Stochastic Environmental Research and Risk Assessment (2026) 40:132
https://doi.org/10.1007/s00477-026-03248-5

ORIGINAL PAPER

Robust fitting of the generalized Pareto distribution for extreme precipitation modeling: a case study in Japan

Muhammad Aslam Mohd Safari[1,2,3] **· Tosiyuki Nakaegawa**[2] **· Nurulkamal Masseran**[4]

Received: 9 October 2024 / Accepted: 13 April 2026

Abstract

This study introduces a robust-efficient method for estimating the parameters of the generalized Pareto distribution (GPD), based on the probability integral transform. The probability integral transform estimator (PITE) is designed to enhance robustness in the presence of outliers, a frequent challenge in modeling extreme events. We also study the properties of PITE, including its efficiency and its robustness based on score function and breakdown point, demonstrating its ability to handle extreme data with minimal sensitivity to outliers. In addition, PITE offers computational simplicity, making it accessible for practical applications. Monte Carlo simulations indicate that the PITE family provides a flexible balance between robustness and efficiency: high-efficiency versions perform comparably to traditional estimators for uncontaminated data, while more robust versions often yield smaller deficiencies in simulated scenarios with data contamination. Applied alongside the peaks over threshold approach, PITE effectively models the tail behavior of extreme precipitation events. The method is applied to daily precipitation data from 12 meteorological stations in southern Japan, a region highly susceptible to extreme rainfall due to typhoons and complex climatic factors. Using PITE, GPD parameters are estimated, and return levels for 5-, 10-, 25-, 50-, 100-, and 200-year periods are calculated. The results reveal significant regional variability in extreme precipitation, with stations such as Naze, Miyazaki, and Kumamoto displaying particularly high return levels, indicative of their vulnerability to intense rainfall. The robust application of PITE provides critical insights for flood risk management in southern Japan, particularly in typhoon-prone settings where storm-driven extremes can act as potential outliers, highlighting the importance of localized strategies to mitigate the impact of extreme precipitation events.

Keywords Generalized Pareto distribution · Robust statistics · Extreme precipitation · Monte Carlo simulation · Return level

1 Introduction

The generalized Pareto distribution (GPD) is a key tool in extreme value theory (EVT), extensively used to model the tails of distributions through the peaks over threshold (POT) approach, where it analyzes exceedances over a specified threshold (Coles 2001). This makes the GPD invaluable across various fields where understanding and predicting rare, extreme events is critical. In finance, for instance, the GPD is used to estimate value-at-risk by modeling the distribution of extreme losses beyond a certain threshold, offering a robust approach to assessing the risk of significant financial downturns (Park and Kim 2016; He et al. 2022). In engineering, the GPD is used to assess the probability of material failure in high-stress conditions, particularly by modeling extreme fracture densities to improve predictions

✉ Muhammad Aslam Mohd Safari
aslam.safari@upm.edu.my

1 Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

2 Department of Applied Meteorology, Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan

3 Institute for Mathematical Research, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

4 Department of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

Published online: 22 May 2026

and risk management in highly fractured zones (Tanaka et al. 2024). In environmental sciences, the GPD is pivotal in analyzing extreme weather events, including high temperatures (Lazoglou et al. 2019), wind speeds (Jiang et al. 2023), and particularly precipitation (Mo et al. 2019; Tuel and El Moçayd 2023). By accurately modeling the frequency and magnitude of these extremes, the GPD provides essential insights for risk management and mitigation strategies, making it an indispensable tool in managing the impacts of extreme events.

In the context of Japan, where extreme weather events are frequent due to complex topography and climatic conditions, the GPD has been particularly effective for modeling extreme precipitation events. These events are critical for managing flood risks, planning water resources, and designing resilient infrastructure. Over the years, several studies have applied the GPD within the POT framework to analyze extreme precipitation in Japan, providing valuable insights into return levels and supporting flood management strategies. Tanaka and Takara (2001) conducted a comparative study between the annual maximum series (AMS) and partial duration series (PDS) in flood frequency analysis. They highlighted the advantages of using the PDS method (aligned with POT) over AMS for capturing extreme flood events. By employing the GPD for PDS data, they demonstrated its superiority in accurately modeling the tail behavior of the distribution, thus reinforcing the importance of the GPD in flood risk assessment. Chikamori and Nagai (2012) further utilized the POT method with GPD to estimate return levels of daily precipitation across 21 observatories in Japan. Their study underscored the effectiveness of the GPD in modeling extreme precipitation events and the critical role of precise threshold selection within the POT approach. Sugi et al. (2017) proposed the extended regional frequency analysis (ERFA) to improve the estimation of extreme rainfall probabilities across Japan. By aggregating data from meteorologically homogeneous regions, ERFA combined with GPD provided a reliable fit for return periods of 10 to 1000 years, enhancing the robustness of extreme precipitation modeling. Minakawa et al. (2018) further expanded this approach by generating numerous pseudo-climate scenarios and applying the POT method with GPD to assess the probabilistic distribution of extreme rainfall under future climate conditions, offering valuable insights for flood management, landslide prevention, and reservoir safety evaluations across Japan.

While the GPD has been widely adopted and proven effective in modeling extreme events, particularly through the POT approach, its application is not without challenges. One significant issue is the sensitivity of traditional GPD parameter estimation methods, such as maximum likelihood estimation (MLE) and method of moments (MOM), to outliers and data irregularities (de Zea Bermudez and Kotz 2010a, b). Outliers, which are data points that deviate significantly from the rest of the dataset (Aggarwal 2017), can occur due to rare, extreme occurrences or anomalies within the data. In extreme event datasets, such as those involving precipitation, these outliers can disproportionately influence the estimation of GPD parameters, leading to biased return level estimates and undermining the reliability of the model. Given these challenges, it is crucial to employ robust estimators that are less sensitive to outliers, ensuring more accurate and reliable modeling of extreme events. When outliers are present, as is often the case in extreme precipitation datasets, robust estimation methods must be used to maintain the integrity and accuracy of the GPD model.

Over the years, a range of robust estimation methods have been developed to address the sensitivity of traditional methods to outliers in extreme value data. Among these, the probability-weighted moments (PWM) by Hosking and Wallis (1987) and the method of medians (MM) originally proposed by He and Fung (1999) for lifetime data with Weibull models, and later adapted by Peng and Welsh (2001) for GPD, are notable early attempts to improve robustness. However, the PWM estimators, while straightforward, lack robustness as they are significantly influenced by outliers, as shown by Davison and Smith (1990). The MM, though robust, is relatively complex and has been noted to sacrifice efficiency. Another approach is the minimum-divergence procedure by Juárez and Schucany (2004), which, despite its robustness, can present computational challenges, particularly in large datasets. A notable development is the method of trimmed moments (MTM) proposed by Brazauskas and Kleefeld (2009), which enhances robustness by trimming extreme values. While MTM provides a good balance between robustness and efficiency, it faces limitations when applied to small sample sizes, where trimming outliers can further reduce the sample size and increase variance, potentially leading to less reliable estimates. In other work, Ahmad et al. (2011) introduced trimmed L-moments, particularly TL-moments (1,0), for the GPD, which reduces the influence of small or less relevant events and can improve bias and RMSE for many quantiles. However, their results also show that TL-moments (1,0) may yield poorer RMSE for some high quantiles and small samples, and the choice of trimming scheme is not always straightforward. In parallel, Ruckdeschel and Horbenko (2013) developed optimally robust one-step estimators i.e., most bias robust estimator (MBRE) and optimal MSE-robust estimator (OMSE), for generalized Pareto models, designed to minimize worst-case bias or mean squared error under small contamination by using bounded influence functions built on a high-breakdown initial estimator, at the cost of more involved optimization and tuning.

Despite these advancements, there remains a need for a robust estimator that not only effectively addresses the influence of outliers but also maintains computational efficiency and ease of implementation. This is particularly critical in the context of extreme precipitation modeling in Japan, where data variability and the presence of anomalies are common challenges. To address this, our study introduces the probability integral transform estimator (PITE) for the GPD. PITE leverages the probability integral transform, a method that transforms a continuous random variable into a standard uniform random variable, thereby enhancing robustness against outliers while remaining computationally simple and easy to implement. We evaluate the relative efficiency of PITE in comparison to MLE and assess its robustness through measures such as the breakdown point and score function. Additionally, through Monte Carlo simulations, we compare the performance of PITE with several well-known estimators, both in the presence and absence of outliers, demonstrating its reliable performance across various scenarios.

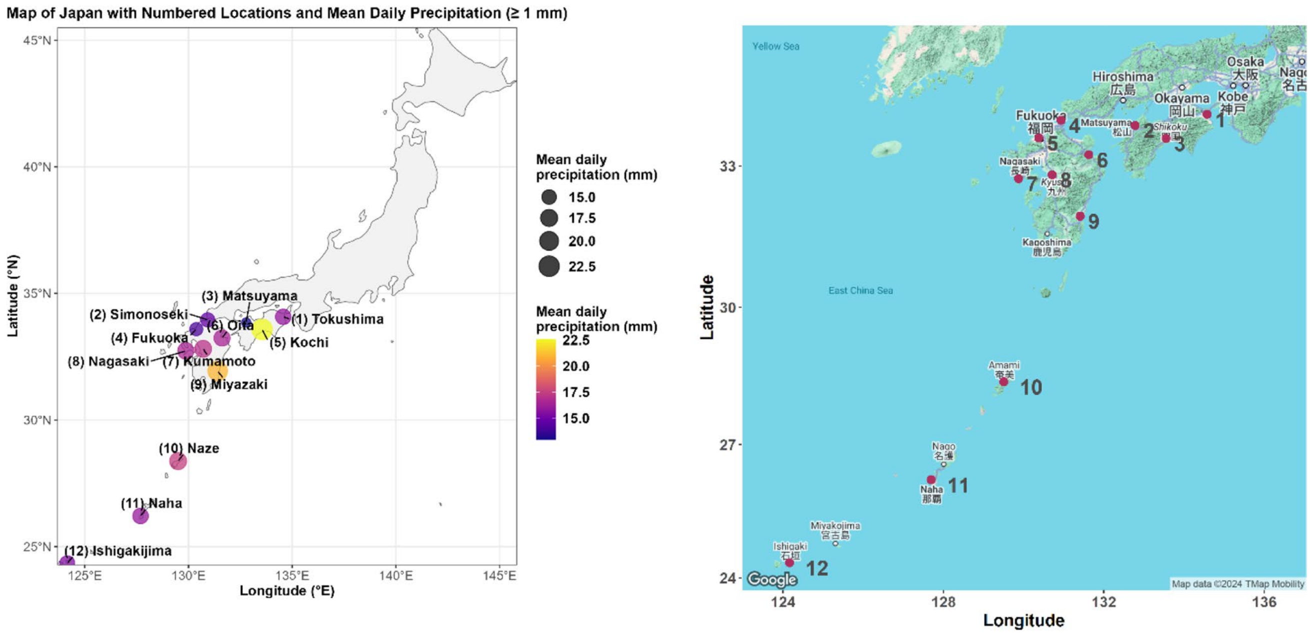
In the context of this study, we apply the PITE to analyze extreme precipitation events in the southern region of Japan, an area particularly susceptible to intense rainfall due to its intricate climatic and topographical characteristics (Nakaegawa and Murazaki 2022; Endo 2023). Using the POT approach, we identify exceedances above a selected threshold to model the tail behavior of precipitation distributions. The PITE is employed to estimate the GPD parameters for each location under consideration, ensuring robustness against outliers and variability within the data. Following the estimation of GPD parameters, we compute return levels for various return periods across the selected sites. These return levels provide critical insights into the potential magnitude of future extreme precipitation events, which are essential for informing flood risk management strategies and enhancing the resilience of infrastructure in the region. The application of PITE in this context not only validates its robustness and efficiency but also underscores its utility in practical environmental risk assessment.

2 Data and study area

Japan experiences a wide range of climatic conditions due to its geographical location and complex topography, with distinct seasonal weather patterns that vary across the country. The southern region of Japan, particularly areas like Kyushu and Shikoku, is highly susceptible to extreme precipitation events (Tsuguti et al. 2019). This vulnerability is compounded by seasonal phenomena such as the Baiu and Akisame rainy seasons, as well as frequent typhoons, which bring intense rainfall and severe weather conditions (Endo 2023; Kawase et al. 2023). Typhoons, in particular, contribute significantly to the region's extreme precipitation, often resulting in floods, landslides, and infrastructure damage (Nayak and Takemi 2020; Fujiwara et al. 2024). Given the severity of these events, studying this region is crucial for improving flood risk management and developing strategies to mitigate the impact of extreme rainfall.

The dataset used in this study consists of daily precipitation records obtained from the Japan Meteorological Agency (JMA), covering a 120-year period from 1901 to 2020. The data were collected from twelve meteorological stations located in the southern region of Japan. Each location provides a continuous time series of precipitation measurements, allowing for a detailed analysis of extreme precipitation events. These stations were selected based on their comprehensive historical records and strategic locations in areas frequently affected by extreme precipitation, including rainfall induced by typhoons. Although all stations are located in southern Japan, they span a relatively wide latitudinal range and include both mainland and island sites, which leads to differences in rainfall regimes. The Kyushu and Shikoku stations are influenced by the Baiu and Akisame seasonal rain fronts and by typhoon landfalls, with local topography contributing to rainfall enhancement in some areas. In contrast, the southern island stations in the Nansei island chain experience more frequent tropical cyclone influence. Previous studies have shown that tropical cyclone activity is concentrated in the western North Pacific and that the contribution of tropical cyclones to precipitation is particularly important over southern Japan and its surrounding islands (Takaya et al. 2010; Kamahori and Arakawa 2018). The locations and detailed information of all stations considered in this study are provided in Fig. 1 and Table 1.

Summary statistics of the precipitation data for each station, focusing on days with daily precipitation ≥ 1 mm, are presented in Table 1. This threshold of 1 mm is based on the simple precipitation intensity index (SDII), which defines wet days as those with at least 1 mm of rainfall, ensuring that the analysis focuses on meaningful precipitation events (Oguchi and Fujibe 2012). Key metrics include the mean, median, maximum daily precipitation, variability, skewness, and excess kurtosis. These statistics provide valuable insights into the general precipitation patterns and distribution characteristics at each location. In all stations, the mean exceeds the median, indicating a right-skewed distribution, which is further corroborated by the positive skewness values observed at each station. This reflects the predominance of moderate precipitation events, with a smaller number of extreme precipitation events pulling the distribution to the right. The high excess kurtosis values, suggest that the precipitation data are heavy-tailed, meaning there is a higher



(a) (b)

Fig. 1 **a** Map of Japan showing the numbered locations of the 12 meteorological stations used in this study, with colors and symbol sizes representing mean daily precipitation (≥1 mm). **b** Zoomed-in map of southern Japan highlighting the specific station locations in typhoon-prone areas (Google 2024)

Table 1 Location and summary statistics of daily precipitation data (≥1 mm) for 12 meteorological stations in the southern region of Japan (1901–2020)

Location	Longitude (°E)	Latitude (°N)	Elevation (m)	Mean (mm)	Median (mm)	Maximum (mm)	Variance	Skewness	Excess Kurtosis
Tokushima	134.57	34.07	2	15.89	7.50	463.40	633.20	4.82	39.23
Matsuyama	132.78	33.84	32	12.91	7.20	215.10	274.14	3.46	19.28
Kochi	133.55	33.57	2	22.55	11.00	628.50	1051.58	3.89	28.23
Simonoseki	130.93	33.95	3	14.72	7.30	336.70	417.21	3.50	19.68
Fukuoka	130.38	33.58	3	14.06	7.00	307.80	409.24	3.95	24.72
Oita	131.62	33.23	5	16.44	8.40	443.70	594.69	4.84	40.60
Nagasaki	129.87	32.73	27	16.55	7.70	448.00	597.42	4.19	31.53
Kumamoto	130.71	32.81	38	17.04	8.50	480.50	629.87	4.56	37.60
Miyazaki	131.41	31.94	9	20.95	10.50	587.20	880.97	4.14	32.23
Naze	129.5	28.38	3	17.42	8.00	622.00	843.92	5.56	54.67
Naha	127.69	26.21	3	15.92	6.50	468.90	682.77	4.73	37.66
Ishigakijima	124.16	24.34	11	15.43	6.10	378.90	685.53	4.44	28.37

likelihood of extreme precipitation events compared to what would be expected in a normal distribution. The presence of these extremes in the data highlights the variability and intensity of precipitation in the southern region of Japan, which is critical for the subsequent analysis of extreme rainfall and the estimation of return levels.

3 POT approach and GPD

In EVT, the POT approach is used to model rare and extreme events by focusing on exceedances over a specified threshold. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (IID) random variables representing daily precipitation values, governed by an unknown distribution function F. We are primarily interested in values of X_i that exceed a high threshold u.

Let $Y = X - u$ represent the exceedances over the threshold u, i.e., the amount by which a precipitation value X exceeds u. The conditional distribution of Y, given that

$X > u$, is referred to as the conditional excess distribution function and is given by:

$$F_u(y) = P(X - u \leq y | X > u) = \frac{F(u + y) - F(u)}{1 - F(u)}, y \geq 0. \quad (1)$$

This equation describes the distribution of exceedances above the threshold u, conditional on the fact that an exceedance has occurred. The conditional distribution $F_u(y)$ represents the distribution of extreme precipitation values that exceed the threshold.

According to EVT, for sufficiently high thresholds u, the limiting distribution of exceedances $Y = X - u$, as $u \to \infty$ as can be approximated by a GPD. More formally, Pickands (1975) and Balkema and De Haan (1974) showed that if F is in the domain of attraction of an extreme value distribution, then for a sufficiently large threshold u, the distribution of the exceedances follows a GPD. The cumulative distribution function (CDF) of the GPD for the exceedances can be expressed as follows for the random variable X:

$$F(x;\sigma,\xi) = \begin{cases} 1 - \left(1 + \frac{\xi(x-u)}{\sigma}\right)^{-\frac{1}{\xi}} \text{for } \xi \neq 0, \\ 1 - exp\left(-\frac{(x-u)}{\sigma}\right) \text{for } \xi = 0. \end{cases} \quad (2)$$

In Eq. (2), the scale parameter $\sigma > 0$ controls the spread or dispersion of the exceedances, determining the degree of variability in the extreme values. The shape parameter ξ, governs the tail behavior and dictates the nature of the distribution. Specifically, when $\xi > 0$ the distribution has a heavy tail, meaning extreme values can be very large, as seen in the Pareto distribution. When $\xi = 0$, the distribution has a light, exponentially decaying tail. For $\xi < 0$, the distribution is bounded, with extreme values capped at a certain limit. The support of X depends on the value of ξ : for $\xi \geq 0$, $x \geq u$; and for $\xi < 0$, the support is bounded as $u \leq x \leq u - \sigma/\xi$.

A key result from fitting the GPD to exceedances is the estimation of return levels. A return level is the value expected to be exceeded on average once every t observations. For practical purposes, we often express return levels on an annual scale. Let $n_y = 365.25$ represent the number of daily observations per year (considering leap years). The return period t in terms of years, denoted as r, is given by $t = r \times n_y$. The return level $\widehat{z}_t$ for a given return period t can be expressed in terms of the GPD parameters as follows (Masseran and Safari 2020):

$$\widehat{z}_t = \begin{cases} \widehat{u} + \frac{\widehat{\sigma}}{\widehat{\xi}}\left[\left(t\widehat{\lambda}\right)^{\widehat{\xi}} - 1\right] \text{for } \xi \neq 0, \\ \widehat{u} + \widehat{\sigma} log(t\widehat{\lambda}) \text{ for } \xi = 0. \end{cases} \quad (3)$$

To express the return level in terms of years, where r is the number of years, we adjust for the number of observations per year. The return level $\widehat{z}_r$ for an r-year return period can then be written as:

$$\widehat{z}_r = \begin{cases} \widehat{u} + \frac{\widehat{\sigma}}{\widehat{\xi}}\left[\left(rn_y\widehat{\lambda}\right)^{\widehat{\xi}} - 1\right] \text{for } \xi \neq 0, \\ \widehat{u} + \widehat{\sigma} log(rn_y\widehat{\lambda}) \text{ for } \xi = 0. \end{cases} \quad (4)$$

Here, $\widehat{u}$, $\widehat{\sigma}$, and $\widehat{\xi}$ are the estimated threshold, scale and shape parameters, respectively. $\widehat{\lambda}$ is the rate of exceedances which defined as the number of observations above the threshold u divided by the total number of observations. The return level provides insights into the magnitude of extreme precipitation events expected over specific return periods (e.g., 10, 50, or 100 years), crucial for flood risk management and infrastructure resilience.

4 A robust-efficient estimator for GPD parameters

This section introduces a robust and efficient estimator for GPD parameters, known as PITE. Building on the probability integral transform statistic, PITE is designed to improve robustness while maintaining computational efficiency. Our focus is primarily on the general case where the GPD shape parameter $\xi \neq 0$, which presents unique challenges due to the heavy-tailed nature of the distribution, especially when $\xi > 0$. The key properties of this new estimator, such as its efficiency, bounded score function, and high breakdown point are examined in detail, highlighting its capability in handling outliers in GPD modeling.

4.1 PITE for GPD

Let $x_1, x_2, \ldots, x_n$ be a random sample from the GPD, which is characterized by its CDF given in Eq. (2). The CDF of the GPD is continuous and strictly increasing, which allows us to apply the probability integral transform. Specifically, when we transform the random variables using the GPD CDF, $F(x_i)$, the transformed variables $F(x_1), F(x_2), \ldots, F(x_n)$ follow the standard uniform distribution $U(0,1)$, i.e., $F(x_i) \sim U(0,1)$. In the same manner, the complementary CDF, $1 - F(x_i)$, also follows a uniform distribution, i.e., $1 - F(x_i) \sim U(0,1)$. Building upon this, we introduce the PITE for the GPD. Given a known threshold u, PITE is generally defined as:

$$P_{n,\tau}(\sigma,\xi) = \frac{1}{n}\sum_{i=1}^{n}\left(1 + \frac{\xi(x_i - u)}{\sigma}\right)^{-\frac{\tau}{\xi}}. \quad (5)$$

In this equation, τ is a positive tuning parameter that balances efficiency and robustness during estimation. When $\tau = 1$, the probability integral transform simplifies, and the transformed variables can be treated as standard uniformly distributed random variables. Using the strong law of large numbers, we can demonstrate that

$$\frac{1}{n}\sum_{i=1}^{n} u_i^{\tau} \to E\left[U^{\tau}\right] = \frac{1}{1+\tau}.$$

Thus, the expected value of $U^{2\tau}$ is

$$E\left[U^{2\tau}\right] = \frac{1}{1+2\tau}.$$

Based on this, we define PITE by solving the following equations with respect to σ and ξ:

$$P_{n,\tau}\left(\sigma,\xi\right) = \frac{1}{n}\sum_{i=1}^{n}\left(1+\frac{\xi(x_i-u)}{\sigma}\right)^{-\frac{\tau}{\xi}} = \frac{1}{1+\tau}, \tag{6}$$

$$P_{n,2\tau}\left(\sigma,\xi\right) = \frac{1}{n}\sum_{i=1}^{n}\left(1+\frac{\xi(x_i-u)}{\sigma}\right)^{-\frac{2\tau}{\xi}} = \frac{1}{1+2\tau}. \tag{7}$$

To estimate σ and ξ, these equations are solved concurrently using numerical methods. It is worth noting that this technique can also be applied in the case where $\xi = 0$ (i.e., the exponential case). In this case, the PITE is simplified to a single equation, and the complementary CDF needs to be adjusted for the exponential distribution.

Table 2 Tuning parameter τ values for the PITE estimators corresponding to different shape parameters ξ and target ARE levels, where the ARE is computed with respect to the MLE

ARE (%)	Tuning parameter (τ)				
	$\xi = -0.2$	$\xi = -0.1$	$\xi = 0.1$	$\xi = 0.3$	$\xi = 0.5$
99	0.0001	0.02	0.09	0.17	0.23
90	0.09	0.16	0.25	0.36	0.45
85	0.14	0.22	0.33	0.45	0.55
80	0.19	0.28	0.4	0.53	0.65
75	0.25	0.34	0.47	0.62	0.75
70	0.3	0.4	0.55	0.71	0.85
65	0.36	0.47	0.64	0.81	0.96
60	0.43	0.55	0.73	0.92	1.08
55	0.5	0.63	0.83	1.04	1.21
50	0.59	0.73	0.95	1.17	1.36
45	0.68	0.84	1.08	1.33	1.53
40	0.8	0.97	1.24	1.51	1.72

4.2 Relative efficiency of PITE

The MLE is widely recognized for its efficiency, making it an ideal benchmark for comparing the efficiency of alternative estimators. The asymptotic relative efficiency (ARE) measures the efficiency of the PITE for GPD parameters in comparison to the MLE. The ARE of PITE with respect to MLE can be expressed as follows:

$$ARE\left(\left(\widehat{\sigma}_{PITE},\widehat{\xi}_{PITE}\right),\left(\widehat{\sigma}_{MLE},\widehat{\xi}_{MLE}\right)\right) = \left(\frac{|\sum_{MLE}|}{|\sum_{PITE}|}\right)^{1/2}, \tag{8}$$

where $|\sum|$ denotes the determinant of the covariance matrix. The covariance matrix for the MLE is typically available from the literature (Brazauskas and Kleefeld 2009), while for PITE, we estimate it using a Monte Carlo simulation approach. This allows us to compute the ARE for various values of the tuning parameter τ, which controls the balance between robustness and efficiency.

As mentioned earlier, the balance between efficiency and robustness in PITE can be adjusted by modifying the value of τ. Increasing τ enhances robustness but reduces the relative efficiency, as indicated by the ARE. Conversely, by selecting a τ value close to zero, the ARE of PITE can be brought closer to 1, making it nearly as efficient as the MLE. It is important to note that the optimal choice of τ also depends on the value of the shape parameter ξ. For different values of ξ, the tuning parameter τ must be adjusted accordingly to maintain the desired balance between robustness and efficiency. Table 2 presents the tuning parameter values for different shape parameters, along with their corresponding ARE levels, as obtained through simulation.

For comparison, the ARE of PITE relative to the L-moment (LMOM) estimator was computed using the same expression in Eq. (12) but replacing the MLE covariance matrix $\sum_{MLE}$ with the covariance matrix of the LMOM estimator $\sum_{LMOM}$; the resulting values are reported in Supplementary Table S1. These results show similar patterns, with high-efficiency PITE variants retaining substantial efficiency relative to LMOM while allowing increased robustness for larger values of τ. In general, for a given target ARE, the corresponding tuning parameters τ in Table S1 tend to be larger than those in Table 2, reflecting the lower asymptotic efficiency of LMOM compared with MLE. Equivalently, for a fixed τ, the ARE of PITE relative to LMOM is typically higher than the ARE relative to MLE, indicating that similar nominal efficiency can be achieved while applying stronger downweighting when efficiency is measured with respect to LMOM.

4.3 Score function of PITE

A desirable robustness criterion for an estimator is that it has a bounded score function. The score functions for PITE are as follows:

$$\Psi_1\left(x;\sigma,\xi\right)=\left(1+\frac{\xi\left(x-u\right)}{\sigma}\right)^{-\frac{\tau}{\xi}}-\frac{1}{1+\tau}, \tag{9}$$

$$\Psi_2\left(x;\sigma,\xi\right)=\left(1+\frac{\xi\left(x-u\right)}{\sigma}\right)^{-\frac{2\tau}{\xi}}-\frac{1}{1+2\tau}. \tag{10}$$

Both score functions, given by Eqs. (9) and (10), are bounded because the terms $(1+\xi(x-u)/\sigma)^{(-\tau/\xi)}$ and $(1+\xi(x-u)/\sigma)^{(-2\tau/\xi)}$ approach finite limits as x increases. The constants $1/(1+\tau)$ and $1/(1+2\tau)$ further ensure that the score functions do not tend to infinity or negative infinity, thereby guaranteeing robustness.

To further illustrate the robustness of PITE, we analyze the behavior of its score functions as x increase from 1 to an upper bound, which varies depending on the shape parameter ξ. For this analysis, the shape parameter takes values $\xi=-0.1$, 0.1, 0.3, 0.5, the scale parameter is set to $\sigma=2$ and the threshold is set to $u=1$. Additionally, we plot six variations of PITE, corresponding to ARE of 50%, 60%, 70%, 80%, 90%, and 99%. In Figs. 2 and 3, the score functions for PITE are shown for different values of ξ, with x starting from 1. The results confirm that the score functions remain bounded and do not exhibit extreme or unbounded behavior, demonstrating that PITE is not overly sensitive to individual data points, even as the outlier values increase. The figures clearly illustrate that PITE's score functions remain within a finite range as x increases, validating the theoretical expectation of boundedness. Regardless of the value of the shape parameter ξ, this behavior demonstrates the estimator's ability to handle extreme values effectively, contributing to its overall stability.

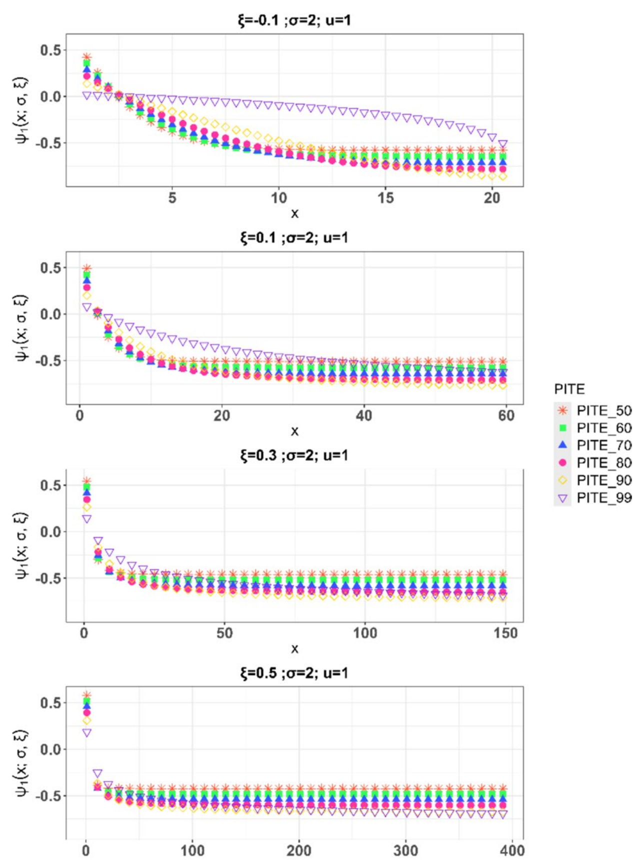


Fig. 2 The graph of $\Psi_1\left(x;\sigma,\xi\right)$ for several AREs with parameters $\xi=-0.1, 0.1, 0.3, 0.5$, $\sigma=2$, and $u=1$. The bounded curves illustrate that the influence of individual large observations on the estimator remains limited, highlighting the robustness of PITE

4.4 Breakdown point of PITE

The breakdown point (BP) is a key measure of robustness for statistical estimators, reflecting their ability to resist the influence of data contamination. Introduced by Huber (1981), the BP is defined as the maximum proportion of contaminated data that an estimator can handle before it breaks down or produces unreliable results. A higher BP indicates greater robustness to outliers and data anomalies. In the context of estimating GPD parameters, robustness is assessed through two types of breakdown points: the upper breakdown point (UBP) and the lower breakdown point (LBP). The UBP refers to the maximum fraction of upper contamination (i.e., large values) that the estimator can handle before it loses accuracy. Similarly, the LBP represents the highest fraction of lower contamination (i.e., small values) that the estimator can tolerate before it begins to fail.

The derivation of the UBP and LBP for the PITE begins with the fundamental equations that define the estimator. Specifically, for any integer m between 1 and n, the PITE is represented by two equations:

$$\frac{1}{n}\sum_{i=1}^{m}\left(1+\frac{\widehat{\xi}(x_i-u)}{\widehat{\sigma}}\right)^{-\frac{\tau}{\widehat{\xi}}}+\frac{1}{n}\sum_{i=1+m}^{n}\left(1+\frac{\widehat{\xi}(x_i-u)}{\widehat{\sigma}}\right)^{-\frac{\tau}{\widehat{\xi}}}=\frac{1}{1+\tau}$$

$$\frac{1}{n}\sum_{i=1}^{m}\left(1+\frac{\widehat{\xi}(x_i-u)}{\widehat{\sigma}}\right)^{-\frac{2\tau}{\widehat{\xi}}}+\frac{1}{n}\sum_{i=1+m}^{n}\left(1+\frac{\widehat{\xi}(x_i-u)}{\widehat{\sigma}}\right)^{-\frac{2\tau}{\widehat{\xi}}}=\frac{1}{1+2\tau}$$

These equations balance the contributions of the smallest m observations and the largest $n-m$ observations in the sample, scaled by the parameter τ and the estimated parameters $\widehat{\sigma}$ and $\widehat{\xi}$.

To derive the UBP, consider the case where all $x_1, x_2, \ldots, x_m \to \infty$. When this happens, the terms

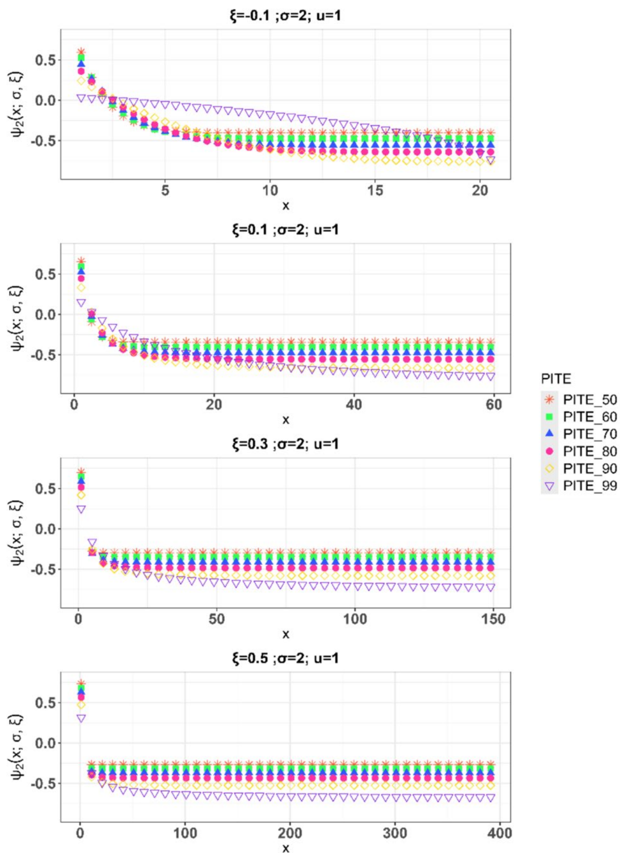


Fig. 3 The graph of $\Psi_2\left(x;\sigma,\xi\right)$ for several AREs with parameters $\xi=-0.1,\,0.1,\,0.3,\,0.5$, $\sigma=2$, and $u=1$. As in Fig. 2, the bounded behaviour of the score function as x increases demonstrates that PITE is not overly sensitive to extreme values and thus has good robustness properties

$(1+\widehat{\xi}(x_i-u)/\widehat{\sigma})^{(-\tau/\widehat{\xi})}$ and $(1+\widehat{\xi}(x_i-u)/\widehat{\sigma})^{(-2\tau/\widehat{\xi})}$ approach 0 for all $i=1,2,\ldots,m$. This leads to the first parts of both equations becoming negligible, leaving us with the following simplified forms:

$$\frac{1}{\tau+1}\approx\frac{1}{n}\sum_{i=1+m}^{n}\left(1+\frac{\widehat{\xi}(x_i-u)}{\widehat{\sigma}}\right)^{-\frac{\tau}{\widehat{\xi}}}<\frac{n-m}{n},$$

$$\frac{1}{2\tau+1}\approx\frac{1}{n}\sum_{i=1+m}^{n}\left(1+\frac{\widehat{\xi}\,(x_i-u)}{\widehat{\sigma}}\right)^{-\frac{2\tau}{\widehat{\xi}}}<\frac{n-m}{n}.$$

Solving these inequalities for m gives us $m<n\tau/(1+\tau)$ and $m<2n\tau/(1+2\tau)$. Since $n\tau/(1+\tau)<2n\tau/(1+2\tau)$, the tighter bound is $m<n\tau/(1+\tau)$. Thus, the finite sample UBP is given by:

$$UBP=\frac{\lceil n\tau/(1+\tau)\rceil}{n}, \tag{11}$$

where $\lceil\cdot\rceil$ denotes the ceiling function, which rounds up to the nearest integer. As n becomes large, this expression converges to $\tau/(1+\tau)$, which represents the UBP for large sample sizes.

To find the LBP, consider the opposite scenario where all $x_1,x_2,\ldots,x_m\to u$. In this case, the terms $(1+\widehat{\xi}(x_i-u)/\widehat{\sigma})^{(-\tau/\widehat{\xi})}$ and $(1+\widehat{\xi}(x_i-u)/\widehat{\sigma})^{(-2\tau/\widehat{\xi})}$ close to 1 for all $i=1,2,\ldots,m$, which affects the first parts of the equations significantly. Under these conditions, the equations become:

$$\frac{1}{\tau+1}\approx\frac{m}{n}+\frac{1}{n}\sum_{i=1+m}^{n}\left(1+\frac{\widehat{\xi}\,(x_i-u)}{\widehat{\sigma}}\right)^{-\frac{\tau}{\widehat{\xi}}}>\frac{m}{n},$$

$$\frac{1}{2\tau+1}\approx\frac{m}{n}+\frac{1}{n}\sum_{i=1+m}^{n}\left(1+\frac{\widehat{\xi}\,(x_i-u)}{\widehat{\sigma}}\right)^{-\frac{2\tau}{\widehat{\xi}}}>\frac{m}{n}.$$

Solving for m gives $m<n/(1+2\tau)$. Therefore, the finite sample LBP is calculated as:

$$LBP=\frac{\lceil n/(1+2\tau)\rceil}{n}, \tag{12}$$

which converges to $1/(1+2\tau)$ as n becomes large, representing the LBP for large samples.

In most typical applications, including POT analysis, contamination related to the upper end of the distribution is a key concern, which is reflected in the UBP. In the context of modeling extreme events, such as heavy precipitation, a high UBP ensures that the estimator remains robust even when dealing with potential outliers from the upper tail of the distribution. Although infrequent, these outliers can have a substantial impact on the analysis. By increasing the tuning parameter τ, we can further strengthen the robustness of PITE, raising the UBP to better handle these extreme outliers without compromising the reliability of the estimates. It is worth noting, however, that the theoretical maximum BP of any estimator is 50%: once more than half of the observations are contaminated, no procedure can reliably distinguish signal from noise. Thus, while PITE can be tuned to achieve a high degree of robustness, its BP, like that of any estimator, cannot exceed this fundamental upper limit.

5 Simulation study

In this section, we evaluate the performance of PITE at various ARE levels: 99%, 90%, 85%, 80%, 75%, 65%, 55%, and 45%, which are referred to as PITE 1–8, respectively. Additionally, we compare these with MLE, PWM, LMOM, and three variations of the MTM method (MTM 1–3), which differ based on the degree of trimming. The MTM variations differ based on the trimming parameters: MTM 1 uses trimming parameters $(a_1, b_1) = (0.05, 0.70)$ and $(a_2, b_2) = (0.70, 0.05)$, MTM 2 applies $(a_1, b_1) = (0.15, 0.65)$ and $(a_2, b_2) = (0.80, 0.10)$, and MTM 3 uses $(a_1, b_1) = (0.33, 0.50)$ and $(a_2, b_2) = (0.70, 0.15)$. These trimming parameters control the degree of trimming applied to the data for robust estimation, with each variation having different ARE levels based on the shape parameter ξ. For further details on MTM variations, refer to Brazauskas and Kleefeld (2009). The evaluation is conducted using Monte Carlo simulations, both with and without the presence of outliers.

5.1 Simulation design

The Monte Carlo simulation follows a structured approach. First, a random sample is generated from a GPD with sample sizes n = 50, 100, 500, 1000. The shape parameter is set to $\xi = -$ 0.1, 0.1, 0.3, 0.5, and the scale parameter is fixed at $\sigma = 2$, while the threshold is set to $u = 1$. Note that these shape parameter values reflect typical values in hydrological applications (de Zea Bermudez and Kotz 2010a). Next, outliers are introduced by randomly replacing some of the observations with values drawn from a normal distribution $N(\mu,\sigma_*^2)$. These outliers are generated with a mean μ = 13.04, 31.24, 99.99, 397.00, and a standard deviation $\sigma_* =$ 1. The outliers correspond to the 99.99% quantile of the GPD for the true values of ξ= – 0.1, 0.1, 0.3, 0.5, with $\sigma =$ 2 and $u = 1$. The contamination proportions for the outliers are set to ω = 0%, 1%, 3%, 5%, 8%. These proportions represent none to high levels of contamination, allowing the impact of replacing a small fraction of observations on estimator performance to be systematically examined. After generating the data and introducing outliers, the parameters σ and ξ are estimated using the considered methods (PITE, MLE, PWM, LMOM, and MTM). This process is repeated N = 10,000 times to ensure reliable simulation results.

The performance of each estimator is then assessed using the mean square error (MSE) and deficiency (Def) criterion. Given the true parameters σ and ξ, the MSE and Def are calculated as follows:

$$MSE(\widehat{\sigma}) = \frac{1}{N}\sum\nolimits_{i=1}^{N}(\widehat{\sigma}_i - \sigma)^2, \tag{13}$$

$$MSE(\widehat{\xi}) = \frac{1}{N}\sum\nolimits_{i=1}^{N}\left(\widehat{\xi}_i - \xi\right)^2, \tag{14}$$

$$Def\left(\widehat{\sigma}, \widehat{\xi}\right) = \frac{MSE\left(\widehat{\sigma}\right) + MSE\left(\widehat{\xi}\right)}{2}. \tag{15}$$

In these equations, $\widehat{\sigma}_i$ and $\widehat{\xi}_i$ are the estimated scale and shape parameters for the i—th simulated dataset, and N is the total number of simulations. Smaller values of MSE and Def indicate higher accuracy and precision.

5.2 Simulation results

The simulation results are presented in Supplementary Figures S1 and S2, which show the MSE values for the scale and shape parameters, and in Fig. 4, which presents heatmaps of the Def for each estimator across all combinations of sample size, tail index and contamination level. Table 3 provides a summary of which estimators perform best in each regime. Several general patterns emerge.

First, for all estimation methods, the MSE values for both σ and ξ, as well as the Def values, decrease as the sample size n increases, while they tend to increase as the shape parameter ξ becomes larger. This behaviour reflects the usual gain in precision with larger samples and the increased variability associated with heavier tails.

Second, in the absence of contamination (ω = 0%), the classical estimators MLE, PWM and LMOM perform well in estimating both σ and ξ, often achieving the smallest Def values, especially for small and moderate sample sizes. In these uncontaminated settings, the high-efficiency PITE variants (PITE 1 and PITE 2) behave similarly to the classical methods, whereas the MTM estimators perform less well because their trimming is mainly designed for contaminated data.

Third, when a small to moderate proportion of outliers is introduced (ω = 1–3%), the relative performance changes. For short-tailed cases ($\xi = -0.1$) and small sample sizes ($n \leq$ 100), PITE 1–3 remain competitive with PWM and LMOM. As the tail becomes heavier ($\xi \geq 0.1$) and the sample size increases, mid- to high-efficiency PITE variants (PITE 3–8) frequently attain the lowest Def values, while MLE, PWM and LMOM begin to deteriorate, particularly for $\xi \geq 0.3$.

Finally, for higher contamination levels ($\omega = 5 - 8\%$) , robust estimators become essential. For small samples with $\xi = -0.1$, PITE 1–4, PWM and LMOM still provide reasonable performance, but for heavier tails and larger samples (n = 500 and 1000), the most robust PITE variants (PITE 6–8) and the MTM 2–3 generally show the best performance, with PWM and LMOM remaining competitive mainly in the light-tailed case $\xi = -0.1$. Overall, Fig. 4 and Table 3 show that the PITE family provides a flexible range of estimators that can be tuned to balance efficiency for clean data against robustness when a small proportion of observations are outliers. These findings guide the choice of

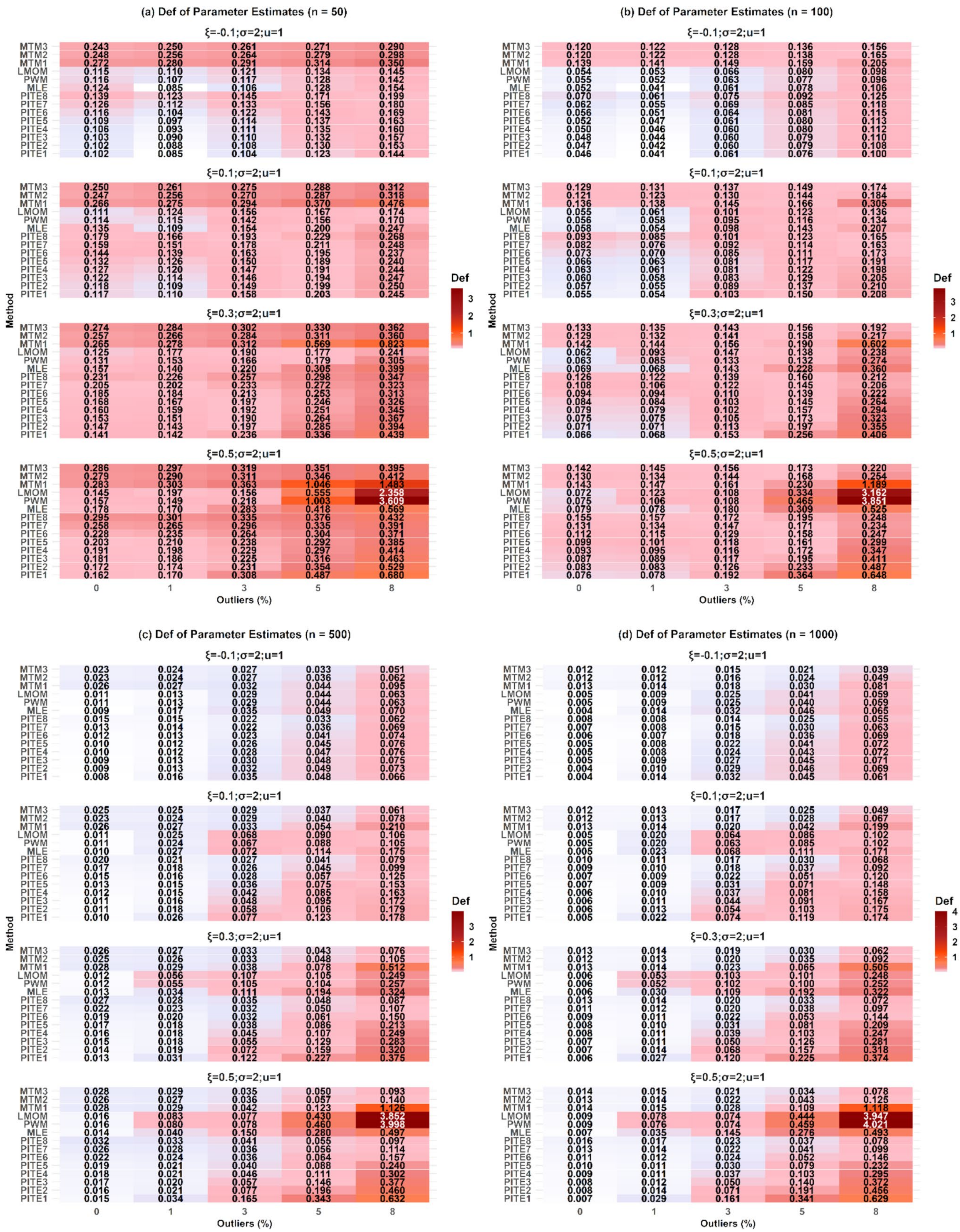


Fig. 4 Def values corresponding to $\xi = -0.1, 0.1, 0.3, 0.5$, $\sigma = 2$, and $u = 1$ with ω = 0%, 1%, 3%, 5%, 8% for different estimators. Results are shown for sample sizes **a** $n = 50$, **b** $n = 100$, **c** $n = 500$, and **d** $n = 1000$. Smaller Def values (lighter shading) indicate better accuracy

Table 3 Summary of simulation results (Def) for estimator performance across contamination levels, sample sizes, and tail indices. "Best" denotes estimators that most frequently achieve the smallest Def within each regime, while "comp." indicates comparable performance

Contamination proportion (ω)	Small ($n = 50$) and moderate ($n = 100$) samples		Large sample ($n = 500, 1000$)	
	$\xi = -0.1$	$\xi \geq 0.1$	$\xi = -0.1$	$\xi \geq 0.1$
0% (none)	PITE 1–2, MLE, PWM, LMOM best	PITE 1–2, MLE, PWM, LMOM best	PITE 1–2, MLE, PWM, LMOM best	PITE 1–2, MLE, PWM, LMOM best
1% (small)	PITE 1–3, MLE, PWM, LMOM best	PITE 1–3, MLE, PWM, LMOM best	PITE 3–5, PWM, LMOM best	PITE 3–5 best
3% (moderate)	PITE 1–4 best; MLE, PWM, LMOM comp	PWM, LMOM best for $n = 50$; PITE 3–5 best for $n = 100$	PITE 6–8, MTM 1–3 best; LMOM, PWM comp. at $n = 500$	PITE 6–8, MTM 1–3 best
5–8% (high)	PITE 1–4, PWM, LMOM best; MLE comp	At $n = 50$ PWM, LMOM best for $\xi \leq 0.3$, PITE 4–5 best for $\xi = 0.5$; at $n = 100$ PITE 5–7 best	PITE 7–8, MTM 2–3 best; PWM, LMOM comp	PITE 7–8, MTM 2–3 best

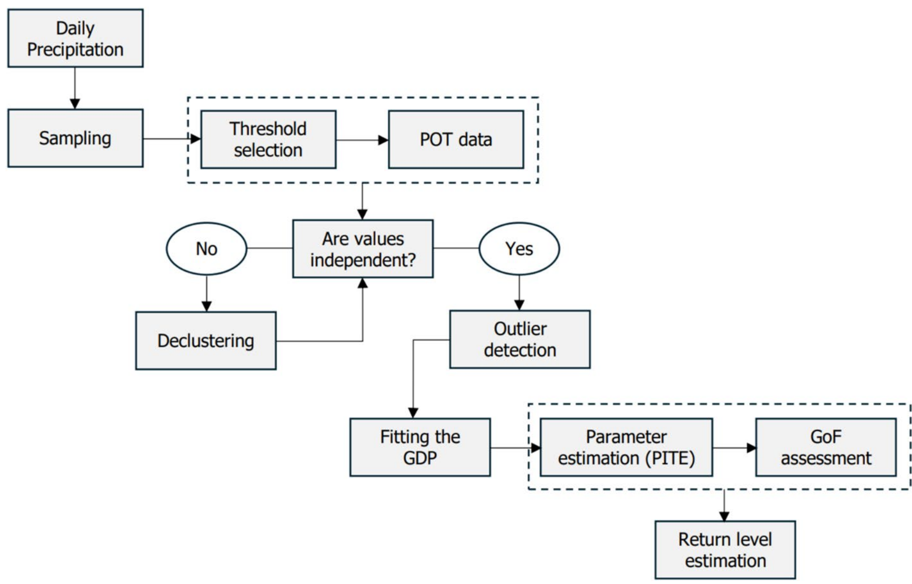


Fig. 5 Flowchart of the procedure for analyzing extreme precipitation events using the POT–GPD framework. The steps include threshold selection and extraction of exceedances, dependence checking and declustering, outlier detection using the generalized boxplot, robust parameter estimation of the GPD with PITE and GoF assessment, followed by estimation of return levels for selected periods

PITE variants in the subsequent precipitation application, where contamination from a small number of very large events (such as typhoon-related rainfall) is a concern.

6 Modeling extreme precipitation in Japan

In this section, we outline the procedure for analyzing extreme precipitation events in the southern region of Japan using the POT method, combined with the robust parameter estimation technique, PITE. This analysis aims to model the frequency and magnitude of extreme daily precipitation events, offering insights into the risk and return levels of heavy rainfall.

6.1 Procedure for the analysis of extreme precipitation

The procedure for the extreme precipitation analysis is outlined below and illustrated in Fig. 5:

1. Sampling: Selecting an appropriate threshold is pivotal in the POT method (Tanaka 2021). The threshold should be set high enough to ensure that only the most extreme precipitation events are considered for further analysis, allowing for effective modeling of the tail of the precipitation distribution. For this purpose, we utilize graphical tools such as the mean residual life (MRL) plot. The MRL plot aids in visually identifying a threshold that balances between capturing sufficient data points and maintaining the extremity required for accurate modeling. Once a suitable threshold is selected, data points exceeding this threshold are classified as POT data.

2. Dependence check and declustering: Extreme precipitation events often exhibit temporal dependence, where

exceedances close in time may result from the same meteorological event, leading to serial correlations. To ensure the independence of exceedances, we apply the Ljung-Box test. If the results indicate temporal dependence, a declustering method is applied. We use the runs declustering approach, where clusters of exceedances are identified, and only one representative value from each cluster is selected (Fawcett and Walshaw 2007). A cluster is deemed to have ended when at least κ consecutive observations fall below the threshold. The maximum (or 'peak') value from each cluster is then extracted for analysis. Once the series of exceedances is confirmed to be independent, we proceed to the next steps in the analysis.

3. Outlier detection: To identify potential outliers in the data, we apply the generalized boxplot method (Bruffaerts et al. 2014). This method is specifically designed for skewed and heavy-tailed distributions, which are common in extreme precipitation data. The generalized boxplot offers significant advantages over the standard boxplot, including robustness against outliers and improved performance in identifying extreme values. By using this approach, we can more effectively detect and manage outliers, ensuring that the subsequent parameter estimation is not unduly influenced by anomalous data points.

4.Fitting the GPD model: This step involves both parameter estimation and the goodness-of-fit (GoF) assessment. The parameters of the GPD, i.e., σ and ξ, are estimated using the PITE, which is known for its robustness, especially when dealing with outliers. The goal is to accurately capture the tail behavior of extreme precipitation events. Following the parameter estimation, the GoF assessment is conducted to ensure that the fitted GPD model is suitable. This involves using diagnostic tools such as the Anderson–Darling (AD) test, relative root mean squared error (RRMSE), and quantile–quantile (QQ) plot to compare the empirical distribution of the exceedances with the theoretical.

5. Return level estimation: Once a satisfactory GPD fit has been obtained, return levels are estimated for a range of return periods T (i.e. 5, 10, 25, 50, 100 and 200 years). We compute the T-year return level using the explicit expression given in Eq. (4). These return levels summarize the expected magnitude of rare precipitation events at each station and provide a convenient basis for comparing the intensity of extremes across locations and time scales.

6.2 Results and discussion

6.2.1 Sampling, dependency testing, and outlier detection

Based on the meteorological context in Japan, a daily precipitation of 100 mm is generally considered a heavy precipitation event and often triggers flood warnings or other public safety measures (Fujibe et al. 2006; Miyasaka et al. 2020). However, the selection of a suitable threshold is a trade-off between bias and variance. A lower threshold could introduce bias by including too many moderate events, which would reduce the precision of the GPD model. On the other hand, a higher threshold might result in fewer exceedances, leading to higher variance in the parameter estimates due to insufficient data. To ensure the suitability of the threshold for the data, we use the MRL plot as a graphical validation tool.

The MRL plot is developed by plotting the average exceedances over a range of thresholds. A suitable threshold is indicated when the MRL plot exhibits stability or a roughly linear trend, suggesting that the selected threshold adequately captures the extreme values necessary for accurate modeling with the GPD. The MRL plot for this analysis is shown in Supplementary Figure S3. Based on this analysis, a threshold of 100 mm was selected for all stations, ensuring consistency in capturing extreme precipitation events. Therefore, data exceeding the threshold $\widehat{u} =$ 100 mm is recognized as POT data.

To ensure the independence of exceedances, the Ljung-Box test was applied to the POT data. Initially, several stations exhibited temporal dependence, as indicated by low p-values, suggesting that consecutive exceedances were likely part of the same meteorological event. This temporal dependence reflects serial correlations between exceedances that occur close in time, violating the independence assumption required for the POT approach. To address this issue, runs declustering was implemented, where exceedances occurring close together were grouped, and only one representative peak value from each cluster was extracted. As outlined by Fawcett and Walshaw (2012), the choice of κ is crucial for balancing independence and preserving sufficient data points. If κ is too small, the peaks will not be far enough apart to safely assume independence, while a larger κ value may result in too few exceedances, reducing the amount of data available for inference. In our study, we used κ value of 2, which was found to provide a suitable balance between independence and data retention. If the Ljung-Box test indicated that the data were already independent without declustering, we did not proceed with further declustering to avoid unnecessary data reduction.

Table 4 shows the results of the Ljung-Box test for dependence, the number of independent exceedances, and outlier detection across different stations in Japan. Most stations exhibited independence in their POT data without the need for declustering, as indicated by p-values greater than 0.05. Some stations, such as Tokushima, Fukuoka, and Ishigakijima, demonstrated particularly high p-values (e.g., p-value$>$0.9), further confirming the independence of exceedances in these locations. In contrast, Kochi was

Table 4 Results of the Ljung-Box test for dependence, number of independent exceedances, and outlier detection across different locations in Japan

Location	Ljung-Box test (p-value)		No. of independent exceedances	No. of outliers	Proportion of outliers (%)
	Without declustering	$\kappa = 2$			
Tokushima	0.9915	–	215	12	5.6
Matsuyama	0.1501	–	61	1	1.6
Kochi	0.007916	0.447	416	8	1.9
Simonoseki	0.6266	–	137	2	1.5
Fukuoka	0.9445	–	147	2	1.4
Oita	0.5709	–	171	1	0.6
Nagasaki	0.8603	–	224	2	0.9
Kumamoto	0.05483	–	220	3	1.4
Miyazaki	0.7379	–	373	3	0.8
Naze	0.6373	–	449	5	1.1
Naha	0.2224	–	276	5	1.8
Ishigakijima	0.9033	–	314	1	0.3

the only station that initially showed temporal dependence, with a p-value of 0.007916. In this case, runs declustering with $\kappa = 2$ was applied, raising the p-value to 0.447, indicating that temporal dependence was removed. Overall, declustering was only necessary for Kochi, ensuring a balance between maintaining data independence and retaining sufficient exceedances for modeling.

Using the generalized boxplot, outliers were identified for the independent POT data, and the results are presented in Supplementary Figure S4. The proportion of outliers varied across stations, ranging from 0.3% in Ishigakijima to 5.6% in Tokushima. Tokushima exhibited the highest proportion of outliers (5.6%), which may indicate localized meteorological phenomena causing rare but extreme precipitation events. While Tokushima showed frequent deviations in precipitation patterns, stations like Naze and Miyazaki, despite having fewer outliers (1.1% and 0.8%, respectively), experienced higher magnitudes of extreme precipitation. This suggests that a higher proportion of outliers does not necessarily correspond to more intense events. In Naze and Miyazaki, fewer but more extreme events, likely driven by larger meteorological phenomena such as typhoons, highlight the need for robust flood risk management.

6.2.2 GPD model fitting and validation

The GPD was fitted to the independent exceedances at each station using the PITE estimator, selected for its robustness, particularly in handling outliers. The scale parameter (σ) and shape parameter (ξ) were estimated for each station. To determine the best fit, three different ARE levels were considered, guided by the number of outliers detected during the earlier analysis. The choice of ARE level helps strike a balance between efficiency and robustness, with higher ARE levels being more efficient but less robust, and lower ARE levels being more robust but potentially less efficient. The corresponding tuning parameter (τ) was approximated through simple interpolation based on the shape parameter (ξ), using values from Table 2. While this interpolation approach does not provide an exact method for determining τ, it offers a reasonable approximation, allowing for flexible adjustments across stations with varying outlier profiles and tail behaviors. To assess the adequacy of the fitted models, the AD test was applied, with p-values greater than 0.05 indicating a satisfactory fit. Once the model passed the GoF assessment, the RRMSE was used to select the best-fitting GPD model for each station. The RRMSE evaluates the overall performance of the model by comparing the fitted values to the observed data, with the lowest RRMSE value indicating the best fit. The detailed results for model comparisons across ARE levels, including intermediate fits and p-values, are provided in Supplementary Table S2. The best-fitting models for each location, based on the lowest RRMSE and satisfactory AD test p-values, are summarized in Table 5.

The results in Table 5 show variation across stations in terms of the σ and ξ parameters, reflecting differences in the extremity and variability of precipitation events at each location. For example, stations such as Kumamoto, Miyazaki, and Naze exhibit relatively higher shape parameter values ($\xi \approx 0.17$, 0.18 and 0.17, respectively), suggesting heavier tails and a higher likelihood of extreme precipitation events. In contrast, stations like Kochi, Simonoseki and Ishigakijima display negative shape parameter values ($\xi \approx -0.06$, -0.09 and -0.05, respectively), indicating bounded tails and a lower probability of very extreme events. The AD test p-values for all stations exceed 0.05, confirming that the GPD models provide a good fit for the data, while the low RRMSE values indicate high accuracy in the model estimations across all locations.

Further support for these results is provided by the QQ plots in Fig. 6, which are based on the best-fitting GPD for

Table 5 Best fitted GPD model for each location

Location	ARE (%)	Tuning parameter (τ)	Scale ($\widehat{\sigma}$)	Shape ($\widehat{\xi}$)	AD (*p*-value)	RRMSE (%)
Tokushima	85	0.20	46.1395	0.0609	0.9876	0.01149
Matsuyama	75	0.34	25.7449	0.1162	0.9282	0.08841
Kochi	60	0.58	51.2777	–0.0562	0.9388	0.00411
Simonoseki	80	0.29	34.6175	–0.0914	0.7693	0.03147
Fukuoka	75	0.47	30.7879	0.1323	0.9697	0.02087
Oita	80	0.4	50.4249	0.0901	0.9664	0.01716
Nagasaki	75	0.47	36.3617	0.1187	0.8892	0.01549
Kumamoto	70	0.55	40.3557	0.1714	0.9953	0.00933
Miyazaki	90	0.31	39.1239	0.1822	0.9564	0.00497
Naze	90	0.31	49.4550	0.1666	0.5777	0.00685
Naha	70	0.47	50.8124	0.0130	0.8991	0.00971
Ishigakijima	75	0.36	55.4265	–0.0495	0.8579	0.00877

each location. The QQ plots compare the theoretical quantiles of the fitted GPD model against the sample quantiles of the precipitation data for each station. For all stations, the data points align closely with the diagonal reference line, indicating that the GPD model fits well for the majority of the data. However, deviations are observed at the right tail of the plots, where potential outliers or anomalies are apparent. These deviations suggest that while the GPD model captures the overall tail behavior well, there are extreme precipitation events in the dataset that are further into the tail than predicted by the model. Such extreme observations are consistent with rare high-impact rainfall events, often associated with intense tropical storms or typhoons affecting southern Japan. While detailed event-based attribution is beyond the scope of this study, these deviations highlight the practical importance of robust estimation methods such as PITE, which are designed to limit the influence of a small number of extreme observations without discarding physically meaningful events. This reinforces the need for robust estimation methods like PITE, which can handle such anomalies effectively without unduly affecting parameter estimates.

In practical applications, such extreme precipitation observations may arise from different physical mechanisms, including rare high-impact tropical cyclones, long-term climatic variability or change, and local factors such as urbanization. Recent studies have shown that changes in precipitation extremes can be associated with evolving storm characteristics and nonstationary climate drivers, and may also motivate alternative modeling strategies such as mixture distributions or nonstationary extreme-value models (e.g., Yan et al. 2017, 2023; Xu et al. 2023). In this study, however, the term "outliers" is used in a statistical sense to describe observations that lie far into the tail and exert strong influence on parameter estimation under standard GPD assumptions. These values are treated as physically meaningful extremes rather than errors, and the objective is not to attribute them to specific physical causes, but to assess the robustness of the proposed estimator when such events are present. Within this context, PITE provides a flexible and practical alternative for handling extreme observations without requiring explicit assumptions about regime changes or multiple precipitation populations.

6.2.3 Return level of extreme precipitation

Using the GPD model fitted to the exceedances, return levels were estimated for 5-, 10-, 25-, 50-, 100-, and 200-year periods at each station. These return levels provide crucial insights into the expected magnitude of extreme precipitation events over different timescales, helping to inform flood risk assessments and infrastructure planning. Figure 7 summarizes the return levels for each station, based on the best-fitting GPD model parameters ($\widehat{\sigma}$ and $\widehat{\xi}$) determined in Sect. 6.2.2.

The results reveal significant regional variability in extreme precipitation patterns across Japan. Stations such as Naze, Miyazaki, and Kumamoto exhibit particularly high return levels for longer periods, with Naze reaching approximately 697 mm for the 200-year period. This pattern is consistent with tropical cyclone rainfall asymmetry. In the Northern Hemisphere, precipitation is often enhanced on the right-hand side of a tropical cyclone track due to the superposition of the storm's tangential winds and translational motion, which increases low level wind speed and moisture supply (Chen et al. 2006). After land interaction, frictional convergence and orographic lifting by nearby terrain can further intensify and sustain rainfall (Bender et al. 1985). Although station elevations are generally low (Table 1), nearby topography such as the mountain range west of Miyazaki and mountainous islands in the surrounding region may contribute to locally enhanced totals during tropical cyclone and frontal events. This indicates a higher potential for severe rainfall events, consistent with these regions' susceptibility to typhoons and large scale

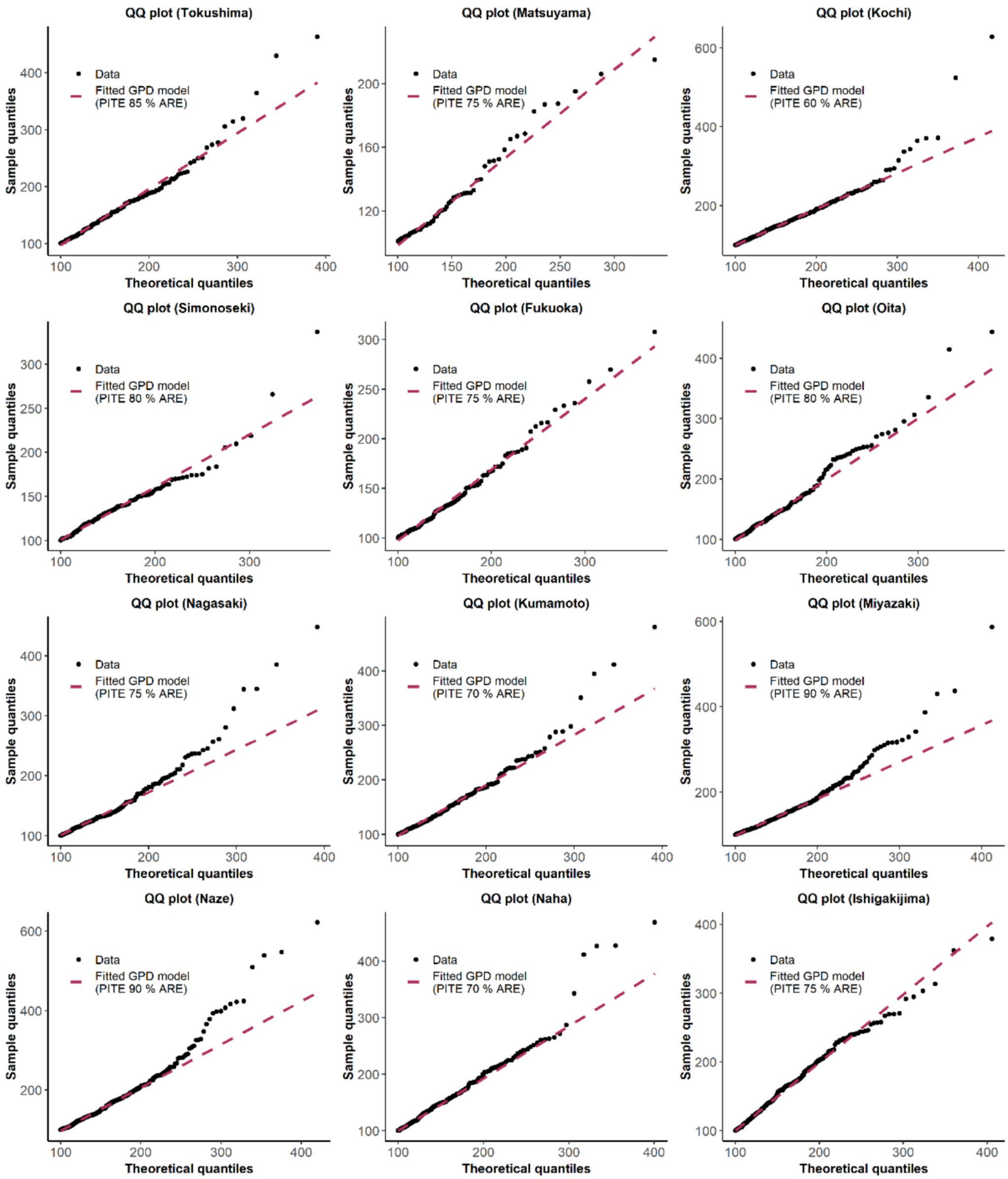


Fig. 6 QQ plots comparing theoretical quantiles of the best-fitting GPD model estimated using PITE with sample quantiles of exceedances at the 12 stations in southern Japan. The close alignment with the diagonal line for most points indicates an overall good fit, while deviations in the upper tail highlight potential outliers and motivate the use of robust estimation

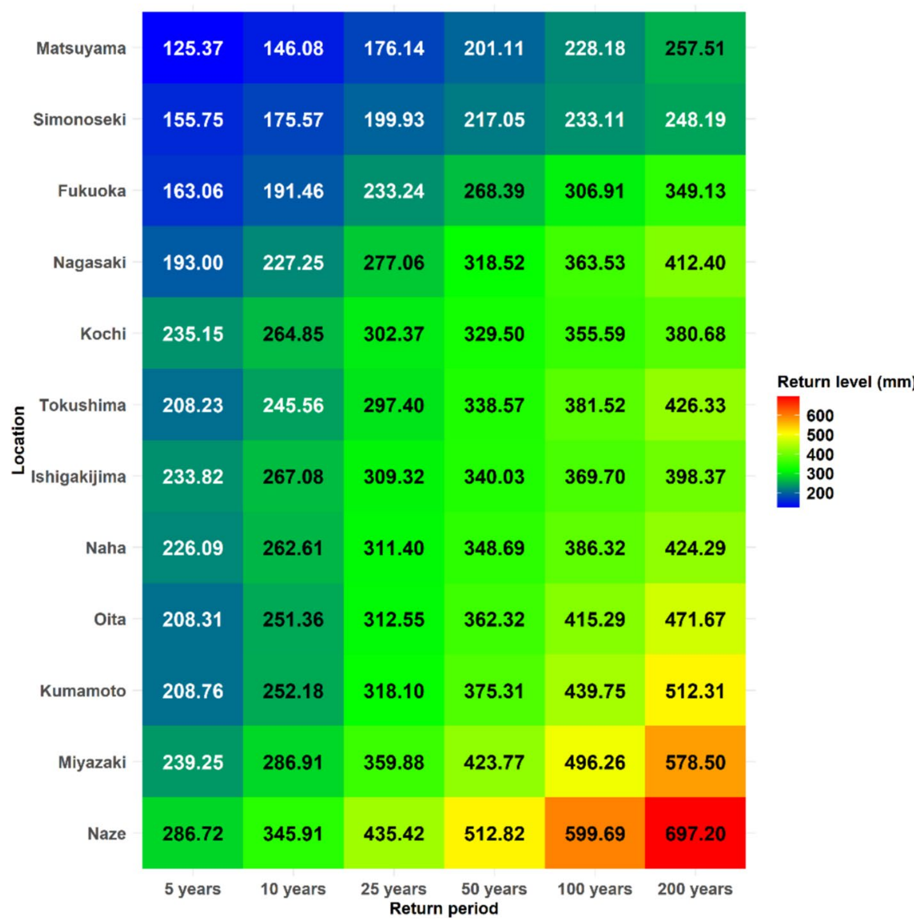


Fig. 7 Heatmap of GPD-based return levels for the POT series at 12 stations in southern Japan. Each cell shows the estimated return level (mm) for 5-, 10-, 25-, 50-, 100-, and 200-year return periods, and the colour scale indicates the magnitude of extreme precipitation, highlighting spatial differences between stations and the increase in intensity with longer return periods

atmospheric systems. Naze's return level is comparable to the historical record of 622 mm per day on October 20, 2010, which occurred when a stationary front was enhanced by Typhoon 13 (Megi), whose minimum central pressure reached 885 hPa (Japan Meteorological Agency 2011). This event severely affected the Amami region including Naze, underscoring the need for robust flood management systems, including improved drainage infrastructure and emergency preparedness plans.

On the other hand, while Matsuyama exhibits lower return levels compared to Naze, it has experienced significant rainfall events in its history. For instance, Matsuyama recorded 215.1 mm of daily precipitation on July 23, 1943 (Japan Meteorological Agency n.d.). This historical event closely aligns with Matsuyama's 100-year return level of 228.18 mm, as estimated in this analysis. Although Matsuyama's lower return levels indicate a reduced probability of extreme precipitation events compared to high-risk southern regions like Naze, they still underscore the potential for significant, albeit less frequent, rainfall events.

Overall, the return level estimates underscore the importance of localized flood risk management. Regions with higher return levels, particularly for longer return periods, should focus on enhancing flood defenses and emergency planning. Meanwhile, areas with lower return levels, though facing less extreme events, should still maintain infrastructure resilience to mitigate potential risks. These findings reinforce the necessity for tailored flood management strategies across Japan, based on the specific risks and meteorological characteristics of each region.

7 Conclusion

This study introduced a robust-efficient method, the PITE, for estimating the parameters of the GPD in extreme value modeling using the POT approach. Designed to enhance robustness in the presence of outliers, PITE proved to be effective and computationally simple, providing reliable estimates even in challenging conditions with high data variability. The properties of PITE, such as its efficiency, score function, and breakdown point, were explored, demonstrating its robustness against extreme values. Through Monte Carlo simulations, we compared PITE with several estimators under various combinations of sample size, tail heaviness, and contamination. The results indicate that high-efficiency PITE variants behave similarly to conventional methods for uncontaminated data, while more robust variants often achieve smaller deficiencies in contaminated scenarios, particularly when a small proportion of outliers is present. The application of PITE, combined with the POT method, to daily precipitation data from 12 meteorological stations in southern Japan, a region highly susceptible to extreme rainfall due to typhoons and complex climatic conditions where storm-driven extremes can act as potential outliers, showcased its practical utility. By estimating GPD parameters and return levels for various return periods, the study revealed significant regional variability in extreme precipitation, with stations such as Naze, Miyazaki, and Kumamoto exhibiting particularly high return levels. However, because the empirical analysis focuses on a relatively small and geographically homogeneous set of 12 southern stations, the findings may not directly generalize to other regions of Japan, especially areas where typhoon influence is weaker. The robust application of PITE and the POT method highlights their relevance in flood risk management and infrastructure resilience, emphasizing the importance of localized strategies to mitigate the impacts of extreme weather events. As a computationally simple and highly robust estimator, PITE offers a valuable tool for practitioners and researchers in fields where modeling extreme events is critical, paving the way for more accurate and reliable predictions in the face of climate change and increasing weather extremes.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00477-026-03248-5.

Acknowledgements The first author gratefully acknowledges the support of the Japan Society for the Promotion of Science (JSPS) as an International Research Fellow under the Postdoctoral Fellowships for Research in Japan (Standard). The authors also extend their sincere thanks to the editor and reviewers for their valuable time and effort in reviewing this manuscript.

Author contributions Conceptualization: Muhammad Aslam Mohd Safari; Methodology: Muhammad Aslam Mohd Safari; Formal analysis: Muhammad Aslam Mohd Safari; Investigation: Tosiyuki Nakaegawa, Nurulkamal Masseran; Writing—original draft preparation: Muhammad Aslam Mohd Safari; Writing—review and editing: Tosiyuki Nakaegawa, Nurulkamal Masseran; Software: Muhammad Aslam Mohd Safari; Supervision: Tosiyuki Nakaegawa; Validation: Tosiyuki Nakaegawa, Nurulkamal Masseran; Funding acquisition: Tosiyuki Nakaegawa; Data curation: Tosiyuki Nakaegawa.

Funding Open access funding provided by The Ministry of Higher Education Malaysia and Universiti Putra Malaysia. This study was supported by the Japan Society for the Promotion of Science KAKENHI (24KF0123) and SENTAN program (Grant Number JPMXD0722680734).

Data availability The research data supporting the results of this manuscript are derived from the Japan Meteorological Agency (JMA). Due to data usage policies and confidentiality agreements with JMA, the data are not publicly shareable. Researchers interested in accessing the data must directly request permission from the JMA.

Declarations

Conflict of interest The authors declare no competing interests.

References

Aggarwal CC (2017) An Introduction to Outlier Analysis BT - Outlier Analysis. In: International S (ed) Aggarwal CC. Publishing, Cham, pp 1–34

Ahmad UN, Shabri A, Zakaria ZA (2011) Trimmed L-moments (1, 0) for the generalized Pareto distribution. Hydrol Sci J 56:1053–1060

Balkema AA, De Haan L (1974) Residual life time at great age. Ann Probab 2:792–804

Bender MA, Tuleya RE, Kurihara Y (1985) A numerical study of the effect of a mountain range on a landfalling tropical cyclone. Mon Weather Rev 113:567–583

Brazauskas V, Kleefeld A (2009) Robust and efficient fitting of the generalized Pareto distribution with actuarial applications in view. Insur Math Econ 45:424–435

Bruffaerts C, Verardi V, Vermandele C (2014) A generalized boxplot for skewed and heavy-tailed distributions. Stat Probab Lett 95:110–117

Chen SS, Knaff JA, Marks FD (2006) Effects of vertical wind shear and storm motion on tropical cyclone rainfall asymmetries deduced from TRMM. Mon Weather Rev 134:3190–3208

Chikamori H, Nagai A (2012) Estimation of return level of daily rainfall using peak over threshold theory. Trans Japanese Soc Irrig Drain Rural Eng 80:167–175

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

Davison AC, Smith RL (1990) Models for exceedances over high thresholds. J R Stat Soc Ser B Methodol 52:393–425

de Zea Bermudez P, Kotz S (2010a) Parameter estimation of the generalized Pareto distribution—Part I. J Stat Plan Inference 140:1353–1373

de Zea Bermudez P, Kotz S (2010b) Parameter estimation of the generalized Pareto distribution—Part II. J Stat Plan Inference 140:1374–1388

Endo H (2023) Long-term precipitation changes in the Baiu and Akisame seasons in Japan over the past 120 years (1901–2020). J Meteorol Soc Japan Ser II 101:309–322

Fawcett L, Walshaw D (2007) Improved estimation for temporally clustered extremes. Environmetrics 18:173–188

Fawcett L, Walshaw D (2012) Estimating return levels from serially dependent extremes. Environmetrics 23:272–283

Fujibe F, Yamazaki N, Kobayashi K (2006) Long-term changes of heavy precipitation and dry weather in Japan (1901–2004). J Meteorol Soc Japan Ser II 84:1033–1046

Fujiwara K, Takemi T, Mori N (2024) Impact of historical and future warming on heavy rainfall over eastern Japan induced by Typhoon Hagibis (2019) in dynamical downscaling simulations. Nat Hazards. https://doi.org/10.1007/s11069-024-06875-9

Google (2024) Map of Japan. https://maps.googleapis.com/maps/api/staticmap?center=30,130&zoom=6&size=640x640&scale=2&maptype=terrain&language=en-EN&key=xxx. Accessed 1 Oct 2024

He X, Fung WK (1999) Method of medians for lifetime data with Weibull models. Stat Med 18:1993–2009

He Y, Peng L, Zhang D, Zhao Z (2022) Risk analysis via generalized Pareto distributions. J Bus Econ Stat 40:852–867

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

Huber PJ (1981) Robust statistics. John Wiley & Sons, New York

Japan Meteorological Agency (2011) Annual Report on the Activities of the RSMC Tokyo – Typhoon Center 2010

Japan Meteorological Agency The top 10 values in the history of observation (values throughout the year). https://www.data.jma.go.jp/stats/etrn/view/rank_a.php?prec_no=73&block_no=1521&year=&month=&day=&view=. Accessed 30 Aug 2024

Jiang F, Zhang J, Zhang M et al (2023) Determination of extreme wind speed under different wind directions and attack angles with mixed wind climates in mountain terrain. Stoch Environ Res Risk Assess 37:4589–4606

Juárez SF, Schucany WR (2004) Robust and efficient estimation for the generalized Pareto distribution. Extremes 7:237–251

Kamahori H, Arakawa O (2018) Tropical cyclone induced precipitation over Japan using observational data. SOLA 14:165–169

Kawase H, Nosaka M, Watanabe SI et al (2023) Identifying robust changes of extreme precipitation in Japan from large ensemble 5-km-grid regional experiments for 4K warming scenario. J Geophys Res Atmos 128:e2023JD038513

Lazoglou G, Anagnostopoulou C, Tolika K, Kolyva-Machera F (2019) A review of statistical methods to analyze extreme precipitation and temperature events in the Mediterranean region. Theor Appl Climatol 136:99–117

Masseran N, Safari MAM (2020) Risk assessment of extreme air pollution based on partial duration series: IDF approach. Stoch Environ Res Risk Assess 34:545–559

Minakawa H, Kudo R, Masumoto T (2018) An evaluation method of the probability of heavy rainfall intensity based on projection uncertainties of climate scenarios. Trans Jpn Soc Irrig Drain Rural Eng 86:I_163-I_173

Miyasaka T, Kawase H, Nakaegawa T et al (2020) Future projections of heavy precipitation in Kanto and associated weather patterns using large ensemble high-resolution simulations. SOLA 16:125–131

Mo C, Ruan Y, He J et al (2019) Frequency analysis of precipitation extremes under climate change. Int J Climatol 39:1373–1387

Nakaegawa T, Murazaki K (2022) Historical trends in climate indices relevant to surface air temperature and precipitation in Japan for recent 120 years. Int J Climatol 42:8950–8970

Nayak S, Takemi T (2020) Typhoon-induced precipitation characterization over northern Japan: a case study for typhoons in 2016. Prog Earth Planet Sci 7:1–12

Oguchi S, Fujibe F (2012) Seasonal and regional features of long-term precipitation changes in Japan. Pap Meteor Geo phys 63:21–30

Park MH, Kim JHT (2016) Estimating extreme tail risk measures with generalized Pareto distribution. Comput Stat Data Anal 98:91–104

Peng L, Welsh AH (2001) Robust estimation of the generalized Pareto distribution. Extremes 4:53–65

Pickands III J (1975) Statistical inference using extreme order statistics. Ann Stat 119–131

Ruckdeschel P, Horbenko N (2013) Optimally robust estimators in generalized Pareto models. Statistics 47:762–791

Sugi M, Imada Y, Nakaegawa T, Kamiguchi K (2017) Estimating probability of extreme rainfall over Japan using extended regional frequency analysis. Hydrol Res Lett 11:19–23

Takaya Y, Yasuda T, Ose T, Nakaegawa T (2010) Predictability of the mean location of typhoon formation in a seasonal prediction experiment with a coupled general circulation model. J Meteorol Soc Japan Ser II 88:799–812

Tanaka APB, Renard P, Straubhaar J (2024) Fracture density reconstruction using direct sampling multiple-point statistics and extreme value theory. Appl Comput Geosci 22:100161

Tanaka S (2021) Comparison of AMS and POT Analysis with Long Historical Precipitation and Future Change Analysis Using “d4PDF” BT - Pioneering Works on Extreme Value Theory: In Honor of Masaaki Sibuya. In: Hoshino N, Mano S, Shimura T (eds). Springer Singapore, Singapore, pp 93–112

Tanaka S, Takara K (2001) Comparison of AMS and PDS in flood frequency analysis. Proc Hydraul Eng 45:205–210

Tsuguti H, Seino N, Kawase H et al (2019) Meteorological overview and mesoscale characteristics of the heavy rain event of July 2018 in Japan. Landslides 16:363–371

Tuel A, El Moçayd N (2023) Evaluating extreme precipitation in gridded datasets with a novel station database in Morocco. Stoch Environ Res Risk Assess 37:3085–3097

Xu W, Xiong L, Yan L et al (2023) Estimating the scale parameter of the norming constants method in analysing nonstationary annual maximum floods. Hydrol Process 37:e14797

Yan L, Xiong L, Liu D et al (2017) Frequency analysis of nonstationary annual maximum flood series using the time-varying two-component mixture distributions. Hydrol Process 31:69–89

Yan L, Zhang L, Xiong L et al (2023) Flood frequency analysis using mixture distributions in light of prior flood type classification in Norway. Remote Sens 15:401