MATHEMATICAL MODELING AND COMPUTING, Vol. 12, No. 3, pp. 894–905 (2025)

Comparison of parametric and nonparametric estimation methods for annual precipitation in Kuala Lumpur

Ilias I. S. C.[1], Mustafa M. S.[2], Sidi N. S.[2]

[1]*Institute for Mathematical Research, Universiti Putra Malaysia, Malaysia*
[2]*Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, Malaysia*

(Received 11 February 2025; Revised 11 September 2025; Accepted 13 September 2025)

Flash floods are becoming a critical issue as they occur more frequently in recent years. Managing watersheds and water resources, researching floods and droughts, and monitoring climate change are all connected to annual precipitation. Therefore, discovering the most accurate method for calculating annual precipitation is crucial. This study compares two basic approaches to estimating annual precipitation parameters: parametric and nonparametric. The research focuses on fitting the distribution of annual precipitation for fifteen strategically located rain gauge stations scattered around Kuala Lumpur. These stations play a crucial role in providing comprehensive data for the study. The Generalized Extreme Value (GEV) distribution is utilized for parametric approaches with Maximum Likelihood Estimation (MLE) as the parameter estimator. Meanwhile, the kernel function using the Gaussian distribution is applied for the nonparametric method. Two approaches are used to compute the smoothing parameter: Silverman's Rule of Thumb (ROT) and the Adamowski Criterion (AC). The goodness-of-fit of the proposed models is assessed using the Mean Relative Deviation (MRD) and Mean Squared Relative Deviation (MSRD) statistics to evaluate nonparametric and parametric models. The results show that ROT was the best method compared to AC and MLE in fitting the distribution for the fifteen rainfall stations in Kuala Lumpur. According to the study, nonparametric approaches can be an alternative for estimating the annual precipitation in Kuala Lumpur.

Keywords: *kernel density estimation; maximum likelihood; generalized extreme value.*
2010 MSC: 62G05 **DOI:** 10.23939/mmc2025.03.894

1. Introduction

Precipitation investigation is a basic perspective of hydrological and natural things, especially in places like Kuala Lumpur, where the tropical climate results in considerable variability in precipitation. Understanding and anticipating precipitation designs are vital for water asset administration, flood forecasting, and urban arranging. One of the primary challenges in such studies is choosing the most appropriate statistical methods to model precipitation. Traditionally, these methods are divided into two broad categories: parametric and nonparametric estimation techniques.

This paper compares parametric and non-parametric approaches to estimating annual precipitation for Kuala Lumpur. Parametric strategies expect a particular distribution of the information, allowing more structured models with defined parameters. On the other hand, nonparametric strategies do not assume any distribution, making them more adaptable but possibly less efficient when strong assumptions about the data hold true. Selecting a suitable statistical model is basic in districts like Kuala Lumpur, where rainstorms and other climatic variables impact precipitation designs. The profoundly variable and extraordinary precipitation in such ranges requires vigorous modeling approaches to capture regular patterns, extremes, and fundamental distributions accurately. Extreme Value Analysis (EVA), utilizing strategies such as the Gumbel distribution or the Generalized Extreme Value (GEV) model, is broadly utilized to assess extraordinary precipitation occasions [1, 2]. By evaluating both approaches, we can offer insight into the preferences and limitations of each strategy for hydrological estimation in tropical climates.

Precipitation is significant in hydrological processes, affecting water accessibility, agribusiness, urban waste systems, and flood risks. Precise precipitation gauges can lead to way better flood management, forward agrarian arranging, and more productive plans of urban seepage frameworks. Kuala Lumpur encounters substantial annual precipitation, with articulated rainstorm seasons, making it inclined to flood and waterlogging [3]. Hence, modeling annual precipitation is critical to the city's planning efforts. Understanding the distinction between these approaches and their qualities and limitations is fundamental for selecting the most suitable show in Kuala Lumpur's interesting precipitation behavior.

Parametric estimation methods are widely used for modeling annual precipitation, particularly in regions like Kuala Lumpur where monsoon-driven rainfall exhibits distinct patterns and extremes. These methods assume the data follows a specific probability distribution, such as the Normal, Log-Normal, Gamma, or Generalized Extreme Value (GEV) distribution, and estimate parameters such as the mean, variance, or shape using techniques like Maximum Likelihood Estimation (MLE) or the Method of Moments [1, 4]. For instance, the GEV distribution is as often as possible applied in hydrology to model extreme precipitation events, providing insights into flood risks during monsoon seasons. The Log-Normal and Gamma distributions are particularly successful for highly skewed precipitation data, capturing inconstancy and long-tail characteristics common in yearly precipitation designs [5]. These parametric approaches are computationally effective and give a theoretical premise for understanding precipitation patterns, making them essential tools in hydrological modeling and foundation arranging.

Non-parametric estimation methods differ from parametric approaches in that they do not require assumptions about the underlying data distribution, making them highly flexible, particularly when the data distribution is unknown or complex. This flexibility makes non-parametric methods a preferred choice in scenarios where parametric models may fail or be unsuitable due to the data's complexity [6]. Kernel Density Estimation (KDE), a widely used non-parametric technique, estimates the probability density function of a random variable without assuming a specific distribution, instead constructing a smooth density estimate based on observed data [7]. In the context of precipitation estimation in Kuala Lumpur, KDE can effectively capture the distribution's tails, where extreme rainfall events are represented. Another outstanding non-parametric method is the Empirical Cumulative Distribution Function (ECDF), which gives a step function that increments at each observed data point, offering a direct likelihood estimate without depending on distributional assumptions. This approach is especially valuable for hydrologists who want to include sporadic or non-stationary data, such as Kuala Lumpur's precipitation patterns [8].

Despite their preferences, non-parametric strategies have limitations. They typically require larger sample sizes to achieve the same accuracy as parametric models, especially when estimating tail probabilities or extreme events. Additionally, non-parametric methods may be less efficient when data follows a known distribution, as they do not leverage the structural advantages provided by parametric models [9].

2. Application of parametric and non-parametric methods in Kuala Lumpur

In Kuala Lumpur, both parametric and non-parametric methods have been utilized in hydrological studies to estimate precipitation and predict flood risks. Parametric approaches, such as using Generalized Extreme Value (GEV) or Gamma distributions, are often employed to model extreme rainfall events and assess the likelihood of flooding [10]. These methods are particularly viable when a theoretical framework for understanding precipitation extremes is required. In any case, the tropical climate of Kuala Lumpur, characterized by rainstorm and convective storms, leads to complex and profoundly variable precipitation patterns, which in some cases challenge the assumptions of parametric models. As a result, non-parametric methods, such as Kernel Density Estimation (KDE), have gained popularity. These methods are used to represent precipitation intensity and frequency, particularly when parametric models fail to capture the full extent of precipitation variability [11]. Non-parametric methods offer expanded flexibility, enabling them to better handle anomalies and complexities inherent in tropical precipitation data.

3. Comparison of parametric and non-parametric methods

The choice between parametric and non-parametric methods for precipitation estimation in Kuala Lumpur depends on several factors, including the size of the dataset, the need for model flexibility, and the specific goals of the analysis. Parametric methods are generally more efficient when the expected distribution closely matches the data, making them appropriate for estimating central tendencies and assessing extremes. In any case, they can be biased if the wrong distribution is chosen. Non-parametric methods, whereas more adaptable, require larger datasets and may be less efficient in some contexts. In any case, they are invaluable when the data does not conform standard distributions or when flexibility in modeling is required. In Kuala Lumpur, where precipitation patterns are characterized by high variability and non-normal distributions, non-parametric methods offer a more precise representation of the fundamental precipitation distribution. These strategies do not rely on assumptions about the data's distribution, making them especially valuable in regions with complex climatic conditions, such as Kuala Lumpur, where rainstorm and convective storms make unpredictable precipitation patterns [6]. For example, Kernel Density Estimation (KDE) allows for the estimation of the probability density function directly from observed data, offering flexibility in modeling the full range of precipitation variability without assuming a specific distribution [7]. Studies have shown that non-parametric approaches are better suited for capturing extreme rainfall events and irregular distributions, which are common in tropical climates like that of Kuala Lumpur [11]. Additionally, non-parametric methods such as the Empirical Cumulative Distribution Function (ECDF) provide a straightforward way to estimate probabilities from data, further enhancing their applicability in hydrological studies in regions with complex rainfall patterns [8].

4. Materials and methods

This study used the annual precipitation data from rain gauges in Kuala Lumpur. About 25 rain gauge stations surround Kuala Lumpur; however, only 15 stations were active from 2012 to 2022, as shown in Table 1, which presents the annual maximum series for Kuala Lumpur. The data was obtained from the Department of Irrigation and Drainage (DID), Ministry of Environment and Water. The locations of the rain gauge stations are shown in Figure 1, and all rain gauge stations are distributed across Kuala Lumpur at different locations.

Table 1. Annual maximum series of 15 rain gauges in Kuala Lumpur.

Station \ Year	2010	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
S01	–	–	147	90	86.5	104	116.5	90	102	109.5	253.5	117
S02	–	–	84.5	95	76	65.5	108	105.5	79.5	86	127	94.5
S03	73	94	92	88.5	89	88	106.5	98	92	108.5	199.5	79.5
S04	–	–	100.5	82.5	105.5	70.5	104	88.5	86.5	181	223.5	86
S05	–	–	125.5	100	105.5	95.5	137.5	65	95.5	152.5	203	71.5
S06	–	–	91.5	90.5	85.5	75	95.5	69	127	112.5	194	78.5
S07	–	–	155	80.5	116	112	105	108.5	99.5	133.5	237	112
S08	–	–	145	103	134	91.5	105	77.5	139	155	262	108.5
S09	–	–	113	103	100	103.5	–	120	124.5	151	197.5	105
S10	–	–	146.5	84.5	111	108.5	117.5	95.5	143.5	140	251	140.5
S11	–	94	93	101	85	94	94.5	104	84.5	99.5	208	83.5
S12	–	–	135	92	76.5	74	78	115	117	156	214.5	84.5
S13	–	–	188.5	133	101.5	82.5	102.5	91.5	87.5	115.5	170.5	113
S14	–	–	116.5	100.5	104	97.5	67.5	90	75	103.5	236.5	122
S15	–	–	95.5	89.5	89	100	100.5	78.5	104.5	137.5	182.5	81

The choice of the Generalized Extreme Value (GEV) distribution and Kernel Density Estimation (KDE) for this study is well-justified given the nature of rainfall patterns in Kuala Lumpur. The GEV distribution is particularly suited for modeling extreme rainfall events, which are critical for flood risk assessment in monsoon-prone regions like Kuala Lumpur. This distribution allows for the estimation of the tail behavior of the rainfall data, capturing extreme precipitation events that can have significant hydrological impacts [4]. It is especially effective when the data exhibits heavy tails, as is typical of extreme weather events in tropical climates [10]. On the other hand, KDE provides a non-parametric

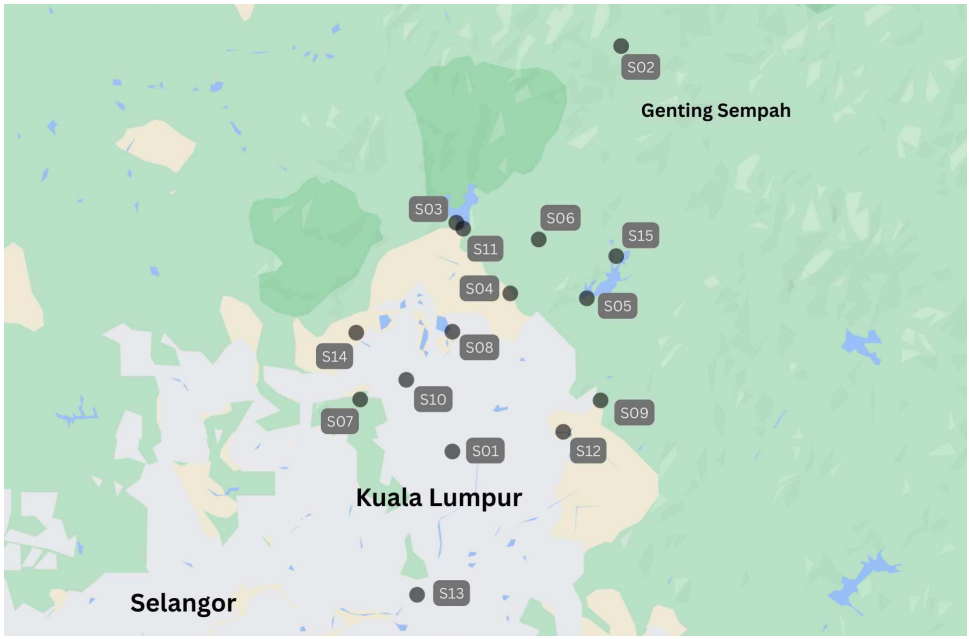

Fig. 1. Locations of 15 rain gauges stations in Kuala Lumpur.

approach to estimate the probability density function of precipitation without assuming a specific distribution. This flexibility is invaluable for capturing the irregularities and variability inherent in tropical rainfall, especially when the data deviates from common parametric distributions (Silverman, 1986). By combining GEV for extreme event modeling and KDE for a more flexible, assumption-free estimation, this study can provide a comprehensive understanding of the precipitation distribution and enhance flood prediction accuracy in Kuala Lumpur.

4.1. Parametric method

There are many methods for estimating parameters, such as maximum likelihood estimation (MLE), method of moments (MOM) and L-moments. Each parameter estimation technique has its benefits and drawbacks. However, it may be described in terms of unbiasedness, efficacy, and consistency to provide an optimum parameter estimate. When the estimated parameter is near the actual parameter, and the parameter estimation is accurate, it is claimed that it must be unbiased [12]. The approach that produces the minimum root mean square error (RMSE) demonstrates an effective estimator. This study considers parametric estimation methods most compatible with GEV distribution. Reference [13] explains that compared to other parametric approaches, MLE may provide the optimum parameter value since it increases the chance or combined likelihood of occurrence of the detected sample. In addition, MLE has ideal properties in estimation, including sufficiency (containing all relevant information about the parameter of interest), consistency (asymptotically data generated from the actual parameter value) and efficiency (achieves the lowest variance of parameter estimates and parameterization invariance [14]). Furthermore, [12] claims that MLE is the most often applied approach to estimate the GEV parameters as it has good asymptotic properties such as consistency and efficiency.

The Maximum Likelihood Estimator (MLE) has been chosen to estimate the Generalised Extreme Value (GEV) distribution parameter. Let $y_1, \ldots, y_n$ denote the annual maximum rainfall distribution; the probability density function of GEV is

$$L(y) = \prod_{i=1}^{n} f(y_i; \xi, \sigma, \mu) = \prod_{i=1}^{n} \frac{1}{\sigma} t(y)^{\xi+1} e^{-t(y)}, \tag{1}$$

where

$$t(y) = \begin{cases} \left(1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right)^{\frac{-1}{\xi}}, & \xi \neq 0, \\ e^{\frac{-(y-\mu)}{\sigma}}, & \xi = 0, \end{cases}$$

$$l(\xi,\sigma,\mu;y) = \ln(L(\xi,\sigma,\mu|y)) \\ = -n\ln\sigma + \sum_{i=1}^{n}\left[\left(\frac{1}{\xi}-1\right)\ln(Z_i) - (Z_i)^{\frac{1}{\xi}}\right], \tag{2}$$

where $Z_i = 1+\xi\left(\frac{y-\mu}{\sigma}\right)$. The MLE of $\hat{\xi}$, $\hat{\sigma}$, $\hat{\mu}$ can be achieved by differentiating $l(\xi,\sigma,\mu;y)$ with respect to respective parameters and equating the result to zero. This estimation technique is widely used because of its desirable properties, such as consistency and asymptotic efficiency, when the sample size is large [15].

Kolmogorov–Smirnov test The Kolmogorov–Smirnov (K-S) test is a non-parametric test used to determine whether a sample comes from a population that follows a specified distribution. One key advantage of the K-S test is that the distribution of its test statistic is independent of the underlying cumulative distribution function (CDF) being tested, which makes it particularly versatile and appealing in various applications [5, 16]. Another benefit is its precision, as it does not rely on large sample sizes for the approximations to hold, unlike the chi-square goodness-of-fit test, which requires sufficiently large sample sizes to ensure validity [16]. The K-S test is based on comparing the observed cumulative distribution function (ECDF) of the sample with the expected CDF of the hypothesized distribution. The test can be defined as H_o: The data follow a certain distribution and H_1: The data does not follow a certain distribution, where the test statistics can be calculated by

$$D = \max_{1<i<N}\left(F(Y_i) - \frac{i-1}{N}, \frac{i}{N} - F(Y_i)\right), \tag{3}$$

where F is the theoretical cumulative distribution of the distribution being tested which must be a continuous distribution, Y_i are ordered from smallest to largest value and N is the number of samples.

4.2. Non-parametric method

The kernel method. Nonparametric methods, by definition, impose little to no assumptions on the underlying data distribution. These methods are particularly useful when the form of the distribution is unknown or complex. The kernel density estimates $\hat{f}(x)$ at a point x is given by

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right), \tag{4}$$

where n is the number of observations, $h>0$ is the smoothing parameter, and $K(x)$ is the kernel function which should satisfy three conditions, namely [6] $\int_{-\infty}^{\infty} K(x)\,dy = 1$, $K(x)$ should be symmetrical and $K(x) \geqslant 0$.

Then, the kernel estimator of the distribution function by

$$F_h(x) = \int_{-\infty}^{x} f_h(x)\,dt = \frac{1}{n}\sum_{i=1}^{n} H\left(\frac{x-x_i}{h}\right), \tag{5}$$

where $H(u) = \int_{-\infty}^{x} K(x)\,dt$.

The accuracy of the predicted kernel density estimate is influenced by two key factors: the choice of the kernel function and the selection of the bandwidth. First, the kernel function determines the shape and smoothness of the estimated density, while the bandwidth controls the degree of smoothing applied to the data. A smaller bandwidth results in a more sensitive estimate that may overfit the data (too much sensitivity to noise), while a larger bandwidth results in a smoother estimate that may underfit the data (overly generalized). The optimal bandwidth is crucial to achieving a balance between these extremes [6].

For this study, the Gaussian kernel function was selected due to its desirable properties, particularly its symmetry and smoothness. The Gaussian kernel is widely used because it provides a smooth and continuous estimate, with no sharp discontinuities, making it well-suited for modeling complex data distributions, such as those found in rainfall data [7]. Its bell-shaped form ensures that it assigns higher weights to data points that are closer to the point of interest, which is particularly useful in capturing local features in the data, such as peaks or clusters in precipitation patterns. Additionally,

the Gaussian kernel is mathematically tractable and computationally efficient, which adds to its appeal in practical applications [6]. By combining this kernel with an appropriately selected bandwidth, we can achieve an accurate and smooth estimate of the underlying precipitation distribution, essential for reliable rainfall prediction and flood risk assessment. Therefore,

$$K(x) = \frac{1}{\sqrt{2\pi h}}\, e^{-\frac{(x-x_i)^2}{2h^2}}. \tag{6}$$

Choosing the smoothing parameter. The fundamental issue in applying KDE is choosing the smoothing parameter, known as bandwidth selection. The optimal value of h is based on the value that minimises Integrated Mean Squared Error (IMSE), known as the optimality criterion. IMSE is given by [6, 9]:

$$\text{IMSE} = \int_{-\infty}^{\infty} \mathbb{E}\big[\hat{f}(x) - f(x)\big]^2 dx, \tag{7}$$

where $\hat{f}(x)$ is the estimate of $f(x)$ unknown density function.

An excessively smoothed probability density, with suppressed modes and excessively accentuated tails, is produced by a broad bandwidth. On the other hand, a narrow bandwidth may provide density estimates with observable peaks in the probability density's tails. In this study, two selection methods will be compared between Adamowski criterion (AC) proposed by Adamowski (1989) and Rule of Thumb (ROT) proposed by Silverman (1987).

Adamowski criterion. Adamowski and Labatiuk (1987) discover that using IMSE criteria, all different numerical techniques for calculating h_{AC} behave similarly and are all somewhat near to the optimal value anticipated by theory. The optimal value of h_{AC} may be represented as:

$$h_{AC} = \frac{\sum_{i=2}^{n}\sum_{j=1}^{i-1}(x_i - x_j)}{\sqrt{5}n\left(n - \frac{10}{3}\right)}, \tag{8}$$

where x_i and x_j are order statistics of observations, and n is the number of samples. Then the kernel estimator with Gaussian kernel function can be calculated from equation below:

$$\frac{1}{n}\sum_{i=1}^{n} \frac{1}{h\sqrt{2\pi}}\, e^{-0.5\left(\frac{x-x_i}{h}\right)^2}. \tag{9}$$

Rule of thumb. To reduce the value of Asymptotic Mean Integrated Squared Error (AMISE), Silverman (1987) developed ROT. The optimum combination between asymptotic variance and bias is provided by

$$h_\infty = \left(\frac{R(K)}{\mu_2^2(K)R(f^2)}\right)^{\frac{1}{5}} n^{-\frac{1}{5}}, \tag{10}$$

where h_∞ is the optimal bandwidth (also known as the "asymptotic bandwidth"), $R(K)$ is the integral of the squared kernel function $R(K) = \int_{-\infty}^{\infty} K^2(x)\,dx$, $mu_2(K)$ is the second moment of the kernel function $K(x)$, i.e., $\mu_2(K) = \int_{-\infty}^{\infty} x^2 K(x)\,dx$, $R(f^2)$ is the integral of the squared second derivative of the true density function $f(x)$, i.e., $R(f'') = \int_{-\infty}^{\infty}(f''(x))^2\,dx$ and n is the sample size.

Assuming the data follows the normal distribution with Gaussian kernel function such that $R(f^2) = \frac{3\sigma^{-5}}{8\sqrt{\pi}}$ and $R(K) = \frac{1}{2\sqrt{\pi}}$, we obtain $h = 1.3643\delta\, n^{-\frac{1}{5}}$. However, to address his concern that the value h above smoothes non-unimodal distributions excessively, Silverman (1987) suggests that the smoothing parameter be set at a little lower value in the equation below:

$$h_{\text{ROT}} = 0.9\min\left(\hat{\sigma}, \frac{\text{IQR}}{1.34}\right) n^{-\frac{1}{5}}. \tag{11}$$

The ROT offers a practical and widely used approach for selecting the bandwidth in kernel density estimation. However, its accuracy may be compromised if the underlying population is not normally distributed. In such cases, the ROT may not provide the optimal bandwidth, leading to suboptimal smoothing and inaccurate density estimation.

4.3. The goodness of fit test

Olofintoye and Adeyemo (2012) claim that the best-fit model was selected using statistical tests (goodness of fit test) based on the mathematical expression produced for each parametric and nonparametric density function. This research employed the Mean Relative Deviation (MRD) and Mean Squared Relative Deviation (MSRD) statistics to evaluate the fit of constructed models, comparing the nonparametric and parametric techniques. The following are the definitions of these statistical terms:

$$\text{MRD} = \frac{1}{n}\sum_{i=1}^{n}\frac{|x_i - \hat{x}_i|}{x_i} \times 100, \tag{12}$$

$$\text{MSRD} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \hat{x}_i}{x_i} \times 100\right)^2, \tag{13}$$

where n is the number of samples, x_i represents the observed highest annual precipitation values and $\hat{x}_i$ represents the calculated highest annual precipitation values, while x_i denotes the observed values, respectively. Lower MRD and MSRD values generally implied a better fit [17]. Hence, MRD and MSRD values were calculated for each constructed model, and the best-fit model was selected based on the lowest values of these two metrics.

5. Result and discussion

Various reasons, such as equipment state, site conditions, and maintenance programs, might have caused data loss in rainfall data records leading to only certain years with minimal data loss being selected for the period 2012–2022. Maximum annual precipitation data from fifteen rain gauge stations in Kuala Lumpur were fitted to parametric and nonparametric methods.

Table 2. Parameters estimation by MLE.

Station	Location, $\hat{\mu}$	Scale, $\hat{\sigma}$	Shape, $\hat{\xi}$
S01	97.1071	13.6146	0.7594
S02	85.0124	14.9730	−0.1213
S03	87.6251	12.6766	0.2875
S04	88.3381	18.4943	0.4931
S05	96.2806	28.9161	0.0708
S06	83.8706	15.5779	0.4315
S07	106.5247	22.4012	0.2228
S08	107.7754	28.1326	0.2395
S09	106.0476	8.4240	0.9673
S10	112.6865	25.8847	0.2040
S11	89.1955	7.2796	0.7123
S12	84.3629	15.2276	1.0506
S13	99.0371	18.0855	0.4315
S14	90.8675	21.9722	0.2629
S15	90.0388	12.8254	0.4782

Table 3. Bandwidth coefficient by Silverman's rule of thumb and Adamowski criterion.

Station	Silverman's ROT	AC
S01	10.1177	13.9933
S02	9.3761	6.3560
S03	4.8011	7.1318
S04	8.0518	14.9492
S05	16.5273	14.1845
S06	11.8658	11.3369
S07	9.8528	13.1615
S08	16.9511	16.3848
S09	9.0888	10.3648
S10	14.2495	14.3656
S11	4.6775	7.5301
S12	21.5597	14.9358
S13	14.6733	12.0412
S14	9.1112	13.4500
S15	6.0918	9.7034

For data analysis by parametric methods, the MLE method was used to estimate the parameters of GEV distribution. Table 2 shows the estimated values of $\hat{\mu}$, $\hat{\sigma}$, and $\hat{\xi}$ representing the GEV distribution's location, scale, and shape parameters. The nonparametric approach using the Gaussian kernel function fits the annual maximum precipitation data from the locations of 15 rain gauges described above. Two techniques, namely Silverman's ROT and AC, will be employed to determine the bandwidth coefficient, h, when analysing data using the nonparametric kernel method. The values of h for ROT and AC are tabulated below in Table 3. This result shows that ROT is deemed a much superior bandwidth selector compared to AC due to most of the stations' lower values of MRD and MSRD. Nevertheless, the lowest value of h will be selected to calculate MRD and MSRD to compare the performance of the parametric and nonparametric methods [17]. The optimal bandwidth calculated by both methods yielded a smaller value and will be applied to calculate the MRD and MSRD listed in Table 4.

From Table 3, most stations with lower values of h (roughly 60%) are ROT values, indicating that Silverman's ROT is a more accurate estimator of the smoothing parameter for the Gaussian kernel. Hence, it is deemed that ROT is a much superior bandwidth selector compared to AC and will be selected to calculate the value of MRD and MSRD to compare the performance of the parametric and nonparametric methods [17]. For instance, as shown in Table 4, the MRD and MSRD values for the nonparametric model are significantly lower than those for the parametric model, suggesting that the nonparametric method more accurately represents the observed data. This confirms that the nonparametric approach is superior in terms of model fit, as it minimizes the discrepancy between observed and predicted values more effectively than the parametric approach. Another result of this study is that KDE with Gaussian kernel function and ROT as bandwidth selector is the best among the parametric GEV distributions and KDE with AC as bandwidth selector. Table 4 compares parametric and nonparametric methods for analysing rainfall data across 15 stations, presenting each method's Mean Relative Difference (MRD) and Mean Squared Relative Difference (MSRD). Generally, parametric methods yield higher MRD and MSRD values than their nonparametric counterparts. For example, at Station S01, the parametric MRD is 56.5739 with an MSRD of 10820.4623, while the nonparametric MRD and MSRD are significantly lower at 23.5609 and 940.6689, respectively.

This suggests that parametric methods, which assume a fixed statistical distribution for rainfall data, may struggle to capture the real variability, leading to greater estimation errors when the actual distribution deviates from the assumed model. Nonparametric methods, by contrast, make no such assumptions and seem to perform better, particularly in regions where rainfall patterns are highly irregular.

Table 4. MRD and MSRD using parametric and nonparametric methods.

Station	Parametric		Nonparametric		
	MRD	MSRD	h	MRD	MSRD
S01	56.5739	10820.4623	10.1177	23.5609	940.6689
S02	17.4364	792.3782	6.3560	15.4311	114.2151
S03	20.6871	707.7760	4.8011	16.4163	440.6520
S04	36.9443	2943.0673	8.0518	30.0981	1604.3139
S05	34.9697	2482.1666	14.1845	29.5921	1240.6022
S06	31.9003	2361.2267	11.3369	23.9863	748.9548
S07	44.8352	3991.0241	9.8528	22.3572	654.2060
S08	53.1045	3766.3836	16.3848	27.0167	1222.8588
S09	30.5836	666.7029	9.0888	16.8291	187.9080
S10	29.5271	1314.8314	14.2495	22.7583	727.0147
S11	37.3841	2130.5950	4.6775	15.2836	415.0120
S12	53.4976	7150.9177	14.9358	30.1477	1386.0859
S13	40.5051	4234.4934	12.0412	22.7461	500.7414
S14	58.0500	9670.6768	9.1112	24.2583	1215.2512
S15	20.8210	852.8541	6.0918	19.1042	372.1644

The nonparametric method also includes a bandwidth parameter (h), which determines the degree of smoothing in the data estimation. Higher bandwidth values, like 16.9511 at Station S08, result in smoother estimations, while lower values, such as 4.6775 at Station S11, provide more detailed estimations but might overfit the data. Stations like S08 and S12, with very high MRD and MSRD values, illustrate that parametric models may be inadequate in capturing extreme rainfall behaviors in such regions. Meanwhile, nonparametric models consistently show lower MSRD values, indicating their greater flexibility and suitability for modeling complex and variable rainfall data, such as the annual precipitation patterns in Kuala Lumpur.

The kernel density plots for stations S01 to S15 illustrate in Figure 2 varying rainfall distribution patterns across Kuala Lumpur. In these plots, Silverman's Rule of Thumb (ROT) and the Adaptive Coefficient (AC) are used as bandwidth selection. Each method affects the visualization and interpretation of the data differently.

Silverman's Rule of Thumb (ROT) is a more traditional approach, using a fixed bandwidth derived from the data's standard deviation and the number of data points. This method tends to produce smoother density curves, which can be beneficial for identifying general trends and avoiding overfitting where data is sparse. For example, in the plots for stations like S01 and S04, ROT generates smooth curves, providing a broad overview of rainfall distribution that highlights major trends without catching

up in minor data fluctuations. This can be particularly useful in environmental applications where understanding general weather patterns is more relevant than capturing every small anomaly.

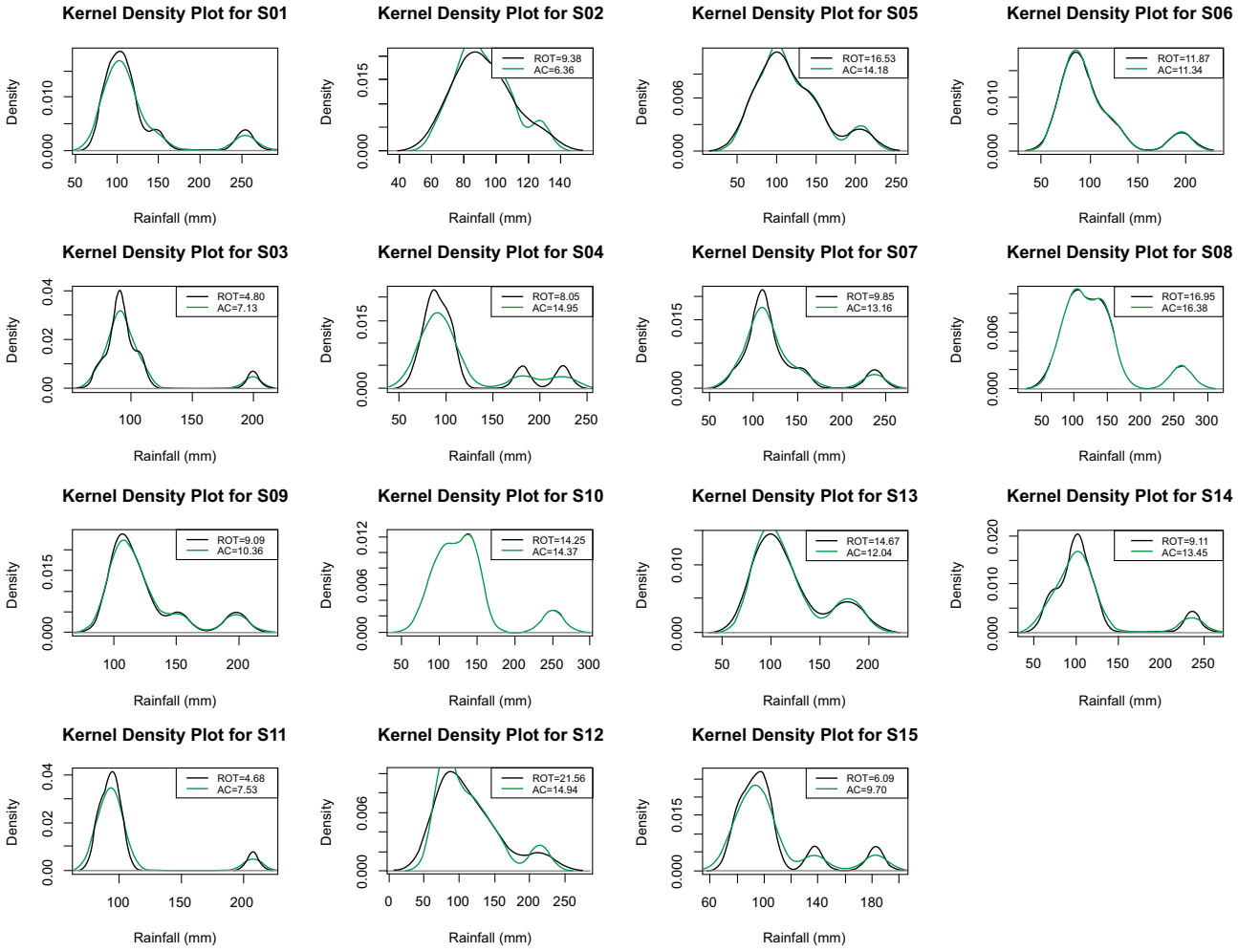

Fig. 2. Kernel density plot of 15 rain gauges in Kuala Lumpur.

On the other hand, the Adaptive Coefficient (AC) adjusts the bandwidth based on local data variability, allowing it to reflect more detailed features in the distribution, such as sharper peaks or multiple modes. This method is evident in its ability to delineate finer details in the rainfall data, as seen in the plots for stations S05 and S12. These detailed plots can reveal multiple rainfall modes or shifts in rainfall intensity that ROT might smooth over. For example, station S15's plot using AC shows a clear peak around 140 mm, which is less pronounced under ROT, suggesting that AC might be better suited for detecting specific rainfall events that could be crucial for local planning and response efforts.

Across the board, these kernel density plots serve as a tool for understanding rainfall behavior at each station. They help meteorologists, urban planners, and environmental researchers discern patterns that could influence flood forecasting, water resource management, and urban planning. The specific needs of the analysis should guide the choice between ROT and AC: ROT for broader trends useful in long-term planning and policymaking, and AC for operational decisions and event-specific analyses where understanding the nuances of rainfall distribution is crucial. Overall, these plots are not just historical data charts but analytical tools that provide insights into how rainfall is distributed over time and space in Kuala Lumpur. They help in making informed decisions about resource allocation, disaster preparedness, and environmental conservation based on the variability and intensity of rainfall captured by different statistical methods.

6. Conclusion

The comparison between parametric and nonparametric approaches in modeling rainfall data across various stations in Kuala Lumpur highlights significant differences in how each method captures rainfall distribution patterns. The parametric approach, typically using models like the Generalized Extreme Value (GEV) distribution, assumes a specific distribution shape, which can limit its ability to capture the full complexity of the data. In contrast, nonparametric methods, specifically Kernel Density Estimation (KDE) with bandwidth selection techniques such as Silverman's Rule of Thumb (ROT) and the Adaptive Coefficient (AC), provide more flexibility in modeling rainfall data without assuming a fixed distribution. ROT offers a smoother, more generalized view of the rainfall distribution. By selecting a fixed bandwidth, it captures the overall trends in the data but may overlook local variations or secondary features. This method provides a broad overview of the rainfall patterns but may not adequately highlight finer details, such as subtle fluctuations in the data. AC, with its adaptive bandwidth approach, is better suited to capturing local density variations. It adjusts the bandwidth according to the data's characteristics, revealing more detailed nuances in the rainfall distribution. This method is particularly effective in stations with higher rainfall variability, as it can identify secondary peaks or multimodal distributions that ROT might miss.

The findings from these methods show that rainfall patterns in Kuala Lumpur differ significantly across stations. Some stations, like S01, S03, and S08, exhibit bimodal or multimodal distributions, indicating irregular rainfall events. These complex distributions suggest the possibility of extreme rainfall events in specific years, which is critical for managing flood risks and planning water resource strategies. The variability in rainfall patterns across stations underscores the need for more accurate modeling and forecasting tools to address the challenges posed by Kuala Lumpur's diverse rainfall characteristics. While both ROT and AC methods provide valuable insights, combining both for different stages of analysis may be more effective. ROT can be used to identify long-term rainfall trends and provide an overall picture of the precipitation distribution, while AC can be employed to detect local variations and extreme weather events. This hybrid approach would provide a more balanced and detailed understanding of rainfall behavior. Multiple peaks in the rainfall distributions at certain stations highlight the possibility of extreme rainfall events. Further research focusing on extreme value analysis or Generalized Extreme Value (GEV) modeling could help better understand and predict these events, which is critical for flood risk management. The observed variability in rainfall patterns, particularly in stations that show high variability in rainfall distribution, suggests that localized weather planning is necessary. Local authorities should tailor flood management and infrastructure development projects based on specific station-level rainfall patterns rather than on generalized city-wide models. Since the AC method is sensitive to local data densities, improvements in the spatial and temporal resolution of rainfall data collection would enhance the accuracy of future rainfall estimations. More frequent rainfall measurements and expanded coverage across additional stations could lead to better rainfall distribution models, improving flood prevention strategies.

By incorporating these suggestions, a more robust framework for rainfall analysis can be developed. This framework would help address both long-term trends and immediate weather-related risks, particularly in the context of Kuala Lumpur's vulnerability to extreme rainfall events and flooding.

[1] Gumbel E. J. Statistics of Extremes. Columbia University Press (1958).

[2] Koutsoyiannis D. Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation. Hydrological Sciences Journal. **49** (4), 575–590 (2004).

[3] Malaysian Meteorological Department (MetMalaysia). Annual Climate Report 2020. Malaysian Meteorological Department (2020).

[4] Hosking J. R. M., Wallis J. R. Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge University Press (1997).

[5] Wilks D. S. Statistical Methods in the Atmospheric Sciences. Academic Press (2011).

[6] Silverman B. W. Density Estimation for Statistics and Data Analysis. Chapman and Hall (1986).

[7] Rosenblatt M. Remarks on some nonparametric estimates of a density function. The Annals of Mathematical Statistics. **27** (3), 832–837 (1956).

[8] Cunnane C. Unbiased plotting positions – A review. Journal of Hydrology. **37** (3–4), 205–222 (1978).

[9] Wand M. P.,Jones M. C. Kernel smoothing. pp. 10–57. Springer US (1995).

[10] Li Y., Fowler H. J., Argüeso D., Blenkinsop S., Evans J. P., Lenderink G., Yan X., Guerreiro S. B., Lewis E., Li X. F. Strong intensification of hourly rainfall extremes by urbanization. Geophysical Research Letters. **47** (14), e2020GL088758 (2020).

[11] Holešovský J., Fusek M., Blachut V., Michálek J. Comparison of precipitation extremes estimation using parametric and nonparametric methods. Hydrological Sciences Journal. **61** (13), 2376–2386 (2016).

[12] Roslan R. A., Na C. S., Gabda D. Parameter estimations of the generalised extreme value distributions for small sample size. Mathematics and Statistics. **8** (2A), 47–51 (2020).

[13] Hamzah F. M., Tajudin H., Jaafar O. A comparative flood frequency analysis of high-flow between annual maximum and partial duration series at Sungai Langat basin. Sains Malaysiana. **50** (7), 1843–1856 (2021).

[14] Myung I. J. Tutorial on maximum likelihood estimation. Journal of Mathematical Psychology. **47** (1), 90–100 (2003).

[15] Coles S. An Introduction to Statistical Modeling of Extreme Values. Springer London (2021).

[16] Stephens M. A. EDF Statistics for Goodness of Fit and Some Comparisons. Journal of the American Statistical Association. **69** (347), 730–737 (1974).

[17] Jou P. H., Akhoond-Ali A. M., Behnia A., Chinipardaz R. A comparison of parametric and nonparametric density functions for estimating annual precipitation in Iran. Research Journal of Environmental Sciences. **3** (1), 62–70 (2009).

Порівняння параметричної та непараметричної оцінки річних опадів у Куала–Лумпурі

Іліас І. С. Ч.[1], Мустафа М. С.[2], Сіді Н. С.[2]

[1]*Інститут математичних досліджень, Університет Путра Малайзія, Малайзія*
[2]*Кафедра математики та статистики, Факультет природничих наук, Університет Путра Малайзія, Малайзія*

Раптові повені стають критичною проблемою, оскільки вони трапляються все частіше в останні роки. Керування водозбірними басейнами та водними ресурсами, дослідження повеней та посух, а також моніторинг зміни клімату — все це пов'язано з річними опадами. Тому пошук найточнішого методу розрахунку річних опадів є надзвичайно важливим. У цьому дослідженні порівнюються два основні підходи до оцінки параметрів річних опадів: параметричний та непараметричний. Дослідження зосереджено на підборі розподілу річних опадів для п'ятнадцяти стратегічно розташованих опадомірних станцій, розкиданих по всьому Куала–Лумпуру. Ці станції відіграють вирішальну роль у наданні комплексних даних для дослідження. Розподіл узагальнених екстремальних значень (GEV) використовується для параметричних підходів з оцінкою максимальної правдоподібності (MLE) як оцінкою параметрів. У той же час, функція ядра з використанням гауссового розподілу застосовується для непараметричного методу. Для обчислення параметра згладжування використовуються два підходи: емпіричне правило Сільвермана (ROT) та критерій Адамовського (AC). Ступінь відповідності запропонованих моделей оцінюється за допомогою статистики середнього відносного відхилення (MRD) та середньоквадратичного відносного відхилення (MSRD) для оцінки непараметричних та параметричних моделей. Результат показує, що ROT був найкращим методом порівняно з AC та MLE для апроксимації розподілу для п'ятнадцяти станцій вимірювання опадів у Куала-Лумпурі. Згідно з дослідженням, непараметричні підходи можуть бути альтернативою пошуку даних про річні опади в Куала–Лумпурі.

Ключові слова: *оцінка щільності ядра; максимальна правдоподібність; узагальнене екстремальне значення.*