ISSN 1022-8594
JEAC

Jahangirnagar University Journal of Science
Vol. 45, No. 2, pp. 27—48

Application of Extreme Value Analysis in Epidemiology: Modeling Dengue Prevalence Extremes

Farabe Khan Alif *

Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, Serdang, Selangor, Malaysia
Department of Statistics and Data Science, Jahangirnagar University

Abstract

This study explores the application of Extreme Value Theory (EVT) to epidemiology, demonstrating its capability to model rare but impactful disease prevalence events. Using monthly dengue prevalence in Bangladesh (January 2008–September 2023) as a case study, the objectives are to: evaluate the suitability of the Generalized Pareto Distribution (GPD) for capturing upper-tail behavior in epidemiological data, assess an automated threshold selection method based on the Anderson–Darling statistic and L-moments estimation (AD-L), and examine the impact of dataset extension on model performance and uncertainty. Non-stationarity is addressed by fitting a Seasonal ARIMA (SARIMA) model, with residuals used for GPD modeling. Return levels for 5-, 10-, and 20-year periods are estimated and expressed on the original prevalence scale, with 95% confidence intervals obtained via bootstrap resampling. A hybrid simulation framework generates synthetic series (N = 500) replicating observed trend, seasonality, and tail features to assess the effect of increased sample size. Results show that the GPD, combined with the AD-L method, effectively models heavy-tailed residual extremes, while simulated data improve goodness-of-fit metrics substantially. However, the scarcity of exceedances limits reductions in return level uncertainty. This work demonstrates a novel, data-driven application of EVT to infectious disease prevalence, offering a transferable framework for epidemiological risk assessment.

Keywords: Extreme value theory, Dengue, Threshold selection, SARIMA, Generalized Pareto distribution.

1. Introduction

Extreme Value Theory (EVT) offers a principled framework for modeling rare but consequential events phenomena that lie in the tails of a distribution and has found extensive use across disciplines such as hydrology, environmental sciences, insurance, and finance [1]. Despite its broad applicability in many scientific fields, EVT has been conspicuously underutilized in public health and epidemiology. Notable exceptions include its use in modeling extreme pneumonia or influenza mortality in France and extreme emergency department visit surges in Paris, which have demonstrated EVT's utility in directly informing health system preparedness and planning [2]. Nevertheless, systematic applications of EVT to infectious disease prevalence, especially mosquito-borne diseases like dengue, remain rare.

*Author for correspondance e-mail: gs67618@student.upm.edu.my

Given that dengue fever exhibits highly variable seasonal peaks and occasional extreme surges in case numbers, this variability poses significant challenges for prediction and control strategies in affected regions [3]. Applying EVT to epidemiological data presents a promising yet underexplored opportunity to enhance public health risk assessment and preparedness.

Dengue fever is an acute mosquito-borne viral infection that poses a growing threat to global public health, particularly in tropical and subtropical regions [4]. Transmitted primarily by Aedes aegypti and Aedes albopictus mosquitoes, it has become the most widespread arboviral disease worldwide [5]. The global burden of dengue has escalated sharply over the past two decades. According to [4], the global burden of disease study reported a 400% increase in dengue cases between 2000 and 2013. This alarming trend has persisted, with over 14.1 million dengue cases reported globally in 2024, double the number recorded in 2023 (7 million) and 12 times higher than in 2014 (1.2 million) [5]. The 2024 outbreak also resulted in 9,508 dengue-related deaths, corresponding to a global case-fatality rate of 0.07% [5]. These statistics reflect dengue's transformation from a regional health issue into a global challenge, with transmission now documented in non-endemic regions such as mainland Europe and the United States [6]. Several factors drive this surge. Globalization has facilitated the cross-border spread of dengue, with international air travel increasing dramatically from 31 million passengers in 1950 to nearly 4.5 billion annually in the post COVID-19 era [7]. Rapid urbanization, particularly in low- and middle-income countries, has created ideal mosquito breeding sites in densely populated and poorly managed urban environments [8]. Climate change has further amplified transmission dynamics by extending vector seasons, accelerating mosquito development, and increasing the frequency of human-mosquito contact [9]. The spread of Aedes albopictus to nearly every continent, combined with the increased severity of secondary infections and increased vulnerability among individuals with chronic conditions such as diabetes, obesity, and hypertension, further amplifies the risk of severe outcomes [5]. Collectively, these factors have intensified the global threat of dengue fever, emphasizing the urgent need for robust surveillance systems and innovative statistical approaches to quantify and manage extreme patterns in its epidemiology.

The dataset used in this study comprises the monthly prevalence of dengue cases in Bangladesh from January 2008 to September 2023, obtained from [10]. Bangladesh, with a very high density of population within just 143,000 km^2, is highly vulnerable to dengue due to its tropical monsoon climate, characterized by high humidity and average monsoon temperatures around 29 degrees Celsius, conditions ideal for mosquito breeding and viral transmission [11]–[13]. Historically, the country experienced its first severe dengue outbreak in 2000, predominantly in Dhaka, Chittagong, and Khulna, resulting in 5,551 reported cases and 93 deaths [14]. Subsequent studies identified the introduction of a virus strain from Thailand and highlighted strong correlations between dengue outbreaks and monsoon rainfall, with cases peaking during September and di- minishing in the dry winter season [15], [16]. In recent years, the burden has escalated dramatically, with Dhaka division alone accounting for 70% of cases and 61% of fatalities in 2022, while Chattogram reported 14% of cases and nearly 24% of deaths [17]. These statistics, coupled with the increasing frequency and intensity of outbreaks, underscore a growing public health challenge. The historical patterns of seasonality and upward trend documented in previous outbreaks are also reflected in our dataset, making it particularly suitable for examining the extremes of dengue prevalence through advanced statistical modeling.

The epidemiological dynamics of dengue, marked by pronounced seasonal peaks and occasional extreme surges, demand statistical methods capable of capturing rare but impactful

events. EVT provides a rigorous framework for modeling such extremes, with the generalized Pareto distribution (GPD) being a cornerstone in its peaks-over- threshold (POT) approach. Unlike the block maxima method, which relies on the generalized extreme value (GEV) distribution and discards much of the data, the POT–GPD framework efficiently uses all exceedances above a selected threshold [1], [18]. This is particularly advantageous for shorter time series, such as monthly dengue prevalence data, where retaining exceedance information is critical. While alternative approaches such as self-exciting threshold autoregressive (SETAR) models or zero-inflated models are effective for capturing regime changes or excess zeros in time-series data, they are not designed to quantify the probability and magnitude of extreme exceedances [19], [20]. The present study focuses on rare, high-impact dengue outbreaks, where reliable estimation of return levels and exceedance probabilities is essential for epidemiological risk assessment. EVA, through the POT framework and the GPD, provides a theoretically grounded methodology for modeling the statistical behavior of the tail and is thus more suitable for this application [1], [21]. The GPD's parameters, scale and shape flexibly describe the spread and tail heaviness of exceedances, enabling ac- curate estimation of extreme risks. However, its performance strongly depends on an appropriate threshold: a low threshold biases estimates downward, while a high threshold inflates variance and destabilizes parameters [22], [23].

The challenge of threshold selection has driven significant research efforts. Traditional methods, such as the mean residual life (MRL) plot, though widely used, are subjective and prone to inconsistent results [24], [25]. To overcome these limitations, various automated approaches have emerged, including weighted exceedance techniques and goodness-of-fit (GOF)-based procedures [24], [26]. Recent advancements further enhance objectivity by integrating asymptotic theory, bootstrap-based diagnostics, and stability criteria [27]–[29]. In this study, we apply an automated threshold selection technique proposed by Alif, Ali, and Safari [30], which combines the Anderson–Darling (AD) GOF test with L-moments parameter estimation. Moving forward, we will refer to this technique as the AD-L method. L-moments are particularly robust in handling skewed, heavy-tailed data and improve estimation stability under sample variability [31], [32]. This combination offers a statistically rigorous yet computationally efficient approach, ideally suited to modeling dengue prevalence extremes in the presence of limited data, strong seasonality, and rare but severe outbreaks.

Building on these methodological advancements, the present study applies EVT, with a focus on the GPD and an automated threshold selection framework, to model extremes in monthly dengue prevalence in Bangladesh. The study aims to evaluate the suitability of GPD in capturing the tail behavior of dengue prevalence residuals after adjusting for seasonality and trend, while also examining the performance of the automated threshold selection method under both real and simulated epidemiological conditions. Additionally, it investigates how increasing sample size affects the behavior of GOF tests, particularly whether larger datasets lead to higher p-values, indicating improved model compatibility. By simulating extended datasets that replicate real-world characteristics of dengue prevalence, including seasonal patterns, long-term trends, and rare extreme spikes, this research provides deeper insight into the statistical properties of extreme events in epidemiology. The findings are expected to enhance public health risk assessment by producing more robust return level estimates and offering a framework that can better inform preparedness and targeted intervention strategies in regions vulnerable to vector-borne diseases.

2. Method and Methodology

2.1 Theoretical Background

The GPD is a central model in EVT for characterizing exceedances over a high threshold u [1]. Within the POT framework, observations exceeding u are modeled directly, enabling precise inference on tail behavior even when the bulk of the data is non-extreme [18]. A critical aspect of this approach is the choice of u; thresholds set too low in- corporate non-extreme values, biasing estimates downward, while very high thresholds reduce the number of exceedances, inflating variance and decreasing parameter stability [22], [23]. Originally introduced by Pickands III [33], the GPD is defined by a scale parameter $\sigma > 0$ and a shape parameter ξ, where ξ governs tail heaviness. Its cumulative distribution function (CDF) is expressed as:

$$G(x;\sigma,\xi)=\begin{cases}1-\left[1+\xi\dfrac{x-u}{\sigma}\right]^{-1/\xi}, & for\ \ \xi\neq 0,\\ 1-\exp\left(-\dfrac{x-u}{\sigma}\right), & for\ \ \xi=0,\end{cases} \tag{1}$$

where $x > u$ for $\xi \geq 0$, and $u < x < u - \sigma/\xi$ when $\xi < 0$. Positive values of ξ indicate heavy-tailed behavior, $\xi = 0$ corresponds to an exponential tail, and $\xi < 0$ implies an upper bound on the extremes [34]. The GPD framework relies on several assumptions: exceedances above a sufficiently high threshold follow an approximate GPD distribution, the data and exceedances are independent and identically distributed (iid), and threshold choice ensures parameter stability [1], [33]. To validate iid stationarity, the dengue prevalence series was tested using the Augmented Dickey–Fuller (ADF) test and Seasonal-Trend decomposition using Loess (STL), after which non-stationary components were removed using SARIMA. The AD-L method provided a statistically principled threshold, while GOF tests confirmed adequacy of the fitted GPD under this threshold (see Section 2.5).

To estimate the parameters σ and ξ, this study employs the L-moments approach. L- moments, discussed by Hosking [35], are linear combinations of order statistics that provide stable estimates of distributional characteristics, particularly under skewed or heavy-tailed conditions. They offer resilience to outliers and sampling variability, making them especially advantageous for extreme value modeling in small or noisy datasets [32]. The r^{th} L-moment can be expressed in terms of the quantile function $Q_Y(p)$ as:

$$L_r=\int_0^1 Q_Y(p)P^*_{r-1}(p)\,dp \tag{2}$$

Where P^*_{r-1} denotes the shifted Legendre polynomial of order $r-1$ [36]. These properties make L-moments particularly well-suited for the GPD estimation in this study, where they also support efficient computation during bootstrap resampling [24]. The theoretical foundation of L-moments is described in Hosking [35], while the detailed derivation of GPD parameter estimation through this approach is discussed in detail by Abdul-Moniem and Selim [37]. As such, we rely on these established formulations rather than rederiving them here.

The threshold selection procedure applied here follows the AD–L method, which integrates the AD GOF test with L-moments-based parameter estimation. The AD test places greater

emphasis on tail discrepancies compared to other GOF statistics, making it particularly appropriate for extreme value applications [38]. For a sample of ordered exceedances $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$, the AD statistic is given by:

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}\left[(2i-1)\log G(x_{(i)}; u, \hat{\sigma}, \hat{\xi}) + \log\left(1 - G(x_{(n+1-i)}; u, \hat{\sigma}, \hat{\xi})\right)\right], \quad (3)$$

where $G(x; \hat{\sigma}, \hat{\xi})$ is the fitted GPD CDF. Models yielding higher p-values from this test indicate stronger compatibility between the theoretical GPD and empirical exceedances [39], [40]. To provide a comprehensive assessment of model fit, the Kolmogorov–Smirnov (KS) and Cramer–von Mises (CVM) tests are also utilized. The KS test measures the maximum distance between the empirical and theoretical CDFs (For detail refer to [41]). The CVM test evaluates the integrated squared difference between the CDFs (details has been discussed by [42]–[44]). Although the AD test is more sensitive to tail deviations, KS and CVM offer valuable complementary diagnostics. For all tests considered, higher p-values indicate better empirical agreement with the fitted GPD, reinforcing the validity of the threshold selection and tail modeling approach.

2.2 The AD-L Threshold Selection Method

The AD-L method introduces an automated procedure to determine an optimal thresh- old for GPD modeling. Unlike traditional graphical techniques, this approach uses a structured evaluation of candidate thresholds to minimize subjectivity and maximize model fit. To understand the AD-L method, it is necessary to clarify the notion of a “significant amount of zeros” as established in its development. The AD-L method was originally designed with the assumption that its applications will be in environmental and climate datasets (among many others), particularly precipitation data, which often contain more than 50% zeros [30], [45]. For the present study, however, we adopted a more conservative assumption: if a dataset contains less than 20% zeros, this proportion is not considered significant, and the standard algorithm proceeds directly from the median-based candidate threshold. This adjustment reflects the limited sample size of epidemiological datasets, where even moderate proportions of zeros could bias the candidate threshold toward the lower tail, potentially affecting the performance of the AD-L method. In contrast, when zeros exceed 20% of the data, an additional adjustment step is required to account for the large proportion of zero values. Since the dengue prevalence dataset analyzed here contained no structural zeros, the simplified procedure was applied. Nonetheless, the proposed modification warrants further evaluation in future studies. The process unfolds as follows:

1. **Data segmentation:**
   - Arrange observations in ascending order.
   - Split the ordered data into 200 segments to capture variations across its range. The choice of 200 segments was based on empirical findings from the simulation study by [30], which demonstrated that this segmentation provides an appropriate balance of granularity and stability across a range of simulated and real datasets.
2. **Identification of a Starting Threshold** u_1
   - When zeros dominate the dataset, calculate the mean for each segment, select those exceeding the first quartile Q1, and take their average as u1.
   - In the absence of a significant quantity of zeros, simply use the median as the initial candidate.

3. **Definition of an Upper Bound u_n:**
   - The endpoint of the candidate thresholds, denoted as un, is generally set around the 5th largest observation in the dataset. At the same time, the number of exceedances above the threshold must be considered, particularly for larger datasets. For datasets with at least 5000 observations, a minimum of 80 exceedances is recommended, whereas for datasets containing 1000 to 5000 observations, at least 20 exceedances should be retained. When fewer than 1000 observations are available, no strict requirement is imposed, allowing greater flexibility in small-sample settings. These criteria help balance sample size with statistical reliability, reflected in comparatively lower uncertainties and preventing excessive bias from too few exceedances, with consistent results demonstrated in the simulation studies by [30].
4. **Construction of Threshold Candidates:**
   - Generate roughly 300 evenly spaced thresholds between u_1 and u_n.
5. **Parameter Estimation via L-Moments:**
   - For each candidate u_k, isolate the exceedances and estimate the GPD parameters (σ, ξ) using the L-moments framework.
6. **Evaluation with the Anderson-Darling Test:**
   - Fit a GPD for each uk and calculate the AD test p-value to assess model adequacy.
7. **Selection of the Best Threshold u_0**
   - Choose the threshold that yields the highest AD p-value:

$$u_0 = \arg\max_{u_k} p_{\mathrm{AD}}(u_k)$$

This systematic approach ensures consistency across datasets and effectively identifies the threshold where the GPD best captures the tail behavior of dengue prevalence extremes.

2.3 Return Level Estimation

Estimating return levels is a crucial component of extreme value analysis, as it provides a direct interpretation of the magnitude of rare events expected over specified periods. After identifying an appropriate threshold u, the return level quantifies the value likely to be exceeded on average once every r observation or r years, depending on the con- text of the dataset. As noted by [1], the accuracy of return level estimates is highly sensitive to the chosen threshold, since both overly low and excessively high thresholds can introduce significant bias or variance.

For the GPD, the r-year return level, denoted by Z_r, is calculated as:

$$Z_r = \begin{cases} u + \dfrac{\hat{\sigma}}{\hat{\xi}}\left[(rn_y\hat{\beta}_u)^{\hat{\xi}} - 1\right], & \text{if } \hat{\xi} \neq 0, \\ u + \hat{\sigma}\log(rn_y\hat{\beta}_u), & \text{if } \hat{\xi} = 0, \end{cases} \tag{4}$$

where $\hat{\sigma}$ and $\hat{\xi}$ are the estimated GPD scale and shape parameters, respectively. The term $\hat{\beta}_u$ represents the empirical exceedance probability, calculated as the proportion of data points exceeding the threshold u, and n_y is the number of observations per year. This formulation

allows return levels to be expressed on an annual scale, which is particularly meaningful for epidemiological applications, where stakeholders often assess risks over multi-year horizons. Because return level estimates are directly influenced by both parameter estimation and threshold uncertainty, it is essential to quantify their reliability. In this study, we adopt a bootstrap-percentile approach to construct confidence intervals, enabling us to capture the variability inherent in the estimation process. This method provides a robust framework to evaluate precisely extreme dengue prevalence levels can be predicted, an important factor in public health planning and preparedness.

1.4 Bootstrap Percentile Method

To assess the uncertainty surrounding our return level estimates, we adopted the bootstrap percentile method, a widely recognized non-parametric resampling technique valued for its simplicity and robustness [46]. Unlike approaches that rely heavily on theoretical distributional assumptions, this method builds an empirical sampling distribution of the estimator by repeatedly resampling from the observed data. In this study, the bootstrap is used to generate 95% confidence intervals (CI) for the estimated return levels. Following the general framework described by Mooney, Duval, and Duvall [47] and Tibshirani and Efron [48], we proceed as follows:

1. Draw B = 1000 bootstrap samples from the original dataset using sampling with replacement.
2. For each sample, reapply the entire modeling process, including threshold selection and estimation of return levels using the same methodology as applied to the original data.
3. Compile the resulting bootstrap estimates and sort them to obtain the empirical distribution of each return level.
4. Define the 95% CI by extracting the 2.5th and 97.5th percentiles of the bootstrap distribution.

By capturing both sampling variability and model sensitivity, this approach yields a transparent measure of uncertainty in return level estimates, strengthening the reliability of our conclusions.

2.5 Analysis Method

The methodology of this research is structured to systematically analyze extreme dengue prevalence values using EVT while accounting for the inherent seasonal and trend components often present in epidemiological data. Stationarity assessment is crucial because many time series methods, including EVT under the POT framework, assume that residuals or transformed data exhibit iid behavior. The process begins with an examination of the raw dataset to understand its statistical properties and temporal patterns, followed by rigorous testing for stationarity:

1. The analytical framework begins by assessing the stationarity of the time series, a crucial property that determines the subsequent modeling approach. To preliminarily assess this property, visual inspection through time series plots is performed, which can reveal potential trends, seasonality, or structural breaks. However, formal statistical testing is required for confirmation. We employ the ADF test and the Kwiatkowski Phillips Schmidt Shin (KPSS) test to verify stationarity. The ADF test, implemented using the *urca* package in R [49], evaluates the null hypothesis that a unit root is present in the series, indicating non-stationarity. The test is based on estimating the following regression:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p} \delta_i \Delta y_{t-i} + \epsilon_t, \quad (5)$$

where yt is the time series at time t, $\Delta y_t = y_t - y_{t-1}$ denotes the first difference, t is a time trend, and p is the lag order chosen to address serial correlation. The null hypothesis $H_0: \gamma = 0$ implies the presence of a unit root (non-stationarity), while rejection of H0 suggests stationarity (details of ADF test are discussed by Mushtaq [50]). In contrast, the KPSS test takes the opposite stance by setting stationarity as the null hypothesis H0. For this we use the package tseries [51]. The KPSS statistics are computed from the residuals of a regression of y_t on a deterministic component (constant or trend), defined

$$KPSS = \frac{1}{T^2} \sum_{t=1}^{T} S_t^2 / \hat{\sigma}^2, \quad (6)$$

as:

where T is the number of observations, St is the cumulative residual sum, and $\hat{\sigma}^2$ is the long-run variance of the residuals. A large KPSS statistic indicates evidence against stationarity (details are discussed by Syczewska [52]).

The complementary use of the ADF and KPSS tests provides a more reliable inference: if the ADF fails to reject H_0 while KPSS rejects its null, strong evidence of non-stationarity is concluded. If the series is confirmed to be stationary, the modeling proceeds directly with the GPD. Otherwise, the non-stationary structure is further examined to identify deterministic components such as trend or seasonality for decomposition and subsequent modeling.

2. If the dataset is identified as non-stationary in Step 1, the next stage involves decomposing the series to detect the underlying deterministic components. STL is applied to separate the series into its trend, seasonal, and residual components [53]. This technique provides a flexible, nonparametric approach to capture both smooth long-term variations and recurring seasonal patterns. STL is implemented through the forecast package in R [54]. The decomposition expresses the observed time series y_t as:

$$y_t = T_t + S_t + R_t, \quad (7)$$

where T_t is the smooth trend, S_t is the seasonal component, and Rt is the remainder capturing short-term irregularities. Visual examination is carried out by plotting each component separately. A gradually increasing or decreasing T_t indicates the presence of a long-term trend, while repetitive oscillations in S_t confirm seasonality. This graphical assessment ensures that the structural patterns in the dataset are appropriately identified before further modeling.

3. After identifying the presence of non-stationary components through STL decomposition, the next step involves removing trend and seasonality to obtain a stationary residual series suitable for extreme value analysis. This is achieved by fitting a seasonal autoregressive integrated moving average (SARIMA) model to the original series. The SARIMA model is expressed as:

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^D X_t = \Theta_Q(B^s)\theta_q(B)\epsilon_t, \quad (8)$$

where B is the backshift operator, s denotes the seasonal period (for monthly data, $s = 12$), and ϵ_t represents white noise. The polynomials $\phi_p(B)$ and $\theta_q(B)$ correspond to the non-seasonal autoregressive and moving average terms, while $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ represent the seasonal counterparts. The differencing parameters d and D remove trend and seasonal integration, respectively (for details refer to [55]). Once the model is fitted, the residuals R_t are extracted as:

$$R_t = X_t - \widehat{T}_t - \widehat{S}_t, \tag{9}$$

where $\hat{T}_t$ and $\hat{S}_t$ denote the estimated trend and seasonal components, respectively. These residuals, free from deterministic patterns, form the basis for subsequent extreme value modeling using the GPD.

4. After obtaining the residual series R_t from SARIMA modeling, we directly apply the AD-L threshold selection method to determine the optimal threshold u for GPD modeling. Since the residuals R_t are approximately iid, they satisfy the fundamental assumptions of EVT, making them suitable for tail analysis. Once the threshold is identified, exceedances above u are extracted, and the GPD parameters, the scale σ and shape ξ, are estimated using the L-moments approach, ensuring stable and efficient parameter estimation.
5. Using the fitted GPD model with parameters estimated from the exceedances, we calculate return level estimates for 5-, 10-, and 20-year return periods. To capture the uncertainty in these estimates, the bootstrap percentile method is applied.
6. In the final step, the calculated return levels and their corresponding 95% CIs, initially obtained on the residual scale, are transformed back to the original prevalence scale. This reconstruction is achieved using

$$X_r = \widehat{X} + Z_r, \tag{10}$$

where $\widehat{X}$ represents the mean fitted value from the SARIMA model for the month of interest, and Z_r denotes the r-year return level on the residual scale. The same transformation is applied to the lower and upper confidence bounds, ensuring that the results are interpretable within the context of the observed prevalence.

2.6 Simulation Study

The simulation study is designed to investigate whether increasing the sample size, while preserving key structural properties of the original data, improves the performance of GPD modeling in terms of reducing uncertainty and enhancing model compatibility. Unlike purely synthetic models, this simulation does not aim to replicate the original dataset exactly, as the complex dynamics of dengue prevalence, influenced by climatic, ecological, and socio-environmental factors, are very hard to reproduce accurately. Instead, it seeks to generate an extended dataset that reflects similar statistical behavior, including trend, seasonality, and occasional extreme deviations. The procedure consists of the following stages:

1. The original dataset is first modeled using a Seasonal ARIMA (SARIMA) framework, which effectively captures both long-term trend and seasonal components. The fitted model provides baseline values representing expected prevalence in the absence of unexpected shocks. Residuals extracted from this model serve as the foundation for simulating new data.

2. Residuals R_t are divided into two segments:
   - Phase 1 (earlier period with fewer extremes).
   - Phase 2 (later period exhibiting more frequent extremes).

   The real dataset also follows similar phases (see section 3.1). Residual simulation is carried out using a hybrid approach:
   - For Phase 1, new residuals are generated through bootstrap resampling.
   - For Phase 2, extreme spikes are injected with an empirically estimated probability p_e:

$$P\left(R_t > r_e\right) = p_e, \tag{11}$$

   where r_e represents an extreme threshold.

   Extreme magnitudes are scaled by parameters derived from observed maxima and minima, ensuring heavy-tailed behavior.
3. The simulated residual series R_t^* is constructed by combining resampled Phase 1

$$R_t^* = \begin{cases} R_t^{(b)}, & t \in \text{Phase 1}, \\ f(R_t^{(b)}, r_e, p_e), & t \in \text{Phase 2}, \end{cases} \tag{12}$$

   residuals with hybrid-generated Phase 2 residuals:

   where $R_t^{(b)}$ denotes bootstrap samples and f represents the transformation that introduces rare extremes.
4. Simulated prevalence X_t^* is obtained by adding simulated residuals to the SARIMA baseline $\widehat{X}_t$:

$$X_t^* = \widehat{X}_t + R_t^*, \tag{13}$$

   with negative values truncated to a small positive constant to ensure epidemiological plausibility.
5. The SARIMA model is used to forecast the baseline for $N = 500$ periods, beyond the original series length. Residuals are simulated as described above, generating an extended dataset $X_{t=1}^{*\ 500}$ that mirrors the original dataset's statistical structure without replicating exact values.
6. The extended simulated dataset undergoes the same analytical pipeline applied to the original data:
   - Fitting SARIMA to remove the trend and seasonality.
   - Extracting residuals for subsequent analysis.
   - Applying the AD-L method to determine the optimal threshold.
   - Estimating GPD parameters using L-moments.
   - Computing return levels for 5-, 10-, and 20-year periods.
7. Bootstrap resampling $(B = 1000)$ is applied to residuals to derive the empirical distribution of return levels and calculate 95% CIs. GOF of the GPD is evaluated using the AD, KS, and CVM tests for comparison.
8. Return levels and their confidence intervals are translated back to the prevalence scale using the average SARIMA baseline. This ensures results are interpretable in the context of public health.

This simulation methodology provides a controlled framework to assess how enlarging the dataset under similar trend-seasonality conditions affects the reliability of GPD estimates. While it cannot replicate the real-world dataset exactly, it offers meaningful insight into the potential benefits of increased data availability in extreme value modeling for epidemiology.

3. Results and Discussion

3.1 Bangladesh Dengue Prevalence Dataset

The dataset consists of 189 monthly observations of dengue prevalence in Bangladesh spanning January 2008 to September 2023. The scatter plot in Figure 1 reveals strong non-stationarity, characterized by an upward trend and recurring seasonal peaks.

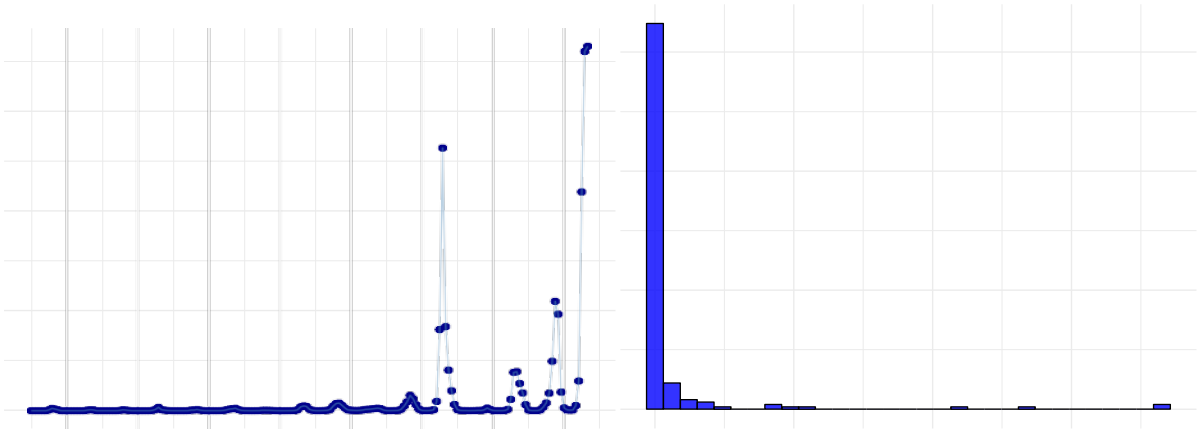

Figure 1: Scatter plot (left) and histogram (right) of monthly dengue prevalence in Bangladesh (January 2008–September 2023), illustrating pronounced seasonality, an upward trend, and a heavy-tailed distribution with extreme values.

Years such as 2019 and 2022 show exceptionally high prevalence levels, aligning with reports of severe dengue activity during these periods. These fluctuations suggest the interplay of climatic variability, vector dynamics, and urbanization in shaping transmission patterns. The histogram Figure 1 shows a right-skewed distribution with a heavy tail, where the majority of months record low prevalence while a small number exhibit extreme values. This distribution underscores the importance of modeling approaches capable of focusing on the tail, as these extremes drive much of the public health impact. To achieve this, deterministic components such as trend and seasonality must first be removed to allow for an accurate application of GPD-based extreme value analysis.

The results of the stationarity tests further confirm the presence of non-stationarity in the dengue prevalence series. The ADF test produced a test statistic of −2.858 with a p-value of 0.217 (> 0.05), leading to a failure to reject the null hypothesis of a unit root. This outcome indicates that the series is non-stationary. In contrast, the KPSS test, which tests the null hypothesis of stationarity, yielded a p-value less than 0.05 (test statistic = 0.161), thereby rejecting the null and reinforcing evidence of non-stationarity. These complementary results clearly establish that the original dengue prevalence data exhibits non-stationary behavior, driven by underlying trends and seasonal patterns, and thus requires further decomposition before applying extreme value analysis.

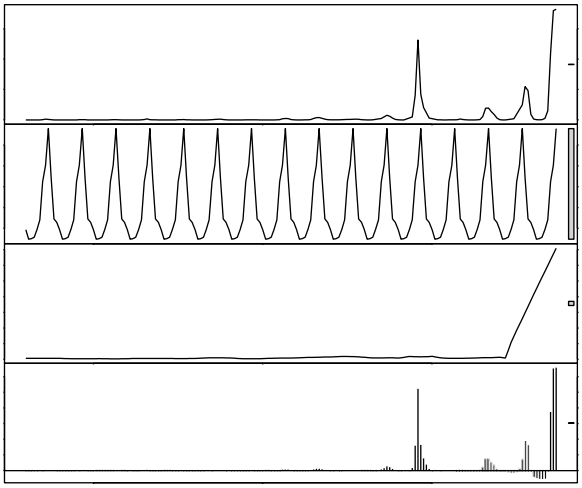

Figure 2: Seasonal-Trend decomposition of the dengue prevalence series using STL, illustrating the observed data (top panel), extracted seasonal component (second panel), estimated trend (third panel), and residual component (bottom panel).

The STL decomposition in Figure 2 highlights the seasonal and trend patterns embedded in the dengue prevalence data. The seasonal component demonstrates continuous cyclical behavior across the years, without any distinct mid-year peaks, indicating a persistent seasonal influence on dengue prevalence. The trend component, on the other hand, remains relatively stable in earlier years but exhibits a sharp upward trajectory after 2020, aligning with the reported surge in dengue cases during recent outbreaks. The residual component primarily captures short-term irregular fluctuations, including sporadic extreme deviations not explained by the deterministic components. Considering these observations, a SARIMA model is applied to isolate and remove the trend and seasonal effects, producing residuals that better approximate stationarity. The selected model, SARIMA (2,1,1) with a seasonal period $s = 12$, incorporates two autoregressive terms ($AR_1 = 1.079$, $AR_2 = -0.386$) and one moving average term ($MA_1 = -0.904$).

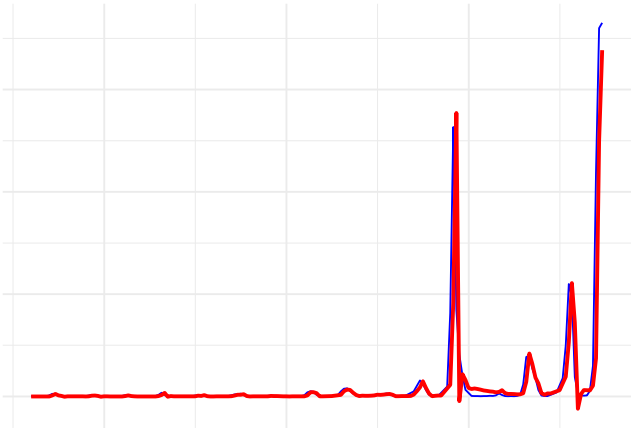

Figure 3: Observed dengue prevalence (blue) and SARIMA (2,1,1) fitted baseline (red), showing close alignment between the model and the observed data.

Figure 3 presents the observed dengue prevalence (blue) alongside the SARIMA fitted baseline (red). The fitted baseline tracks the observed data closely throughout the study period, indicating that the model successfully captures the deterministic trend and seasonal fluctuations. This strong fit validates the SARIMA model as a reliable basis for extracting residuals that will be subjected to extreme value analysis in the subsequent steps.

Figure 4: Time series plot of SARIMA residuals, showing random fluctuations around zero with no visible autocorrelation or deterministic patterns, confirming successful removal of trend and seasonality.

The residual series from the fitted SARIMA (2,1,1) model is evaluated to confirm its suitability for extreme value analysis. The Ljung-Box test (for details see [56]) implemented via the forecast package, tests for autocorrelation. The test yields $Q^{*} = 6.34$ with df = 21 and a p-value of 0.999, indicating no autocorrelation and confirming residual independence. Stationarity is assessed using ADF and KPSS tests. The ADF result (test statistic = −3.45, p = 0.048) rejects the null of a unit root, confirming stationarity, while the KPSS statistic 0.111 is below the 5% critical value 0.146, supporting the null of stationarity. Figure 4 shows the residual series R_t, fluctuating randomly around zero with no visible patterns, confirming that the SARIMA model successfully removed trend and seasonality. These residuals, now satisfying iid assumptions, are used for GPD modeling with the AD-L threshold selection.

Table 1: Determined threshold, estimated GPD parameters, and number of exceedances for the dengue residual series using the AD-L method.

u	$\hat{\sigma}$	$\hat{\xi}$	**Number of exceedances**
2498.211	7664.870	0.312	11

The results, including the selected threshold, estimated GPD parameters, and the number of exceedances, are summarized in Table 1. The positive shape parameter ($\xi > 0$) suggests heavy-tailed behavior in the residual extremes, implying that unexpected spikes in dengue prevalence can reach unusually high magnitudes beyond the fitted baseline.

Table 2: GOF test p-values (KS, AD, and CVM) for the fitted GPD model, assessing the adequacy of tail modeling.

GOF test	*p*-value
KS	0.245
AD	0.225
CVM	0.207

The GOF results for the GPD, presented in Table 2, show that the p-values for the KS, AD, and CVM tests are all above the 0.05 threshold, suggesting that the fitted model cannot be rejected. However, these values are not particularly close to 1, implying only moderate compatibility between the model and the observed exceedances. As emphasized by [40] and [24], higher p-values reflect stronger model conformity to empirical data. This relatively modest performance is likely attributable to the limited number of exceedances identified by the AD-L method, a direct consequence of the small sample size.

Table 3: Estimated return levels and 95% CIs for the residual series R_t, obtained from GPD modeling with the AD-L threshold selection.

Return Period	Return Level	95% CI
5	14221.420	[1533.512,34018.580]
10	22982.370	[7296.656,36997.27]
20	33858.150	[11171.570,44504.050]

Table 4: Return level estimates and corresponding 95% CIs converted from the residual scale to the original dengue prevalence scale, reflecting the expected magnitude of extreme prevalence events over different return periods.

Return Period	Return Level	95% CI
5	16013.920	[3326.010,35811.080]
10	24774.860	[9089.155,38789.770]
20	35650.650	[12964.070,46296.550]

Table 5: Estimated return levels and 95% CIs for extreme dengue prevalence during July, representing an early phase of the high-transmission season when cases begin to rise sharply.

Return Period	Return Level	95% CI
5	15093.760	[2405.854,34890.920]
10	23854.710	[8168.998,37869.610]
20	34730.490	[12043.910,45376.400]

The scarcity of exceedances introduces substantial variability in the tail estimates, a factor that is further reflected in the wide uncertainties observed later in the return level analysis. The scarcity of exceedances introduces substantial variability in the tail estimates, a factor clearly reflected in Table 3, which summarizes return level estimates for the residual series R_t. The 5-, 10-, and 20-year return levels increase steadily, but the corresponding CIs are wide, highlighting uncertainty in predicting extreme events from a limited sample. For instance, the 5-year return level is approximately 14,221 with a CI ranging from 1,534 to 34,019, illustrating the influence of small sample size on precision. When these estimates are converted to the original prevalence scale, as shown in Table 4, the results become more interpretable in a public health context.

The impact of seasonality is further examined in Table 5, which presents return levels specifically for July, the period marking the onset of the monsoon season. When considering July as a strategic point for assessment, the return levels are slightly higher, with the 5-year estimate rising to about 15,094 cases and the 20-year estimate reaching 34,730 cases. The corresponding CIs, although still wide, suggest an upward shift in the predicted magnitude of extremes during this month, reinforcing July's significance as a period conducive to early-stage dengue escalation. This seasonal relevance arises because July marks the onset of monsoon rains in Bangladesh, when increased humidity and temperatures near 30°C provide optimal conditions for mosquito breeding and virus transmission [11]–[13]. Although the intervals in both tables are broad, they remain epidemiologically useful. Predicting dengue extremes inherently involves uncertainty due to fluctuating factors such as future population growth, government control measures, and environmental changes. Thus, the wide CIs should not be interpreted as a lack of reliability; rather, they realistically capture the potential range of outcomes. Importantly, these estimates, despite their variability, offer actionable insights for public health authorities, supporting risk-based resource allocation and early intervention strategies during high-risk months. However, we still prefer narrower uncertainties for their credibility cause and accuracy.

3.2 Simulation Study

As outlined in Section 2.6, the simulation study is designed to investigate whether increasing the dataset length with observations that mimic the original structure improves the performance of the GPD model by reducing uncertainty and enhancing model compatibility. Evidence from Section 3.1 confirms that the original dengue prevalence series exhibits strong seasonality and an upward trend after 2020, necessitating a simulation approach that retains these components. The simulated prevalence series as shown in Figure 5, successfully extends the original data to 500 observations while maintaining its fundamental statistical structure. This series is reconstructed by combining the simulated residuals with the SARIMA baseline, ensuring realistic representation of variability without producing implausible negative values. Although the simulated dataset does not reproduce the exact temporal dynamics of the original dengue prevalence, it retains the essential heavy-tailed characteristics and variability required for extreme value modeling.

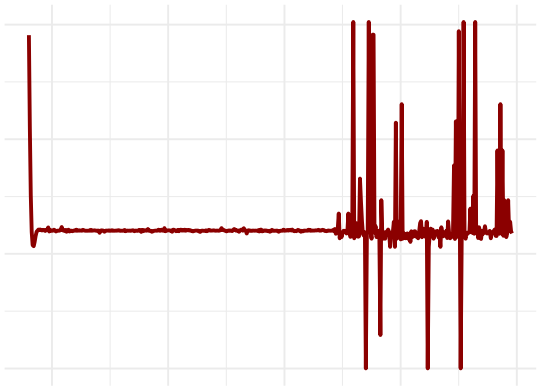

Figure 5: Time series plot of the simulated dengue prevalence (500 observations), preserving key statistical characteristics of the original dataset with controlled variability for extreme value analysis.

The aim was not to completely replicate the Bangladesh dengue prevalence dataset (which is beyond the scope of this paper), but to assess the extremal situation in a larger dataset with similar characteristics. The resulting parameter estimates, along with the identified threshold and number of exceedances, are summarized in Table 6.

Table 6: Estimated GPD parameters, selected threshold, and number of exceedances for the simulated dataset with N = 500.

u	$\hat{\sigma}$	$\hat{\xi}$	**Number of exceedances**
25813.930	30365.945	-3.159	7

Table 6 shows that even with 500 simulated observations, the number of exceedances identified by the AD-L method remains low. This outcome suggests that the simulated data, while preserving the overall statistical characteristics of the original series, did not generate a substantial increase in extreme residuals. The negative shape parameter further implies that the tail of the fitted GPD is bounded, which may restrict the occurrence of very large values.

Table 7: GOF test results for the GPD model fitted to simulated residuals (N = 500), showing p-values for KS, AD, and CVM tests.

GOF test	***p*-value**
KS	0.797
AD	0.973
CVM	0.925

Despite the limited number of exceedances, Table 7 demonstrates a substantial improvement in model compatibility, as reflected by the high p-values from all three GOF tests (p-values are much closer to 1). These values are markedly higher than those obtained from the original dataset, indicating that the fitted GPD aligns well with the simulated residual distribution. This improvement suggests that, even with fewer exceedances, the extended dataset facilitated a more stable fit than previously observed in the original data. The results imply that dataset extensions under similar conditions can enhance the reliability of extreme value modeling, at least in terms of GOF test performance.

Table 8: Estimated return levels and 95% CIs for the simulated residuals, showing expected exceedances over 5-, 10-, and 20-year periods.

Return Period	**Return Level**	**95% CI**
5	18752.420	[4059.688, 34018.350]
10	33559.570	[21312.350, 35215.590]
20	35217.210	[27911.930, 36689.790]

Table 9: Return levels and 95% CIs converted to the original prevalence scale for July, representing the onset of the peak dengue transmission season.

Return Period	**Return Level**	**95% CI**
5	43445.940	[28753.200, 58711.870]
10	58253.080	[46005.870, 59909.100]
20	59910.720	[52605.450, 61383.300]

As shown in Tables 8 and 9, extending the dataset under similar seasonal and trend characteristics did not substantially narrow the confidence intervals of the return levels. This

persistent uncertainty is primarily linked to the very limited number of exceedances in the simulated data, despite increasing the overall series length. Strong seasonality and the sudden sharp upward trajectory (see Figure 2) observed in the original dataset further complicate the modeling, as such dynamics can amplify variability in tail estimates. In real-world contexts, this variability mirrors the unpredictability of dengue prevalence. Outbreak intensities are shaped by environmental factors such as climate fluctuations and stagnant water sources, but they can also be mitigated by sudden positive changes like improved vector control, enhanced sanitation, and effective public health campaigns. These combined influences make future extremes inherently uncertain, explaining why even a longer dataset, when constructed under similar conditions, fails to eliminate the wide intervals observed in the return level estimates.

4. Conclusion

This study demonstrates the applicability of extreme value analysis, particularly the GPD with an automated threshold selection method (AD-L), to model extreme dengue prevalence in Bangladesh. By decomposing the original series through SARIMA modeling, removing trend and seasonality, and focusing on residual extremes, we were able to identify and quantify the rare yet impactful deviations that define severe outbreaks. The findings from the original dataset revealed that, while the GPD can capture the tail behavior, the small number of exceedances and the limited sample size introduced considerable uncertainty in return level estimates. Nevertheless, these estimates remain valuable for public health planning, as they provide a statistical perspective on the potential magnitude of future extremes.

The simulation study sought to evaluate whether extending the dataset under similar seasonal and trend conditions could improve the stability of extreme value modeling. Although model compatibility improved, reflected by markedly higher GOF p-values, the wide confidence intervals persisted. This persistence underscores the challenges posed by complex seasonality, sudden sharp upward trends, and the scarcity of extreme observations. Moreover, the variability seen in the results is consistent with the real- world nature of dengue, where extreme outbreaks are influenced by a combination of unpredictable environmental factors and potential sudden improvements in public health interventions.

Overall, this research highlights both the potential and limitations of GPD modeling for epidemiological extremes. It emphasizes that statistical models, while informative, must be interpreted alongside contextual knowledge of disease dynamics and public health measures. While the study successfully applied the AD-L method to model extremes and tested the impact of dataset extension, certain limitations became apparent. The hybrid simulation approach may not have produced enough upper-tail observations to fully capture rare extreme behavior, and the AD-L method under these conditions tended to select thresholds that limited the number of exceedances. These issues directly influenced the observed variability in return level estimates. These limitations open avenues for future research. Refining the simulation framework to better mimic the stochastic nature of extreme epidemiological events could enhance the realism of extended datasets. Similarly, adapting the AD-L method to account for datasets with strong seasonality and abrupt upward shifts, or developing new threshold selection techniques optimized for such conditions, could lead to more stable tail estimates. Such methodological advancements would not only improve the reliability of extreme value modeling in this study but also strengthen its applicability to broader epidemiological datasets where extremes play a critical role in public health preparedness.

Declarations

Conflict of Interest

There was no conflict of interest.

Ethics/Approval Declarations

Not applicable.

Consent for publication

Not applicable.

Consent to participate

Not applicable.

Availability of data and material/Data availability

Monthly dengue prevalence data for Bangladesh (January 2008–September 2023) are publicly available at

(https://figshare.com/articles/dataset/Monthly_ prevalence_of_dengue_in_Bangladesh_January_2008_-_September_ 2023_/24302470)

References

[1] S. Coles, J. Bawa, L. Trenner, and P. Dorazio, An introduction to statistical modeling of extreme values. Springer, 2001, vol. 208.
[2] M. Thomas, M. Lemaitre, M. L. Wilson, et al., “Applications of extreme value theory in public health,” PloS one, vol. 11, no. 7, e0159312, 2016.
[3] R. Sippy, D. Herrera, D. Gaus, R. E. Gangnon, J. A. Patz, and J. E. Osorio, “Seasonal patterns of dengue fever in rural ecuador: 2009-2016,” PLoS neglected tropical diseases, vol. 13, no. 5, e0007360, 2019.
[4] S. Bhatt, P. W. Gething, O. J. Brady, et al., “The global distribution and burden of dengue,” Nature, vol. 496, no. 7446, pp. 504–507, 2013.
[5] N. Haider, M. N. Hasan, J. Onyango, et al., “Global dengue epidemic worsens with record 14 million cases and 9,000 deaths reported in 2024,” International Journal of Infectious Diseases, p. 107 940, 2025.
[6] N. Haider, M. Hasan, J. Onyango, and M. Asaduzzaman, Global landmark: 2023 marks the worst year for dengue cases with millions infected and thousands of deaths reported. ijid regions 13, 2024.
[7] O. J. Brady, L. S. Bastos, J. M. Caldwell, et al., “Why the growth of arboviral dis- eases necessitates a new generation of global risk maps and future projections,” PLoS computational biology, vol. 21, no. 4, e1012771, 2025.
[8] M. S. Rahman, M. O. Faruk, S. Tanjila, N. M. Sabbir, N. Haider, and S. Chowdhury, “Entomological survey for identification of aedes larval breeding sites and their

distribution in chattogram, bangladesh," Beni-Suef University Journal of Ba- sic and Applied Sciences, vol. 10, no. 1, p. 32, 2021.
[9] N. Haider, "Modeling the vector-borne disease transmission potential in northern europe with a special emphasis on microclimatic temperature," 2018.
[10] P. Hasan, "Monthly prevalence of dengue in Bangladesh (January 2008 - Septem- ber 2023)," Oct.2023.DOI:10.6084/m9.figshare.24302470.v1.[Online]. Available:https://figshare.com/articles/dataset/Monthly_prevalence_of_dengue_in_Bangladesh_January_2008_September_2023_/24302470.
[11] S. Shahid, S. B. Harun, and A. Katimon, "Changes in diurnal temperature range in bangladesh during the time period 1961–2008," Atmospheric Research, vol. 118, pp. 260–270, 2012.
[12] J.-Y. Wu, Z.-R. Lun, A. A. James, and X.-G. Chen, "Dengue fever in mainland china," The American journal of tropical medicine and hygiene, vol. 83, no. 3, p. 664, 2010.
[13] H.-L. Yu, S.-J. Yang, H.-J. Yen, and G. Christakos, "A spatio-temporal climate- based model of early dengue fever warning in southern taiwan," Stochastic Environmental Research and Risk Assessment, vol. 25, no. 4, pp. 485–494, 2011.
[14] E. B. Yunus, A. M. Bangali, M. Mahmood, M. M. Rahman, A. Chowdhury, and K. Talukder, "Dengue outbreak 2000 in bangladesh: From speculation to reality and exercises.," 2001.
[15] S. Sharmin, K. Glass, E. Viennet, and D. Harley, "Interaction of mean temperature and daily fluctuation influences dengue incidence in dhaka, bangladesh," PLoS neglected tropical diseases, vol. 9, no. 7, e0003901, 2015.
[16] M. Rahman, K. Rahman, A. Siddque, et al., "First outbreak of dengue hemorrhagic fever, bangladesh," Emerging infectious diseases, vol. 8, no. 7, p. 738, 2002.
[17] A. S. Bonna, S. R. Pavel, T. Mehjabin, and M. Ali, "Dengue in bangladesh," IJID One Health, vol. 1, p. 100 001, 2023.
[18] A. C. Davison and R. L. Smith, "Models for exceedances over high thresholds," Journal of the Royal Statistical Society Series B: Statistical Methodology, vol. 52, no. 3, pp. 393–425, 1990.
[19] H. Tong, Non-linear time series: a dynamical system approach. Oxford university press, 1990.
[20] D. Lambert, "Zero-inflated poisson regression, with an application to defects in manufacturing," Technometrics, vol. 34, no. 1, pp. 1–14, 1992.
[21] J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels, Statistics of extremes: theory and applications. John Wiley & Sons, 2006.
[22] B. Bader, J. Yan, and X. Zhang, "Automated threshold selection for extreme value analysis via goodness-of-fit tests with application to batched return level mapping," arXiv preprint arXiv:1604.02024, 2016.
[23] B. Liang, Z. Shao, H. Li, M. Shao, and D. Lee, "An automated threshold selection method based on the characteristic of extrapolated significant wave heights," Coastal Engineering, vol. 144, pp. 22–32, 2019.
[24] S. Solari, M. Eguen, M. J. Polo, and M. A. Losada, "Peaks over threshold (pot): A methodology for automatic threshold estimation using goodness of fit p-value," Water Resources Research, vol. 53, no. 4, pp. 2833–2849, 2017.
[25] B. Liu and M. M. Ananda, "A new insight into reliability data modeling with an exponentiated composite exponential-pareto model," Applied Sciences, vol. 13, no. 1, p. 645, 2023.
[26] D. J. Dupuis, "Exceedances over high thresholds: A guide to threshold selection," Extremes, vol. 1, pp. 251–261, 1999.

[27] D. Gaigall and J. Gerstenberg, “Cramer-von-mises tests for the distribution of the excess over a confidence level,” Journal of Nonparametric Statistics, vol. 35, no. 3, pp. 529–561, 2023.
[28] C. Murphy, J. A. Tawn, and Z. Varty, “Automated threshold selection and associated inference uncertainty for univariate extremes,” Technometrics, pp. 1–10, 2024.
[29] R. M´ınguez, “Automatic threshold selection for generalized pareto and pareto–poisson distributions in rainfall analysis: A case study using the noaa ncdc daily rainfall database.,” Atmosphere, vol. 16, no. 1, 2025.
[30] F. K. Alif, N. Ali, and M. A. M. Safari, “An assessment on threshold selection for generalized pareto distribution using goodness of fit,” Malaysian Journal of Mathematical Sciences, 2025.
[31] T. S˘imkova´ and J. Picek, “A comparison of l-, lq-, tl-moment and maximum likelihood high quantile estimates of the gpd and gev distribution,” Communications in Statistics-Simulation and Computation, vol. 46, no. 8, pp. 5991–6010, 2017.
[32] J. R. Hosking, “Moments or l moments? an example comparing two measures of distributional shape,” The American Statistician, vol. 46, no. 3, pp. 186–189, 1992.
[33] J. Pickands III, “Statistical inference using extreme order statistics,” the Annals of Statistics, pp. 119–131, 1975.
[34] P. Embrechts, C. Klu¨ppelberg, and T. Mikosch, “Modelling extremal events springer-verlag,” Applications of mathematics, 1997.
[35] J. R. Hosking, “L-moments: Analysis and estimation of distributions using linear combinations of order statistics,” Journal of the Royal Statistical Society Series B: Statistical Methodology, vol. 52, no. 1, pp. 105–124, 1990.
[36] P. J. van Staden and M. Loots, “Method of l-moment estimation for the generalized lambda distribution,” in Proceedings of the Third Annual ASEARC Confer- ence, 2009, pp. 7–8.
[37] I. Abdul-Moniem and Y. M. Selim, “Tl-moments and l-moments estimation for the generalized pareto distribution,” Applied Mathematical Sciences, vol. 3, no. 1, pp. 43 –52, 2009.
[38] A. Luceno, “Fitting the generalized pareto distribution to data using maximum goodness-of-fit estimators,” Computational Statistics & Data Analysis, vol. 51, no. 2, pp. 904–917, 2006.
[39] S. Greenland, S. J. Senn, K. J. Rothman, et al., “Statistical tests, p values, confidence intervals, and power: A guide to misinterpretations,” European journal of epidemiology, vol. 31, no. 4, pp. 337–350, 2016.
[40] R. L. Wasserstein and N. A. Lazar, The asa statement on p-values: Context, process, and purpose, 2016.
[41] V. W. Berger and Y. Zhou, “Kolmogorov–smirnov test: Overview,” Wiley statsref: Statistics reference online, 2014.
[42] T. W. Anderson, “On the distribution of the two-sample cramer-von mises criterion,” The Annals of Mathematical Statistics, pp. 1148–1159, 1962.
[43] J. Chu, O. Dickin, and S. Nadarajah, “A review of goodness of fit tests for pareto distributions,” Journal of Computational and Applied Mathematics, vol. 361, pp. 13–41, 2019.
[44] V. Choulakian, R. A. Lockhart, and M. A. Stephens, “Cramer-von mises statistics for discrete distributions,” The Canadian Journal of Statistics/La Revue Canadienne de Statistique, pp. 125–137, 1994.
[45] M. M. Hasan, B. Croke, et al., “Filling gaps in daily rainfall data: A statistical approach,” 2013.

[46] P. Thompson, Y. Cai, D. Reeve, and J. Stander, “Automated threshold selection methods for extreme wave analysis,” Coastal Engineering, vol. 56, no. 10, pp. 1013–1021, 2009.
[47] C. Z. Mooney, R. D. Duval, and R. Duvall, Bootstrapping: A nonparametric approach to statistical inference. sage, 1993.
[48] R. J. Tibshirani and B. Efron, “An introduction to the bootstrap,” Monographs on statistics and applied probability, vol. 57, no. 1, pp. 1–436, 1993.
[49] B. Pfaff, Analysis of Integrated and Cointegrated Time Series with R, Second. New York: Springer, 2008, ISBN 0-387-27960-1. [Online]. Available: https://www.pfaffikus.de.
[50] R. Mushtaq, “Augmented dickey fuller test,” 2011.
[51] A. Trapletti and K. Hornik, Tseries: Time series analysis and computational fi- nance, R package version 0.10-58, 2024. [Online]. Available: https://CRAN. R-project.org/package=tseries.
[52] E. Syczewska, “Empirical power of the kwiatkowski-phillips-schmidt-shin test,” 2010.
[53] R. B. Cleveland, W. S. Cleveland, J. E. McRae, I. Terpenning, et al., “Stl: A seasonal-trend decomposition,” J. off. Stat, vol. 6, no. 1, pp. 3–73, 1990.
[54] R. J. Hyndman, G. Athanasopoulos, C. Bergmeir, et al., “Package ‘forecast’,”Online] https://cran. r-project. org/web/packages/forecast/forecast. pdf, 2020.
[55] J. Arlt and P. Trcka, “Automatic sarima modeling and forecast accuracy,” Communications in Statistics-Simulation and Computation, vol. 50, no. 10, pp. 2949–2970, 2021.
[56] G. M. Ljung and G. E. Box, “On a measure of lack of fit in time series models,” Biometrika, vol. 65, no. 2, pp. 297–303, 1978.

This page is intentionally left Blank