

Coles S, Pericchi LR, Sisson S (2003) A fully probabilistic approach to extreme rainfall modeling. J Hydrol 273(1–4):35–50

Cord O (2001) Genetic fuzzy systems: evolutionary tuning and learning of fuzzy knowledge bases, vol. 19, World Scientific

Cunderlik JM, Burn DH (2002) The use of flood regime information in regional flood frequency analysis. Hydrol Sci J 47(1):77–92

Cunderlik JM, Burn DH (2006) Switching the pooling similarity distances: mahalanobis for Euclidean. Water Resource Res. https://doi.org/10.1029/2005WR004245

Cunderlik JM, Jourdain V, Quarda TBMJ, Bobée B (2007) Local non-stationary flood-duration-frequency modelling. Canadian Water Resource J 32(1):43–58

Cunderlik JM, Ouarda TBMJ (2009) Trends in the timing and magnitude of floods in Canada. J Hydrol 375(3):471–480

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

Cunnane C (1979) A note on the poisson assumption in partial duration series models. Water Resources Res 15(2):489–494

Curceac S, Atkinson PM, Milne A, Wu L, Harris P (2020) An evaluation of automated GPD threshold selection methods for hydrological extremes across different scales. J Hydrol 585:124845

Dalrymple T (1960) Flood-frequency analyses, manual of hydrology: Part 3, USGPO

Davison AC, Smith RL (1990) Models for exceedances over high thresholds. J Royal Statist Soci: Series B (Methodological) 52(3):393–425

Deidda R (2010) A multiple threshold method for fitting the generalized Pareto distribution to rainfall time series. Hydrol Earth Syst Sci 14(12):2559

Dupuis DJ (1999) Exceedances over high thresholds: a guide to threshold selection. Extremes 1(3):251–261

Durocher M, Burn DH, 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

Durocher M, Burn DH, Mostofi Zadeh S (2018a) A nationwide regional flood frequency analysis at ungauged sites using ROI/GLS with copulas and super regions. J Hydrol 567:191–202

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

Eastoe EF, Tawn JA (2010) Statistical models for overdispersion in the frequency of peaks over threshold data for a flow series. Water Resource Res. https://doi.org/10.1029/2009WR007757

Edwards PJ, Watson EA, Wood F (2019) Toward a better understanding of recurrence intervals, bankfull, and their importance. J Contemporary Water Res Educat 166(1):35–45

El Adlouni S, Ouarda TBMJ, Zhang X, Roy R, Bobée B (2007) Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. Water Resource Res. https://doi.org/10.1029/2005WR004545

Emmanouil S, Langousis A, Nikolopoulos EI, Anagnostou EN (2020) Quantitative assessment of annual maxima, peaks-over-threshold and multifractal parametric approaches in estimating intensity-duration-frequency curves from short rainfall records. J Hydrol 589:125151

Evin G, Blanchet J, Paquet E, Garavaglia F, Penot D (2016) A regional model for extreme rainfall based on weather patterns subsampling. J Hydrol 541:1185–1198

Gharib A, Davies EGR, Goss GG, Faramarzi M (2017) Assessment of the combined effects of threshold selection and parameter estimation of generalized Pareto distribution with applications to flood frequency analysis. Water (Switzerland) 9(9):692

Gottschalk L, Krasovskaia I (2002) L-moment estimation using annual maximum (AM) and peak over threshold (POT) series in regional analysis of flood frequencies. Norsk Geografisk Tidsskrift - Norwegian J Geography 56:179–187

Green J, Beesley JFC The C (2019) Design Rainfall. In: Australian rainfall & runoff, Commonwealth of Australia

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

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 Env 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, Stedinger JR (2012) Regional flood frequency analysis using Bayesian generalized least squares: a comparison between quantile and parameter regression techniques. Hydrol Processes 26(7):1008–1021

Hamdi Y, Bardet L, Duluc CM, Rebour V (2015) Use of historical information in extreme-surge frequency estimation: the case of marine flooding on the La Rochelle site in France. Nat Hazard 15(7):1515–1531

Hamdi Y, Duluc CM, Bardet L, Rebour V (2019) Development of a target-site-based regional frequency model using historical information. Nat Hazards 98(3):895–913

Heo J-H, Shin H, Nam W, Om J, Jeong C (2013) Approximation of modified Anderson-Darling test statistics for extreme value distributions with unknown shape parameter. J Hydrol 499:41–49

Hosking JRM, Wallis JR (1987) Parameter and quantile estimation for the generalized pareto distribution. Technometrics 29(3):339–349

Hosking JRM, Wallis JR (1993) Some statistics useful in regional frequency analysis. Water Resour Res 29(2):271–281

Hu L, Nikolopoulos EI, Marra F, Anagnostou EN (2020) Sensitivity of flood frequency analysis to data record, statistical model, and parameter estimation methods: an evaluation over the contiguous United States. J Flood Risk Manage 13(1):e12580

Iliopoulou T, Koutsoyiannis D (2019) Revealing hidden persistence in maximum rainfall records. Hydrol Sci J 64(14):1673–1689

Irvine K, Waylen P (1986) Partial series analysis of high flows in Canadian rivers. Canadian Water Resource J 11(2):83–91

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

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

Kang S, Song J (2017) Parameter and quantile estimation for the generalized Pareto distribution in peaks over threshold framework. J Korean Stat Soci 46(4):487–501

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

Keast D, Ellison J (2013) Magnitude frequency analysis of small floods using the annual and partial series. Water 5(4):1816–1829

Kiran KG, Srinivas VV (2021) Distributional regression forests approach to regional frequency analysis with partial duration series. Water Resource Res 57(10):e2021WR029909

Kiriliouk A, Rootzén H, Segers J, Wadsworth JL (2019) Peaks over thresholds modeling with multivariate generalized pareto distributions. Technometrics 61(1):123–135

Komi K, Amisigo BA, Diekkrüger B, Hountondji FCC (2016) Regional flood frequency analysis in the Volta River Basin, West Africa. Hydrology 3(1):5

Koutsoyiannis D (2006) Nonstationarity versus scaling in hydrology. J Hydrol 324(1):239–254

Kuczera, G., & Franks, S. (2019). At-site flood frequency analysis. in *Australian Rainfall & Runoff*, Commonwealth of Australia.

Kumar M, Sharif M, Ahmed S (2020) Flood estimation at Hathnikund Barrage, river Yamuna, India using the Peak-Over-Threshold method. ISH J Hydraulic Eng 26(3):291–300

Kyselý J, Picek J, Beranová R (2010) Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. Global Planet Change 72(1):55–68

Laio F, Di Baldassarre G, Montanari A (2009) Model selection techniques for the frequency analysis of hydrological extremes. Water Resources Res. https://doi.org/10.1029/2007WR006666

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

Lang M, Rasmussen PF, Oberlin G, Bobee B (1997) Over-threshold sampling: modeling of occurrences by renewal processes. Rev Sci Eau 10(3):279–320

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 Resources Res 52(4):2659–2681

Lathouris O (2020) *Lismore declared disaster area after $20m damage in NSW floods*, 2021, <https://www.9news.com.au/national/floods-storms-leave-massive-cleanup-nsw-queensland-weather/106a2e36-879e-4375-8167-200a616d1804>

Le Vine N (2016) Combining information from multiple flood projections in a hierarchical Bayesian framework. Water Resour Res 52(4):3258–3275

Lee O, Sim I, Kim S (2019) Application of the non-stationary peak-over-threshold methods for deriving rainfall extremes from temperature projections. J Hydrol 585:124318

Li Y, Cai W, Campbell EP (2004) Statistical modeling of extreme rainfall in southwest Western Australia. J Climate 18(6):852–863

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, Gregersen IB, Rosbjerg D, Arnbjerg-Nielsen K (2017) Regional frequency analysis of short duration rainfall extremes using gridded daily rainfall data as co-variate. Water Sci Technol 75(8):1971–1981

Madsen H, Mikkelsen PS, Rosbjerg D, Harremoës P (2002) Regional estimation of rainfall intensity-duration-frequency curves using generalized least squares regression of partial duration series statistics. Water Resource Res 38(11):21–11

Madsen H, Rasmussen PF, Rosbjerg D (1997) 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

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

Madsen H, Rosbjerg D (1997b) The partial duration series method in regional index-flood modeling. Water Resour Res 33(4):737–746

Madsen H, Rosbjerg D, Harremoës P (1994) PDS-modelling and regional bayesian estimation of extreme rainfalls. Hydrol Res 25(4):279–300

Madsen H, Rosbjerg D, Harremoöes P (1995) Application of the Bayesian approach in regional analysis of extreme rainfalls. Stoch Hydrol Hydraul 9(1):77–88

Mailhot A, Lachance-Cloutier S, Talbot G, Favre A-C (2013) Regional estimates of intense rainfall based on the Peak-Over-Threshold (POT) approach. J Hydrol 476:188–199

Martins ES, Stedinger JR (2001) Generalized maximum likelihood Pareto-Poisson estimators for partial duration series. Water Resour Res 37(10):2551–2557

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

Mohssen M (2009) 'Partial duration series in the annual domain', pp. 2694–700, Scopus

Mostofi Zadeh S, Burn DH (2017) 'Examination of pooled flood frequency analysis for Canadian catchments', pp. 145–152, Scopus

Mostofi Zadeh S, Durocher M, Burn DH, 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

Nagy BK, Mohssen M, Hughey KFD (2017) Flood frequency analysis for a braided river catchment in New Zealand: comparing annual maximum and partial duration series with varying record lengths. J Hydrol 547:365–374

Navratil O, Albert MB, Breil P (2010) Test of three methods to detect the overbank flow from water level time-series analysis. Hydrol Process 24(17):2452–2464

Ngongondo C, Li L, Gong L, Xu CY, Alemaw BF (2013) Flood frequency under changing climate in the upper Kafue River basin, southern Africa: a large scale hydrological model application. Stoch Env Res Risk Assess 27(8):1883–1898

Ngongondo C, Zhou Y, Xu CY (2020) Multivariate framework for the assessment of key forcing to Lake Malawi level variations in non-stationary frequency analysis. Environ Monitor Assess 192(9):1–23

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

O'Brien NL, Burn DH (2014) A nonstationary index-flood technique for estimating extreme quantiles for annual maximum streamflow. J Hydrol 519(PB):2040–2048

Önöz B, Bayazit M (2001) Effect of the occurrence process of the peaks over threshold on the flood estimates. J Hydrol 244(1–2):86–96

Page KJ, McElroy L (1981) Comparison of annual and partial duration series floods on the Murrumbidgee river 1 JAWRA. J Am Water Res Associat 17(2):286–289

Page K, Read A, Frazier P, Mount N (2005) The effect of altered flow regime on the frequency and duration of bankfull discharge: Murrumbidgee River, Australia. River Res Appl 21(5):567–578

Paixao E, Mirza MQ, Shephard MW, Auld H, Klaassen J, Smith G (2015) An integrated approach for identifying homogeneous regions of extreme rainfall events and estimating IDF curves in Southern Ontario, Canada: Incorporating radar observations. J Hydrol 528:734–750

Pan X, Rahman A (2018) Comparison of annual maximum and peaks-over-threshold methods in flood frequency analysis', In:Hydrology and Water Resources Symposium (HWRS 2018): Water and Communities (p. 614). Engineers Australia, pp. 614–25, Scopus

Pan X, Rahman A (2021) Comparison of annual maximum and peaks-over-threshold methods with automated threshold selection in flood frequency analysis: a case study for Australia. Natural Haz. https://doi.org/10.1007/s11069-021-05092-y

Parent E, Bernier J (2003) Bayesian POT modeling for historical data. J Hydrol 274(1):95–108

Phillips RC, Samadi SZ, Meadows ME (2018) How extreme was the October 2015 flood in the Carolinas? An assessment of flood frequency analysis and distribution tails. J Hydrol 562:648–663

Pickands J (1975) Statistical inference using extreme order statistics. Ann Statist 3(1):119–131

Rahman A, Haddad K, Kuczera G, Weinmann E (2019) 'Regional flood methods', In: Australian rainfall & runoff, commonwealth of Australia

Rahman A (2005) A quantile regression technique to estimate design floods for ungauged catchments in south-east Australia. Austral J Water Resource 9(1):81–89

Rahman AS, 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

Rasmussen PF (2001) Generalized probability weighted moments: application to the generalized Pareto distribution. Water Resource Res 37(6):1745–1751

Rasmussen PF, Rosbjerg D (1991) Prediction uncertainty in seasonal partial duration series. Water Resour Res 27(11):2875–2883

Read LK, Vogel RM (2015) Reliability, return periods, and risk under nonstationarity. Water Resour Res 51(8):6381–6398

Reed DW, Jakob D, Robson AJ, Faulkner DS, Stewart EJ (1999) Regional frequency analysis: a new vocabulary. IAHS AISH Publ 255:237–243

Renima M, Remaoun M, Boucefiane A, Abbes SB, A, (2018) Regional modelling with flood-duration-frequency approach in the middle Cheliff watershed. J Water Land Develop 36(1):129–141

Ribatet M, Sauquet E, Grésillon JM, Ouarda TBMJ (2007) A regional Bayesian POT model for flood frequency analysis. Stoch Env Res Risk Assess 21(4):327–339

Robson A, Reed D (1999) Statistical procedures for flood frequency estimation, flood estimation handbook, Centre for Ecology & Hydrology, Wallingford, UK

Rosbjerg D, Madsen H, Rasmussen PF (1992) Prediction in partial duration series with generalized pareto-distributed exceedances. Water Resource Res 28(11):3001–3010

Roth M, Buishand TA, Jongbloed G, Klein Tank AMG, Van Zanten JH (2012) A regional peaks-over-threshold model in a nonstationary climate. Water Resource Res. https://doi.org/10.1029/2012WR012214

Rustomji P (2009) A statistical analysis of flood hydrology and bankfull discharge for the daly river catchment, northern territory, Australia; 09/2009. CSIRO, Canberra, Australia, p 2009

Rutkowska A, Willems P, Niedzielski T (2017a) Relation between design floods based on daily maxima and daily means: use of the peak over threshold approach in the upper nysa kłodzka basin (SW Poland). Geomat Nat Haz Risk 8(2):585–606

Rutkowska A, Willems P, Onyutha C, Młocek W (2017b) Temporal and spatial variability of extreme river flow quantiles in the Upper Vistula River basin, Poland. Hydrol Process 31(7):1510–1526

Sabourin A, Renard B (2015) Combining regional estimation and historical floods: a multivariate semiparametric peaks-over-threshold model with censored data. Water Resour Res 51(12):9646–9664

Sccarrott C, Macdonald A (2012) A review of extreme value threshold estimation and uncertainty quantification. Stat J 10:33–60

Shu C, Burn DH (2004) Homogeneous pooling group delineation for flood frequency analysis using a fuzzy expert system with genetic enhancement. J Hydrol 291(1–2):132–149

Silva AT, Naghettini M, Portela MM (2016) On some aspects of peaks-over-threshold modeling of floods under nonstationarity using climate covariates. Stoch Env Res Risk Assess 30(1):207–224

Silva AT, Portela MM, Naghettini M (2012) Nonstationarities in the occurrence rates of flood events in Portuguese watersheds. Hydrol Earth Syst Sci 16(1):241–254

Silva AT, Portela MM, Naghettini M (2014) On peaks-over-threshold modeling of floods with zero-inflated Poisson arrivals under stationarity and nonstationarity. Stoch Env Res Risk Assess 28(6):1587–1599

Silva AT, Portela MM, Naghettini M, Fernandes W (2017) A Bayesian peaks-over-threshold analysis of floods in the Itajaí-açu River under stationarity and nonstationarity. Stoch Env Res Risk Assess 31(1):185–204

Solari S, Egüen M, Polo MJ, Losada MA (2017) Peaks Over Threshold (POT): a methodology for automatic threshold estimation using goodness of fit p-value. Water Resource Res 53(4):2833–2849

Solari S, Losada MA (2012) A unified statistical model for hydrological variables including the selection of threshold for the peak over threshold method. Water Resource Res. https://doi.org/10.1029/2011WR011475

Tavares LV, Da Silva JE (1983) Partial duration series method revisited. J Hydrol 64(1):1–14

Thiombiano AN, El Adlouni S, St-Hilaire A, Ouarda TB, El-Jabi N (2017) Nonstationary frequency analysis of extreme daily precipitation amounts in Southeastern Canada using a peaks-over-threshold approach. Theoret Appl Climatol 129(1):413–426

Thiombiano AN, St-Hilaire A, El Adlouni SE, Ouarda TB (2018) Nonlinear response of precipitation to climate indices using a non-stationary Poisson-generalized Pareto model: case study of southeastern Canada. Int J Climatol 38:e875–e888

Thompson P, Cai Y, Reeve D, Stander J (2009) Automated threshold selection methods for extreme wave analysis. Coast Eng 56(10):1013–1021

Tramblay Y, Neppel L, Carreau J, Najib K (2013) Non-stationary frequency analysis of heavy rainfall events in southern France. Hydrol Sci J 58(2):280–294

USWRC 1976, Guidelines for determining flood flow *frequency*, US Department of the Interior, Geological Survey, Office of Water Data Coordination

Villarini G, Smith JA, Serinaldi F, Bales J, Bates PD, Krajewski WF (2009) Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. Adv Water Resour 32(8):1255–1266

Vogel RM, Kroll CN (2020) A comparison of estimators of the conditional mean under non-stationary conditions. Adv Water Resource 143:103672

Vogel RM, McMahon TA, Chiew FHS (1993) Floodflow frequency model selection in Australia. J Hydrol 146(C):421–449

Vogel RM, Thomas WO, McMahon TA (1993b) Flood-flow frequency model selection in southwestern united states. J Water Resour Plan Manag 119(3):353–366

Vogel RM, Yaindl C, Walter MT (2011) Nonstationarity: flood magnification and recurrence reduction factors in the United States. JAWRA J Am Water Resource Associat 47(3):464–474

Walega A, Mlyński D, Bogdal A, Kowalik T (2016) Analysis of the course and frequency of high water stages in selected catchments of the upper Vistula basin in the south of Poland. Water (Switzerland) 8(9):394

Weiss J, Bernardara P, Benoit M (2012) 'Assessment of the regional frequency analysis to the estimation of extreme storm surges', Scopus

Yiou P, Ribereau P, Naveau P, Nogaj M, Brázdil R (2006) Statistical analysis of floods in Bohemia (Czech Republic) since 1825. Hydrol Sci J 51(5):930–945

Zaman M, Rahman A, Haddad K (2012) Regional flood frequency analysis in arid regions: a case study for Australia. J Hydrol 475:74–83

Zhang X, Duan K, Dong Q (2019) Comparison of nonstationary models in analyzing bivariate flood frequency at the Three Gorges Dam. J Hydrol 579:124208

Zhang Z, Stadnyk TA (2020) Investigation of attributes for identifying homogeneous flood regions for regional flood frequency analysis in Canada. Water (Switzerland) 12(9):2570

Zhang Z, Stadnyk TA, Burn DH (2020) Identification of a preferred statistical distribution for at-site flood frequency analysis in Canada. Canadian Water Resource J 45(1):43–58

Zhao L, Chen Z, Liu C, Xu Y (2019a) Flood sequence frequency analysis based on generalized Pareto distribution. Zhongshan Daxue Xuebao/acta Scientiarum Natralium Universitatis Suny-atseni 58(3):32–39

Zhao X, Zhang Z, Cheng W, Zhang P (2019b) A new parameter estimator for the generalized pareto distribution under the peaks over threshold framework. Mathematics 7(5):406

Zhou C, Chen Y, Huang Q, Gu S (2017) 'Higher moments method for generalized Pareto distribution in flood frequency analysis', vol. 82, Scopus

Zhou CR, Chen YF, Huang Q, Gu SH (2017) 'Higher moments method for generalized Pareto distribution in flood frequency analysis', vol. 82, Scopus

Zhou Z, Liu S, Hu Y, Liang Y, Lin H, Guo Y (2017c) Analysis of precipitation extremes in the Taihu Basin of China based on the regional L-moment method. Hydrol Res 48(2):468–479

Zoglat A, El Adlouni S, Badaoui F, Amar A, Okou CG (2014) Managing hydrological risks with extreme modeling: application of peaks over threshold model to the loukkos watershed, Morocco. J Hydrol Eng 19(9):05014010

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.