Journal of Quality Measurement and Analysis *JQMA* 22(1) 2026, 97-122
e-ISSN: 2600-8602 *http://www.ukm.my/jqma*
https://doi.org/10.17576/jqma.2201.2026.06

MODELLING EXTREME RAINFALL PATTERNS IN PENINSULAR MALAYSIA USING SPATIAL EXTREME VALUE APPROACHES

(*Pemodelan Corak Hujan Ekstrem di Semenanjung Malaysia Menggunakan Pendekatan Nilai Ekstrem Ruang*)

FARAH ADIBAH SYAHIZAM & SYAFRINA ABDUL HALIM*

ABSTRACT

Extreme rainfall poses significant challenges to communities, infrastructure, and ecosystems, particularly in tropical regions like Peninsular Malaysia. Recent decades have seen increasing rainfall variability and intensity due to climate change and monsoonal influences. This study models extreme rainfall using the generalized extreme value (GEV) distribution and extends the analysis spatially through max-stable processes (MSP), specifically the Smith and Schlather models. Using block maxima GEV marginals and MSP spatial models Smith and Schlather, this study quantifies spatial dependence in annual maxima from 11 stations (19902023) selected for data completeness and geographic coverage (coastal/inland). Missing data were imputed using inversedistance weighting (IDW). The return levels for 5-, 10-, 20-, and 30-year periods are assessed. The MSP approach enables modeling of spatial extremes by transforming return levels into Frchet margins. Parameters are estimated using maximum pairwise likelihood estimation (MPLE), and model performance is evaluated using the Takeuchi information criterion (TIC) and root mean square error (RMSE). Allowing a timevarying range parameter λ(t) within the Schlather model improves returnlevel estimates relative to stationary alternatives. These results support floodresilient design such as drainage capacity, dam safety and zoning. This study acknowledge nonstationarity due to climate change as a key extension and note that the framework can incorporate climate covariates.

Keywords: extreme rainfall; max-stable processes; Schlather; Smith; generalized extreme value; spatial extreme value

ABSTRAK

Hujan lebat ekstrem memberi cabaran besar kepada komuniti, infrastruktur, dan ekosistem, terutamanya di kawasan tropika seperti Semenanjung Malaysia. Dalam beberapa dekad kebelakangan ini, perubahan iklim dan pengaruh monsun telah menyebabkan peningkatan dalam kepelbagaian dan intensiti hujan. Kajian ini menggunakan taburan nilai ekstrem teritlak (GEV) untuk memodelkan hujan lebat ekstrem, dan mengembangkan analisis secara ruang melalui proses maks-stabil (MSP) dengan menggunakan model Smith dan Schlather. Menggunakan marjin GEV berasaskan blok maksima dan model ruang MSP Smith and Schlather, kajian ini mengkuantifikasi kebergantungan ruang dalam maksimum tahunan daripada 11 stesen (19902023) yang dipilih berdasarkan kelengkapan data dan liputan geografi (pesisir/pedalaman). Data hilang diimput menggunakan pemberat jarak songsang (IDW). Aras pulangan bagi tempoh 5, 10, 20 dan 30 tahun dinilai. Pendekatan MSP membolehkan pemodelan ekstrem ruang dengan menukarkan marjin kepada Frchet. Parameter dianggar menggunakan penganggaran kebarangkalian berpasangan maksimum (MPLE), manakala prestasi model dinilai menggunakan Kriteria Maklumat Takeuchi (TIC) dan ralat punca min kuasa dua (RMSE). Membenarkan parameter julat yang berubah mengikut masa, λ(t), dalam model Schlather meningkatkan anggaran aras pulangan berbanding alternatif pegun. Dapatan ini menyokong reka bentuk berdaya tahan banjir seperti kapasiti saliran, keselamatan empangan dan perzonan. Kajian ini mengakui ketidakpegunan akibat perubahan iklim sebagai peluasan penting dan bahawa rangka kerja ini boleh memasukkan kovariat iklim.

Kata kunci: hujan lebat ekstrem; proses maks-stabil; Schlather; Smith; nilai ekstrem teritlak; nilai ekstrem ruang

1. Introduction

Recent flood events highlight the need for robust modeling of extreme rainfall under monsoonal dynamics. In the last 20 years, Malaysia has seen warming and erratic rainfall (Saimi *et al.* 2020). Reliable returnlevel estimation is essential for planning and resilience. Prior Malaysian studies often use stationary GEV. The GEV distribution is widely used for modelling extreme values. The GEV distribution has three distribution types known as Gumbel, Frchet, and Weibull with three parameters in which the shape parameter is always kept constant (Abdelmoaty & Papalexiou 2023), but one or both of the location and scale parameters are allowed to vary (Agilan & Umamahesh 2017). Extreme value theory (EVT) defines the probability of an extreme event according to a distributions tail behavior (Coles *et al.* 2001). There exist two ways to define the extreme value known as block maxima (BM) and peak over threshold (POT). The POT technique is based on taking the value that above a threshold, whereas the BM approach takes the maximum value in a block, or one period (McNeil 1999). Advances in spatial extremes modeling enable explicit dependence (Smith/Schlather) and nonstationary structures (Huser & Wadsworth (2022); Huser *et al.* (2025)). Spatial extreme modelling could provide insights to the spatial dependence of extreme rainfall across regions (Zeng *et al.* 2022). This approach model the space dependence of extreme value, which is necessary to understand the relations between extreme rainfall of different locations.

2. Data and Methods

A max-stable process (MSP) is a widely known spatial extremes model when addressing BM originating from fixed-spatial observations (Huser & Wadsworth 2022). Based on Deni *et al.* (2010), MSP is an extension of univariate EVT models for BM to infinite dimensions. BM does not always converge to a single parametric family of MSP, which contradicts univariate EVT models. However, a variety of parametric families have been presented, each having a distinct tail-dependence function. Within an EVT framework, the direct calculation of areal-based exceedances is made possible by using MSP model. It is possible to conceptualize MSP model fitting as a two-stage process that involves trend surface and basic MSP model selection, where each phase assumes independence between the fixed unit Frchet margins and the extremes, respectively (Ribatet *et al.* 2018). A study was conducted that also examined how the choice of general MSP model fitting method affects the area exceedance estimate, even though a unique and successful trend surface modelling strategy was used (Love *et al.* 2022; Ribatet *et al.* 2018). Studies done by Karuna *et al.* (2023) used a constant range parameter, λ, that determines the spatial scale of correlation in the Schlather model. However, in this study, λ is allowed to vary in time, allowing $\lambda(t)$ to capture evolving correlation. Thus, the objective of this study is to develop a spatial GEV model using MSP to model the extreme rainfall in Peninsular Malaysia. In particular, the performance of Smith and Schlather with both constant and time-covariate range parameter will be compared. The parameter will be estimated by the maximum pairwise likelihood estimation (MPLE) method. The most appropriate model will be selected based on the Takuechi Information Criterion (TIC) value and the return level of extreme rainfall for 5, 10, 20 and 30 years will be further estimated.

2.1. *Data*

In this study, daily rainfall data from 11 meteorological stations in Peninsular Malaysia from 1990 to 2023 are used for analysis. The data were sourced from the Sistem Pengurusan Rangkaian Hidrologi Nasional (SPRHiN) and the Malaysian Meteorological Department (MetMalaysia). All stations were chosen based on data completeness and quality, ensuring reliable extreme value analysis across different regions. The list of stations is listed in Table 1, while the map of stations can be referred in Figure 1.

Table 1: Study area information

Name of Station	Station ID	(Lat (°), Lon (°))
IBU BEKALAN KAHANG AT JOHOR	0500551RF	(2.23, 103.59)
CHEMBONG AT NEGERI SEMBILAN	0270121RF	(2.61, 102.10)
KLINIK BKT. BENDERA AT P.PINANG	1910051RF	(5.42, 100.27)
SG. GAWI AT TERENGGANU	0670281RF	(5.15, 102.79)
KEDAH PEAK AT KEDAH	0030021RF	(5.79, 100.43)
LDG. LENDU AT MELAKA	0290101RF	(2.36, 102.19)
IBU BEKALAN TALANG AT KUALA KANGSAR PERAK	0180381RF	(4.77, 100.89)
BUKIT BETONG AT PAHANG	0551421RF	(4.23, 101.94)
TAMAN BUKIT RAWANG AT SELANGOR	0210461RF	(3.33, 101.58)
KAKI BUKIT AT PERLIS	0010061RF	(6.64, 100.21)
BLAU AT KELANTAN	0730501RF	(4.76, 101.75)

Inverse distance weighted (IDW) imputation method is used in this study to handle the missing values. It considers the values of neighbouring stations weighted by their proximity. IDW is chosen for computational efficiency and minimal covariate needs on multidecadal daily data (Lu & Wong 2008) for adaptive IDW. It is also to ensure a balanced spatial coverage while maintaining manageable computational complexity. The equation for the weighting factor in the IDW method is given by,

$$w_i = \frac{d_{i,s}^{-p}}{\sum_{j\neq s}^{N} d_{j,s}^{-p}} \tag{1}$$

where w_i is the weighting factor for station i, $d_{i,s}$ is the Euclidean distance between target station s and neighboring station i, p is power parameter that influences the weight; higher values of p give more emphasis to closer stations and $\sum_{j\neq s}^{N} d_{j,s}^{-p}$ is the sum of the inverse distances raised to the power of $-p$ for all neighboring stations $j \neq s$.

2.2. *Generalized extreme value distribution*

In this study, the BM approach is used to extract the annual maximum rainfall that are considered as extreme values. Using BM GEV marginals and MSP spatial models (Huser *et al.* 2025), this study quantifies spatial dependence in annual maxima from 11 stations (19902023) selected for data completeness and geographic coverage (coastal/inland). BM aligns with MSP theory and avoids threshold selection subjectivity (Coles *et al.* 2001). The sequence of BM is then modelled using the GEV distribution (Fisher & Tippett 1928; Gilli & Këllezi 2006). Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables representing the monthly rainfall data. The data is divided into k non-overlapping blocks of size m, such that $n = k \times m$. For each block, the extreme value is determined as follows,

$$M_i = \max\{X_{(i-1)m+1}, X_{(i-1)m+2}, \ldots, X_{im}\} \tag{2}$$

where M_i represents the maximum value in the i-th block.

The GEV distribution has a cumulative distribution function (CDF), as in the following

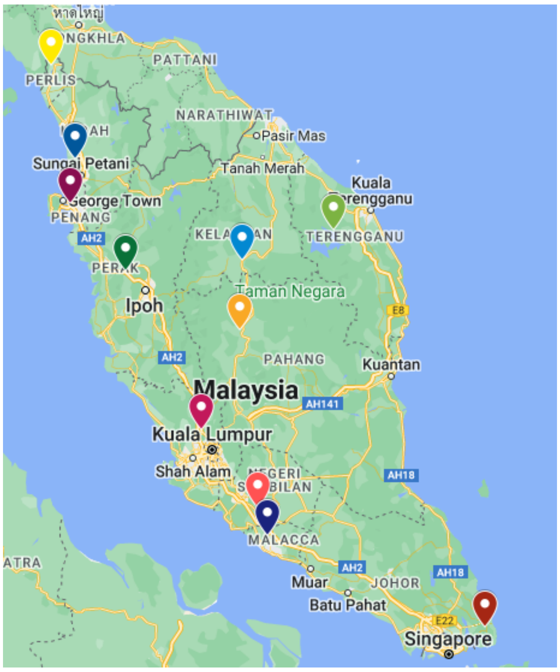

Figure 1: Location of rainfall stations across Peninsular Malaysia

equation,

$$f(x;\mu,\sigma,\xi) = \begin{cases} \exp\left\{-\left(1+\xi(\dfrac{x-\mu}{\sigma})\right)^{-\frac{1}{\xi}}\right\}, & \text{for } -\infty < x < \infty, \xi \neq 0, 1+\xi(\dfrac{x-\mu}{\sigma}) > 0, \\ \exp\left\{-\exp\left(-\dfrac{x-\mu}{\sigma}\right)\right\}, & \text{for } -\infty < x < \infty, \xi = 0, \end{cases} \tag{3}$$

where x is the extreme value obtained from BM, μ is the location parameter with $-\infty < \mu < \infty$, $\sigma > 0$ is the scale parameter, and ξ is the shape parameter.

$$G(x;\mu,\sigma,\xi) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \tag{4}$$

for $1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0$ where μ is the location parameter, σ is the scale parameter, and ξ is the shape parameter.

2.3. *Maximum likelihood estimation*

The parameters of the GEV distribution (μ, σ, and ξ) can be estimated using the maximum likelihood estimation (MLE) method. Given a sample of BM $\{M_1, M_2, \ldots, M_n\}$, the likelihood function $L(\mu,\sigma,\xi)$ is defined as,

$$L(\mu,\sigma,\xi) = \prod_{i=1}^{n} g(M_i;\mu,\sigma,\xi) \tag{5}$$

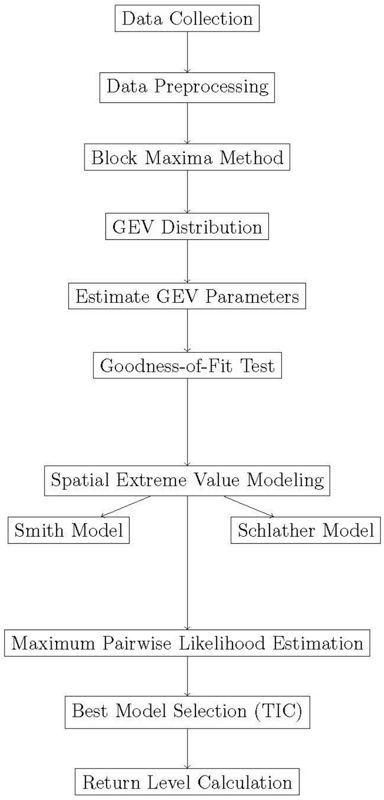

Figure 2: Flow diagram

where $g(M_i; \mu, \sigma, \xi)$ is the probability density function (PDF) of the GEV distribution,

$$g(x; \mu, \sigma, \xi) = \frac{1}{\sigma}\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-(1/\xi+1)} \exp\left(-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right) \tag{6}$$

for $1 + \xi\left(\frac{x-\mu}{\sigma}\right) > 0$.

The log-likelihood function is given by,

$$\ell(\mu, \sigma, \xi) = \log L(\mu, \sigma, \xi) = \sum_{i=1}^{n} \log g(M_i; \mu, \sigma, \xi) \tag{7}$$

Substituting the PDF into the log-likelihood function,

$$\ell(\mu, \sigma, \xi) = -n \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{n} \log\left(1 + \xi(\frac{M_i - \mu}{\sigma})\right) - \sum_{i=1}^{n}\left(1 + \xi(\frac{M_i - \mu}{\sigma})\right)^{-1/\xi} \tag{8}$$

The MLE estimates $\hat{\mu}$, $\hat{\sigma}$, and $\hat{\xi}$ are obtained by solving the following equations,

$$\frac{\partial \ell(\mu, \sigma, \xi)}{\partial \mu} = 0, \quad \frac{\partial \ell(\mu, \sigma, \xi)}{\partial \sigma} = 0, \quad \frac{\partial \ell(\mu, \sigma, \xi)}{\partial \xi} = 0 \tag{9}$$

These likelihood equations are non-linear and do not have closed-form solutions, typically require numerical method such as the BroydenFletcherGoldfarbShanno (BFGS) algorithm to obtain the MLE estimates (Coles *et al.* 2001).

2.4. *Goodness-of-fit test*

The Kolmogorov-Smirnov (KS) test is commonly used to assess the goodness-of-fit of a sample to a specified distribution, such as the GEV distribution. The hypotheses are, H_0: Data follow the GEV distribution. H_a: Data do not follow the GEV distribution. Below are the steps for performing both tests.

Step 1: Calculate the theoretical CDF values

The CDF of the GEV distribution for the sample data is calculated. The CDF of the GEV distribution is given by,

$$G(x;\hat{\mu},\hat{\sigma},\hat{\xi}) = \exp\left\{-\left[1+\hat{\xi}\left(\frac{x-\hat{\mu}}{\hat{\sigma}}\right)\right]^{-1/\hat{\xi}}\right\} \tag{10}$$

for $1+\hat{\xi}\left(\frac{x-\hat{\mu}}{\hat{\sigma}}\right) > 0$, where $\hat{\mu}$ is the location parameter, $\hat{\sigma}$ is the scale parameter, and $\hat{\xi}$ is the shape parameter.

Step 2: Perform the KS Test

The KS test compares the empirical cumulative distribution function (ECDF) of the sample data to the theoretical CDF $G(x;\hat{\mu},\hat{\sigma},\hat{\xi})$. The KS statistic, D is computed as,

$$D = \max\left(\left|F_{\text{empirical}}(x) - G(x;\hat{\mu},\hat{\sigma},\hat{\xi})\right|\right) \tag{11}$$

where $F_{\text{empirical}}(x)$ is the ECDF of the sample data.

2.5. *Spatial extreme value*

SEV models the observed multivariate data at several different locations. Assuming that each component at each location is a GEV distribution, then the transformation is converted into Frchet marginal units. The CDF of the Fréchet distribution is,

$$F(x) = \exp\left(-\frac{1}{x}\right), x > 0 \tag{12}$$

The spatial dependence of extreme rainfall data can be seen by using the extremal coefficients. The extremal coefficient functions of the Smith and Schlather models are defined as follows.

2.5.1. *Smith model*

$$\theta(\mathbf{x}) = 2\Phi\left(\frac{\sqrt{(\mathbf{x}-\mu)^{\top}\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)}}{2}\right) \tag{13}$$

where,

Φ is the standard normal cumulative distribution function,
$\mathbf{x}$ is the vector of the data point,
μ is the mean vector of the distribution,
$\mathbf{\Sigma}$ is the covariance matrix of the distribution, a symmetric positive semi-definite matrix that describes the spread and correlation structure of the data,
$\mathbf{\Sigma}^{-1}$ is the inverse of the covariance matrix, often referred to as the precision matrix. It provides information about the conditional independence relationships between variables in the distribution, and $(\mathbf{x}-\mu)^\top$ denotes the transpose of the vector $(\mathbf{x}-\mu)$.

2.5.2. *Schlather model*

In the Schlather model, we introduce the non-stationary correlation function. $\lambda(t) = \exp(\lambda_0 + \lambda_1 t)$ is specified in Schlather to reflect evolving spatial correlation; this follows nonstationary dependence ideas (Huser & Genton 2016). Parameter estimation uses pairwise composite likelihood Padoan *et al.* (2010).

$$\theta(h) = 1 + \sqrt{\frac{1-\rho(h)}{2}} \tag{14}$$

Eq. (14) shows the correlation function $\rho(h)$, where $h = s_1 - s_2$ is the spatial lag, describes the correlation between the extreme values at locations s_1 and s_2. Common choices for $\rho(h)$ include,

Whittle-Matern

$$\rho(h) = \eta + (1-\eta)\cdot\frac{\Gamma(\nu)}{2^{1-\nu}}(\lambda h)^\nu K_\nu(\lambda h)$$

Non-stationary Whittle-Matern,

$$\rho(h,t) = \eta + (1-\eta)\cdot\frac{\Gamma(\nu)}{2^{1-\nu}}\big(\lambda(t)h\big)^\nu K_\nu\big(\lambda(t)h\big)$$

where $\lambda(t) = \exp(\lambda_0 + \lambda_1 t)$.

Powered Exponential

$$\rho(h) = \eta + (1-\eta)\cdot\exp\left(-\left(\frac{h}{\lambda}\right)^v\right)$$

Non-stationary Powered Exponential

$$\rho(h,t) = \eta + (1-\eta)\cdot\exp\left(-\left(\frac{h}{\lambda(t)}\right)^v\right)$$

where $\lambda(t) = \exp(\lambda_0 + \lambda_1 t)$.

Bessel

$$\rho(h) = \eta + (1-\eta)\left(1+\left(\frac{h}{\lambda}\right)^\nu\right)^{-2}$$

Non-stationary Bessel

$$\rho(h,t) = \eta + (1-\eta)\cdot\left(1+\left(\frac{h}{\lambda(t)}\right)^{v}\right)^{-2}$$

where $\lambda(t) = \exp(\lambda_0 + \lambda_1 t)$.

Cauchy

$$\rho(h) = \eta + \frac{1-\eta}{1+\left(\frac{h}{\lambda}\right)^{\nu}}$$

Non-stationary Cauchy

$$\rho(h,t) = \eta + \frac{1-\eta}{1+\left(\frac{h}{\lambda(t)}\right)^{v}}$$

where $\lambda(t) = \exp(\lambda_0 + \lambda_1 t)$.

There are five parameters need to be estimated known as nugget parameter, η which accounts for noise or small-scale variability, range parameter, λ which determines the spatial scale of correlation, smoothness parameter, ν which controls the shape of the correlation function, and distance between locations. The non-stationary equation allows λ to vary with time. If the extremal coefficients value $\theta(h) = 1$, there are full dependencies between the two regions. If the value is in the interval of $1 < \theta(h) < 2$, there is a dependent relationship between the two regions and if the value of $\theta(h) = 2$, there is an independent relationship.

2.6. *Max-stable process*

MSP is employed to transform all data into the Fréchet distribution. The MSP refers to the transformation of data X as extreme data into the Fréchet distribution using its distribution function in Eq. (15) below,

$$Z(s) = \exp\left\{-\left(1+\xi(\frac{x-\mu}{\sigma})\right)^{-\frac{1}{\xi}}\right\} \tag{15}$$

$$Z(s) = \max_{i=1}^{\infty}\{\xi_i\phi(s-U_i)\} \tag{16}$$

The Smith model has a bivariate cumulative distribution function as follows,

$$F(z_i,z_j) = \exp\left(-\frac{1}{z_i}\Phi\left(\frac{a(h)}{2}+\frac{1}{a(h)}\log\left(\frac{z_j}{z_i}\right)\right)-\frac{1}{z_j}\Phi\left(\frac{a(h)}{2}+\frac{1}{a(h)}\log\left(\frac{z_i}{z_j}\right)\right)\right) \tag{17}$$

$$Z(s) = \max_{i=1}^{\infty} U_i Y_i(s) \tag{18}$$

The bivariate CDF for the MSP under the Schlather model is given by,

$$F(z_i, z_j) = \exp\left(-\left(\frac{1}{z_i} + \frac{1}{z_j}\right)\left(1 + \sqrt{1 - 2(\rho(h)+1)\frac{z_i z_j}{(z_i + z_j)^2}}\right)\right) \tag{19}$$

where $\rho(h)$ is the correlation matrix. The next step is to calculate the MSP parameters using the MPLE.

2.7. *Maximum pairwise likelihood estimation*

MPLE is a parameter estimation method using a pairwise density function of two variables. In the context of the spatial model, the GEV distribution is defined as follows,

$$GEV(\mu(s), \sigma(s), \xi(s)) \tag{20}$$

where GEV distribution parameters follow the trend surface model with the following equation,

$$\mu(s) = \beta_{0,\mu} + \beta_1 lon(s) + \beta_{2,\mu}\, lat(s) \tag{21}$$
$$\sigma(s) = \beta_{0,\sigma} + \beta_{1,\sigma}\, lon(s) + \beta_{2,\sigma}\, lat(s) \tag{22}$$
$$\xi(s) = \beta_{0,\xi} \tag{23}$$

The MPLE is used to estimate the parameters β_μ, β_σ, and β_ξ. The pairwise log-likelihood equation for m location is defined as follows,

$$\ell_p(\beta, z) = \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \log\left[f(z_i, z_j; \beta)\right]$$

with $f(z_i, z_j, \beta)$ is the PDF of the bivariate MSP model. Thus, the MPLE for β is,

$$\hat{\beta} = argmaxl_p(\beta; z). \tag{24}$$

2.8. *Model selection criteria*

The Takuechi Information Criterion (TIC) is used for model selection and is expressed as,

$$\text{TIC} = -2\ell_p(\hat{\beta}) + 2\,\text{p}\left(1 + \frac{p}{n}\right) \tag{25}$$

where,

$-2\ell_p(\hat{\beta})$ is the penalty term reflecting the model's goodness of fit, and
$2\,\text{p}\left(1 + \frac{p}{n}\right)$ is the complexity penalty that adjusts for the number of parameters and sample size.

Integration of these methods involves assessing the adequacy of the GEV distribution fit and selecting the model with the lowest TIC value as the best fit model, balancing goodness of fit and model complexity. This integrated approach ensures robust model selection in statistical modelling, considering both goodness-of-fit and model complexity.

2.9. *Return level*

The maximum threshold achieved within a certain return period, T is referred to as the return level. In extreme spatial cases, particularly for MSP, the return level with the T-year return time is determined using a point-wise quantile as below,

$$Z_p(s) = \hat{\mu}(s) - \frac{\hat{\sigma}(s)}{\hat{\xi}(s)} \left(1 - \left[-\ln\left(1 - \frac{1}{T}\right)\right]^{\hat{\xi}(s)}\right) \tag{26}$$

where,

$\hat{\mu}(s)$ determines the central tendency or baseline level of the distribution,
$\hat{\sigma}(s)$ represents the spread or variability of extreme values,
$\hat{\xi}(s)$ controls the tail behaviour,
$\hat{\xi} > 0$ represents heavy-tailed (frequent extreme events),
$\hat{\xi} = 0$ represents exponential tail (Gumbel distribution, moderate extremes), and
$\hat{\xi} < 0$ represents short-tailed (bounded extremes).

The root mean square error (RMSE) can be used to evaluate the Smith and Schlather models' performance in terms of return levels. The general RMSE formula is given below,

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{27}$$

m is the number of locations,
y_i is the observed value, and
$\hat{y}_i$ is the predicted value.

This formula calculates the square root of the average of the squared differences between observed and predicted values of annual extreme rainfall. A lower RMSE values indicate better agreement between the observed and predicted values, while a higher RMSE suggests poorer model performance. The flow diagram of methodology can be referred in Figure 2.

3. Results and Discussion

The boxplot of annual extreme rainfall for 11 selected stations are illustrated in Figure 3. The y-axis represents the daily rainfall in mm/day, while the x-axis represents Stations ID. This boxplot visualizes the distribution of extreme rainfall events where station 0210461RF shows more variation in extreme rainfall values. Stations 1910051RF, 0500551RF, 0030021RF and 0670281RF receive higher amount of rainfall which exceeding 200 mm/day where station 0670281RF depict the highest extreme value. Stations with shorter whiskers and fewer outliers indicate less variability and fewer extreme rainfall events, whereas stations with longer whiskers and more outliers suggest greater variability and a higher frequency of extreme rainfall events. The varying spread of the boxplots which reflects differences in rainfall variability across stations is more likely due to climate influences, topography or weather patterns. The descriptive statistics are shown in Table 2. Positive values of skewness and kurtosis indicate a heavy-tailed distribution towards the right.

3.1. *GEV fitting*

Table 3 shows the MLE parameters of the GEV distribution. The GEV parameter estimates and their 95% confidence intervals (CI) reveal notable spatial differences in extreme rainfall behaviour across stations. For example, station 0670281RF has a high location parameter estimate of 187.76 with a wide confidence interval (169.35, 206.15), indicating significantly higher

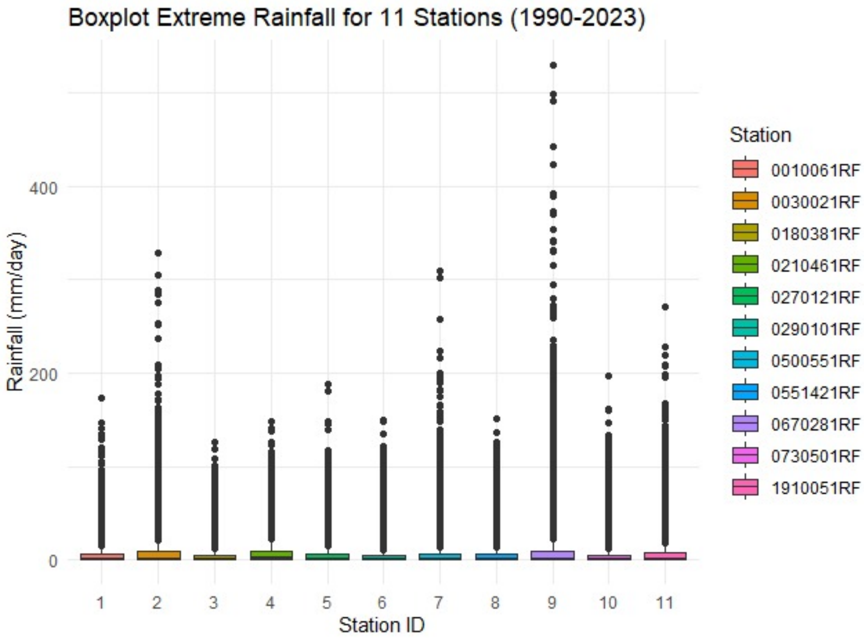

Figure 3: Boxplot of annual extreme rainfall

and more uncertain extreme rainfall levels. Its large scale parameter 100.09 with CI of (90.27, 109.89) reflects substantial variability in extremes. In contrast, station 0270121RF has a lower location 79.99 with CI of (72.14, 87.82) and scale 25.11 with CI of (22.65, 27.57), as well as a negative shape parameter of 0.0907 with CI of (0.09, 0.08), suggesting a bounded distribution with less likelihood of extreme rainfall. The relatively narrow CI across most stations imply stable and precise parameter estimates, supporting reliable inferences about the nature and severity of extreme rainfall events while a relatively wide CI indicates higher uncertainty of the estimates.

3.2. *Diagnostic checking*

Figures 4 and 5 shows the diagnostic plots for the selected stations. The probability plot (top-left) compares the empirical quantiles of the data with the theoretical quantiles of the fitted GEV distribution. A good fit is indicated when the points closely align with the reference (red) 1:1 line, showing minimal deviation. For example, at station 0180381RF, the points are moderately aligned with the red line, suggesting the GEV model fits the data well, with only minor deviations at the extremes which possibly due to underestimation or overestimation of very high values. The Q-Q plot (Quantile-Quantile plot) also compares empirical and theoretical quantiles. A good fit is indicated when points closely follow the blue 1:1 line. For station 0180381RF, the points align closely with the line, with only minimal deviations at the tails, indicating a good fit of the GEV distribution to the extreme rainfall data. The return level plot estimates return levels for different return periods (e.g., 10, 20, 50, and 100 years). A good fit is observed when the plotted points follow the trend of the fitted curve. For station 0180381RF, the purple curve aligns well with the estimated return levels, indicating accurate predictions and stable parameter estimates. Lastly, the density plot compares the histogram of observed data with the theoretical GEV probability density function (PDF). A good fit is indicated when the dark green density curve closely follows the histogram. At station 0180381RF, the curve aligns well with the histogram, especially around the central tendency, showing that the GEV distribution accurately represents the overall distribution of extreme rainfall at this station. Overall,

Table 2: Descriptive statistics

Stations ID	Min	Mean	Max	Skewness	Kurtosis
0500551RF	0	6.6013	308.9	5.74	54.56
0270121RF	0	5.9256	188.7	4.03	24.37
1910051RF	0	7.5413	270.5	4.44	30.89
0670281RF	0	9.7095	530.0	7.62	90.87
0030021RF	0	8.8833	329.0	4.91	38.32
0290101RF	0	5.1819	149.4	4.03	22.39
0180381RF	0	5.0379	125.3	3.59	17.76
0551421RF	0	5.7095	151.6	3.74	19.12
0210461RF	0	6.5322	147.5	3.34	15.43
0010061RF	0	5.1721	173.1	4.63	34.39
0730501RF	0	5.7397	196.5	3.95	23.74

most stations exhibit similar patterns across all diagnostic plots, confirming that the GEV distribution provides a good fit for the extreme rainfall data. However, if any plots suggest misfit, model adjustments may be necessary.

3.3. *Kolmogorov-Smirnov (KS) test*

KS test is further tested on the extreme rainfall data set. As shown in Table 4, the extreme data at all stations follow the theoretical GEV distribution at 95% significance level.

3.4. *Spatial dependencies*

Identification of spatial dependencies is the first step in SEV analysis using the MSP approach, where extremal coefficients are used to assess the presence and strength of dependence in the dataset. Table 5 presents the extremal coefficients for both the Smith and Schlather models. These coefficients, which range from 1 to 2, quantify the strength of dependence between extreme events at different locations whereby values closer to 1 indicate strong dependence, while values near 2 suggest independence. For the Smith model, the extremal coefficient typically increases gradually with distance, reflecting a strong spatial dependence at shorter distances. In contrast, the Schlather model allows for a broader range of dependence structures depending on the chosen correlation function. Although both models yield extremal coefficients within the same theoretical range (1 to 2), the rate of increase in the Schlather model can vary more significantly due to its flexible correlation structure. Based on the results in Table 5, the Schlather model tends to show higher extremal coefficient values at the same distances compared to the Smith model, indicating weaker dependence. This suggests that the Smith model maintains spatial dependence over longer distances, owing to its Gaussian structure, while the Schlather model weakens dependence more quickly and reaches near-independence (coefficient ≈ 2) faster. In summary, both models indicate strong dependence at short distances. However, for longer distances, the Smith model retains some level of dependence, whereas the Schlather model tends to approach independence more rapidly.

The pair of stations showing strong spatial dependence at short distance is 0270121RF and 0290101RF, with a separation of 0.2610 units. The extremal coefficients for this pair are 0.01720 (Smith model) and 1.7912 (Schlather model). The Smith model gives a very low extremal coefficient, indicating very strong spatial dependence at short distances, while the Schlather model suggests moderate dependence, though weaker than that indicated by the Smith model. This difference arises because the Smith model assumes a Gaussian dependence structure that decays smoothly with distance, whereas the Schlather models dependence is governed by its correlation function, which can decay more rapidly. For moderate spatial dependence at

Table 3: MLE parameters of GEV distribution

Stations ID	Parameters	Standard Error	95% CI
0500551RF	$\mu = 116.8516$	5.8425	(105.40,128.30)
	$\sigma = 42.2674$	2.1133	(38.12,46.40)
	$\xi = 0.1067$	0.0053	(0.09,0.11)
0270121RF	$\mu = 79.9864$	3.9993	(72.14,87.82)
	$\sigma = 25.1136$	1.2556	(22.65,27.57)
	$\xi = -0.0907$	0.0045	(-0.09,-0.08)
1910051RF	$\mu = 112.0851$	5.6042	(101.10,123.06)
	$\sigma = 33.2854$	1.6642	(30.02,36.54)
	$\xi = 0.0619$	0.0030	(0.05,0.06)
0670281RF	$\mu = 187.7580$	9.3879	(169.35,206.15)
	$\sigma = 100.0883$	5.0044	(90.27,109.89)
	$\xi = -0.0203$	0.0010	(-0.02,0.01)
0030021RF	$\mu = 149.5911$	7.4795	(134.93,164.25)
	$\sigma = 51.6681$	2.5834	(46.60,56.73)
	$\xi = 0.0071$	0.0003	(0.006,0.007)
0290101RF	$\mu = 82.7737$	4.1386	(74.66,90.88)
	$\sigma = 17.3606$	0.8680	(15.65,19.06)
	$\xi = 0.0559$	0.0027	(0.05,0.06)
0180381RF	$\mu = 72.7675$	3.6383	(65.63,79.89)
	$\sigma = 18.4775$	0.9238	(16.66,20.28)
	$\xi = -0.1778$	0.0088	(-0.19,-0.16)
0551421RF	$\mu = 82.8219$	4.1410	(74.70,90.93)
	$\sigma = 18.4595$	0.9229	(16.65,20.26)
	$\xi = -0.0542$	0.0027	(-0.05,-0.04)
0210461RF	$\mu = 84.1263$	4.2063	(75.88,92.37)
	$\sigma = 27.5507$	1.3775	(24.85,30.25)
	$\xi = -0.2732$	0.0136	(-0.30,-0.24)
0010061RF	$\mu = 68.4576$	0.0136	(61.74,75.16)
	$\sigma = 26.4189$	1.3209	(23.82,29.00)
	$\xi = 0.0373$	0.0018	(0.03.0.04)
0730501RF	$\mu = 81.4340$	4.0717	(73.45,89.41)
	$\sigma = 25.3010$	1.2650	(22.82,27.78)
	$\xi = 0.0663$	0.0033	(0.05,0.07)

medium distance, the pair 0500551RF and 0180381RF is considered, with a distance of 3.7140 units. The extremal coefficients are 0.00848 (Smith) and 1.9959 (Schlather). The Smith model still shows a low value, indicating that some dependence remains even at this medium distance, even though weaker than at shorter distances. In contrast, the Schlather model gives a value very close to 2, suggesting near independence. This reflects the Schlather model's tendency to allow dependence to decay more quickly with distance. Lastly, for weak spatial dependence at long distance, the pair 0500551RF and 0010061RF, with a distance of 4.7200 units, shows an extremal coefficient of 0.00751 for the Smith model and 1.9989 for the Schlather model. The Smith model again indicates that some spatial dependence persists even at long distances due to its Gaussian structure. Meanwhile, the Schlather model suggests that the two stations are almost completely independent, as its extremal coefficient is nearly 2. This again highlights the key difference: the Smith model maintains dependence over longer distances, while the Schlather model's dependence drops off more rapidly. Next, MPLE is used to estimate the spatial GEV parameters. Each parameter estimate of $\hat{\mu}$, $\hat{\sigma}$, and $\hat{\xi}$, is calculated using a trend surface model with a combination of latitude and longitude spatial components. Table 6 shows the parame-

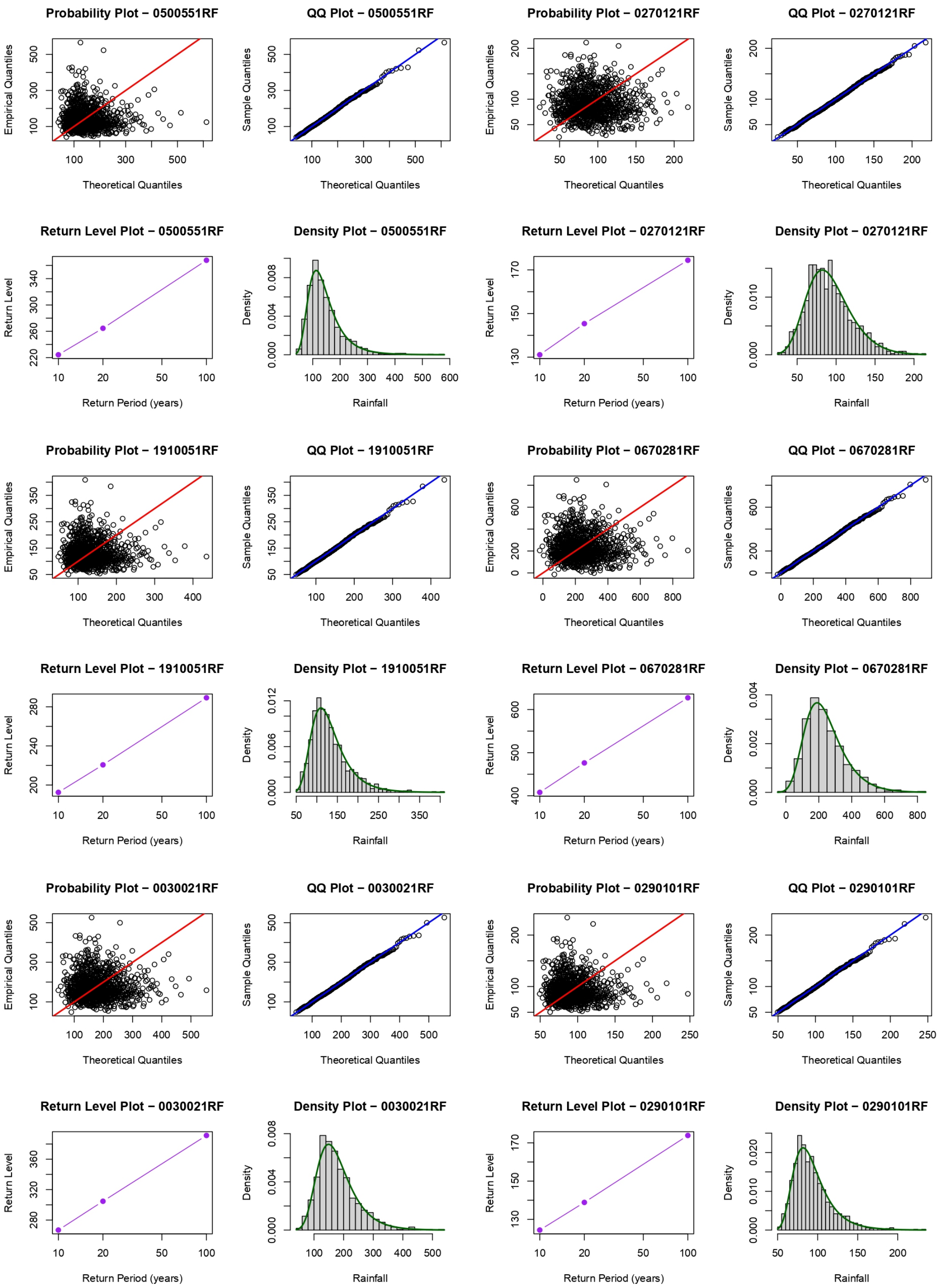

Figure 4: Diagnostic plots for Station 0500551RF, 0270121RF, 1910051RF, 0670281RF, 0030021RF and 0290101RF

ter estimation of the spatial GEV model from the nine combinations of trend surface models formed.

Based on the nine models, the first model is the most appropriate since it has the smallest

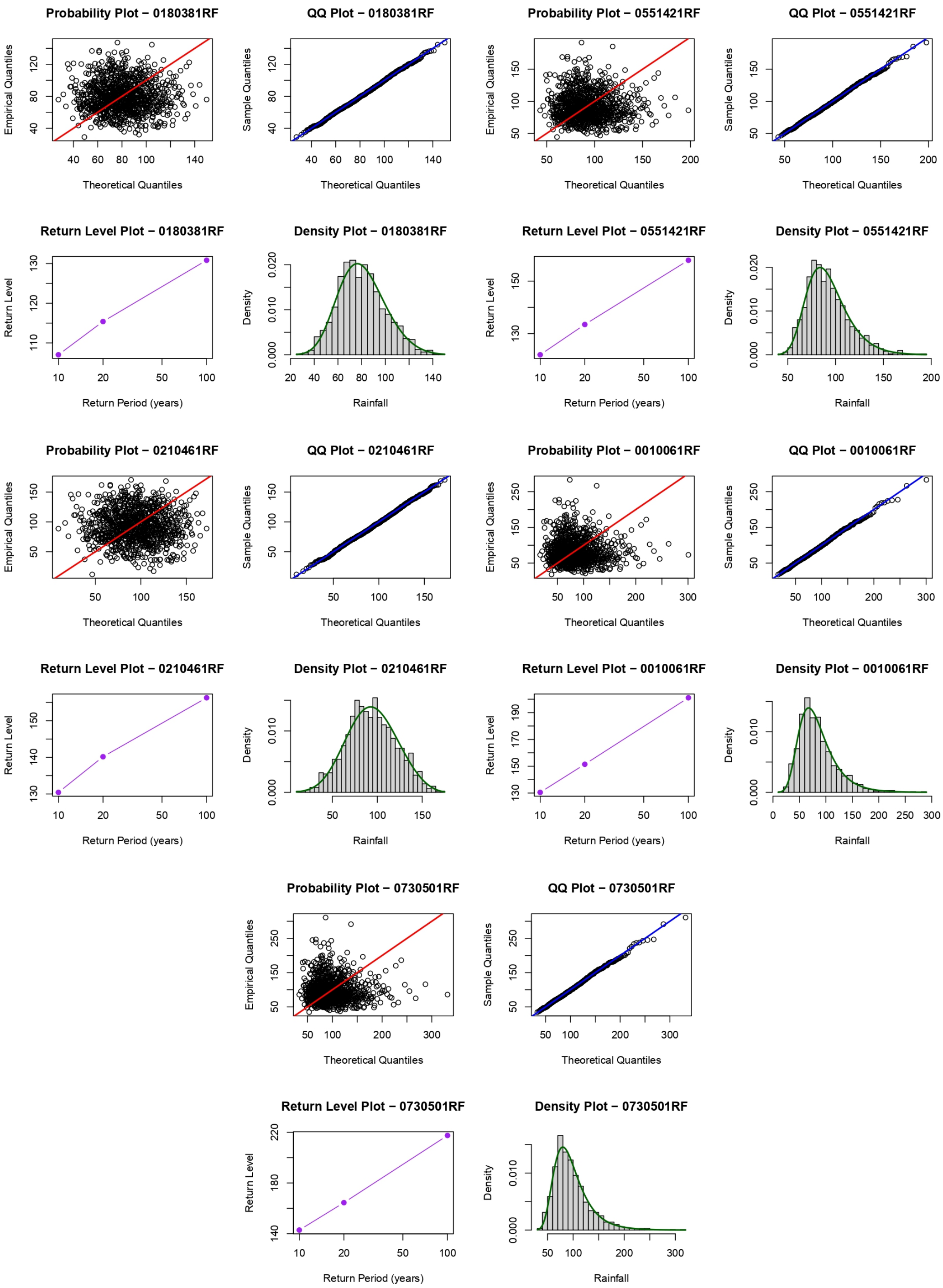

Figure 5: Diagnostic plots for Station 0180381RF, 0551421RF, 0210461RF, 0010061RF and 0730501RF

TIC value, which is 104.3312. Thus, the suitable trend model is as below,

$$\hat{\mu}(s) = 0.3383 - 0.6910 \cdot \text{lon}(s) + 0.6457 \cdot \text{lat}(s)$$

Table 4: KS test

Station ID	KS-Statistic	p-value	Decision
0500551RF	0.1021	0.8350	Fail to reject H_0.
0270121RF	0.1369	0.5468	Fail to reject H_0.
1910051RF	0.0690	0.9969	Fail to reject H_0.
0670281RF	0.1040	0.8555	Fail to reject H_0.
0030021RF	0.0814	0.9640	Fail to reject H_0.
0290101RF	0.0764	0.9887	Fail to reject H_0.
0180381RF	0.0819	0.9764	Fail to reject H_0.
0551421RF	0.0745	0.9915	Fail to reject H_0.
0210461RF	0.1053	0.8067	Fail to reject H_0.
0010061RF	0.0879	0.9551	Fail to reject H_0.
0730501RF	0.1180	0.7304	Fail to reject H_0.

$$\hat{\sigma}(s) = 0.1514 - 0.1688 \cdot \text{lon}(s) + 0.7414 \cdot \text{lat}(s)$$

$$\hat{\xi}(s) = 0.3782$$

This model is further used to estimate the spatial GEV parameters. The parameter estimates obtained will be in the form of $\hat{\mu}$, $\hat{\sigma}$, and $\hat{\xi}$ for each station.

3.5. *Smith model*

By using the BFGS method, and the best trend surface model that has been obtained, the Smith model parameters is obtained as in Table 7.

3.6. *Schlather model*

As for the Schlather model, the correlation functions is calculated and can be seen in Table 8. Overall, powered exponential with constant range parameter is the most suitable correlation function for Schlather with smallest TIC value (2007.99358). In addition, it is also interesting to note that non-stationary Bessel and Cauchy (time-covariate range parameter) produce lower TIC when compared to the constant Bessel and Cauchy, respectively. This also means that the time-varying range parameter is suitable for both Bessel and Cauchy. Meanwhile, the constant range parameter is suitable for both Whittle-Matern and powered exponential. In this model, the nugget parameter is set to zero, indicating that there is no measurement error or spatial discontinuity at the origin. Nevertheless, powered exponential correlation function is further used for estimating the parameters of the Schlather model. In this model, the shape parameter $\hat{\xi}(s)$ is set to 0.5 to ensure numerical stability and reduce the model complexity, as estimating spatially varying shape parameters can lead to over-fitting and unreliable results when data are sparse or noisy (Coles *et al.* 2001). Following to that, the parameter estimates of the Schlather model for each station are obtained as in Table 9.

3.7. *Return level of Smith and Schlather models*

Return level of T = 5, 10, 20 and 30 years is estimated. Table 10 shows the return level estimation for both Smith and Schlather models. Based on Table 10, Schlather model consistently gives higher return levels compared to the Smith model. As the return periods increase from 5,10,20 and 30 years, the difference between the two models increases indicating Schlather predicts much more extreme events at longer return periods. Spatial dependence exhibit critical role in predicting the return periods. Smith model with Gaussian structure lead to moderate return levels compared to Schlather model which based on underlying correlation function, making

Table 5: Extremal coefficient

Pairing Two Stations	Distance	Smith	Schlather
0500551RF - 0270121RF	1.5401	0.01131	1.9538
0500551RF - 1910051RF	4.6129	0.01435	1.9985
0500551RF - 0670281RF	3.0280	0.00848	1.9913
0500551RF - 0030021RF	4.7650	0.00789	1.9987
0500551RF - 0290101RF	1.4124	0.01037	1.9466
0500551RF - 0180381RF	3.7140	0.00848	1.9959
0500551RF - 0551421RF	2.6013	0.00870	1.9860
0500551RF - 0210461RF	2.6714	0.01094	1.9875
0500551RF - 0010061RF	4.7200	0.00751	1.9989
0500551RF - 0730501RF	3.3157	0.00933	1.9972
0270121RF - 1910051RF	3.3599	0.01343	1.9940
0270121RF - 0670281RF	2.6310	0.01043	1.9865
0270121RF - 0030021RF	3.5963	0.00758	1.9954
0270121RF - 0290101RF	0.2610	0.01720	1.7912
0270121RF - 0180381RF	2.4816	0.01756	1.9840
0270121RF - 0551421RF	1.6324	0.01120	1.9585
0270121RF - 0210461RF	0.8898	0.02142	1.9023
0270121RF - 0010061RF	4.4574	0.01108	1.9982
0270121RF - 0730501RF	2.1854	0.01211	1.9778
1910051RF - 0670281RF	2.7700	0.00617	1.9883
1910051RF - 0030021RF	0.4084	0.00953	1.8261
1910051RF - 0290101RF	3.6139	0.01518	1.9955
1910051RF - 0180381RF	0.8990	0.00914	1.9033
1910051RF - 0551421RF	2.0504	0.01276	1.9740
1910051RF - 0210461RF	2.4713	0.01153	1.9838
1910051RF - 0010061RF	1.2222	0.00733	1.9336
1910051RF - 0730501RF	1.6249	0.00859	1.9579
0670281RF - 0030021RF	2.4433	0.00965	1.9831

Pairing Two Stations	Distance	Smith	Schlather
0670281RF - 0290101RF	2.8498	0.00862	1.9894
0670281RF - 0180381RF	1.9368	0.01466	1.9704
0670281RF - 0551421RF	1.2522	0.00977	1.9357
0670281RF - 0210461RF	2.1815	0.01829	1.9776
0670281RF - 0010061RF	2.9843	0.01430	1.9908
0670281RF - 0730501RF	1.1059	0.00895	1.9238
0030021RF - 0290101RF	3.8548	0.00991	1.9965
0030021RF - 0180381RF	1.1179	0.00798	1.9251
0030021RF - 0551421RF	2.1670	0.01119	1.9772
0030021RF - 0210461RF	2.7182	0.00934	1.9877
0030021RF - 0010061RF	0.8783	0.00938	1.9010
0030021RF - 0730501RF	1.6723	0.00939	1.9601
0290101RF - 0180381RF	2.7391	0.01123	1.9880
0290101RF - 0551421RF	1.8872	0.01135	1.9689
0290101RF - 0210461RF	1.1426	0.01257	1.9272
0290101RF - 0010061RF	4.7171	0.00756	1.9986
0290101RF - 0730501RF	2.4428	0.00980	1.9833
0180381RF - 0551421RF	1.1778	0.01250	1.9299
0180381RF - 0210461RF	1.6006	0.02875	1.9569
0180381RF - 0010061RF	1.9899	0.01326	1.9722
0180381RF - 0730501RF	0.8625	0.01584	1.8988
0551421RF - 0210461RF	0.9681	0.0126	1.9109
0551421RF - 0010061RF	2.9663	0.0103	1.9907
0551421RF - 0730501RF	0.5640	0.0118	1.8561
0210461RF - 0010061RF	2.8896	0.0097	1.9899
0210461RF - 0730501RF	1.4448	0.0110	1.9486
0010061RF - 0730501RF	2.4975	0.0056	1.9842

Table 6: Trend surface models and TIC values

No.	Trend Surface Models	TIC
1	$\hat{\mu}(s) = 0.3383 - 0.6910\,\mathrm{lon}(s) + 0.6457\,\mathrm{lat}(s)$ $\hat{\sigma}(s) = 0.1514 - 0.1688\,\mathrm{lon}(s) + 0.7414\,\mathrm{lat}(s)$ $\hat{\xi}(s) = 0.3782$	104.3312
2	$\hat{\mu}(s) = 0.9455 - 0.6910\,\mathrm{lon}(s)$ $\hat{\sigma}(s) = 0.1514 - 0.1688\,\mathrm{lon}(s) + 0.7414\,\mathrm{lat}(s)$ $\hat{\xi}(s) = 0.3782$	180.0031
3	$\hat{\mu}(s) = 0.3687 - 0.6910\,\mathrm{lon}(s) + 0.6457\,\mathrm{lat}(s)$ $\hat{\sigma}(s) = 0.1514 - 0.1688\,\mathrm{lon}(s)$ $\hat{\xi}(s) = 0.3782$	178.8103
4	$\hat{\mu}(s) = 0.3383 + 0.6457\,\mathrm{lat}(s)$ $\hat{\sigma}(s) = 0.1514 - 0.1688\,\mathrm{lon}(s) + 0.7414\,\mathrm{lat}(s)$ $\hat{\xi}(s) = 0.3782$	146.6327
5	$\hat{\mu}(s) = 0.9455 + 0.1105\,\mathrm{lon}(s) + 0.6457\,\mathrm{lat}(s)$ $\hat{\sigma}(s) = 0.1514 + 0.7414\,\mathrm{lat}(s)$ $\hat{\xi}(s) = 0.3782$	123.4973
6	$\hat{\mu}(s) = 0.3687 - 0.1105\,\mathrm{lon}(s)$ $\hat{\sigma}(s) = 0.1514 - 0.1688\,\mathrm{lon}(s)$ $\hat{\xi}(s) = 0.3782$	122.5827
7	$\hat{\mu}(s) = 0.3383 + 0.6457\,\mathrm{lat}(s)$ $\hat{\sigma}(s) = 0.1514 + 0.7414\,\mathrm{lat}(s)$ $\hat{\xi}(s) = 0.3782$	104.4068
8	$\hat{\mu}(s) = 0.9455 + 0.3614\,\mathrm{lon}(s)$ $\hat{\sigma}(s) = 0.1514 + 0.7414\,\mathrm{lat}(s)$ $\hat{\xi}(s) = 0.3782$	119.4200
9	$\hat{\mu}(s) = 0.3687 + 0.6457\,\mathrm{lat}(s)$ $\hat{\sigma}(s) = 0.1514 - 0.1688\,\mathrm{lon}(s)$ $\hat{\xi}(s) = 0.3782$	195.7672

it more sensitive to extreme rainfall resulting in significantly higher return levels. Schlather model need to be used when expecting highly dependent extreme events as it predicts much higher return levels. Meanwhile, Smith model provides a more conservative estimate which will be safer but could underestimate rare events. Schlather model will be more useful when planning for worst-case flood scenarios compared to Smith model. As referred to Table 10, the station that exhibit highest return level values is station 0500551RF. The Smith model predicts a moderate increase in return levels over time , 235.39 mm,346.94 mm, 491.92 mm and 596.12 mm for year 5,10,20 and 30 respectively. The Schlather model predicts extreme values with the 30-year return level being over three times of Smith model which the values is 1870.61 mm. This indicates that extreme rainfall events could be much more severe than Smith model under Schlather model assumption of dependence in extremes. Next, for moderate return level values, the station 0270121RF shows that Smith model estimates are much lower than Schlather's model for each period which are 231.57 mm, 341.41 mm, 484.18 mm and 586.79 mm for period of 5,10,20 and 30 years respectively. The return level at a 30-year for Schlather model with values of 1843.82 mm predicts nearly 3.1 times estimated rainfall by Smith model. This indicates that for extreme events, Schlather model have better accounts for tail heaviness rather than Smith model which might underestimate rare extremes. Lastly, for the lowest return level values, for instance, station 0010061RF, Smith model predicts a steady increase, while Schlather model shows significantly higher estimates for the return periods. Since this station has the lowest Smith model return levels, Smith model underestimating extremes more for this station compare to others.

Table 7: Smith model parameter estimations

Station ID	Parameters	Standard Error	95% C.I.
0500551RF	$\mu = 65.6916$	3.2846	(59.25, 72.12)
	$\sigma = 76.5831$	3.8292	(69.07, 84.08)
	$\xi = 0.3782$	0.0189	(0.34, 0.42)
0270121RF	$\mu = 64.4652$	3.2233	(58.14, 70.78)
	$\sigma = 75.4124$	3.7706	(68.02, 82.80)
	$\xi = 0.3782$	0.0189	(0.34, 0.42)
1910051RF	$\mu = 61.3354$	3.0668	(55.32, 67.34)
	$\sigma = 73.5767$	3.6788	(66.36, 80.78)
	$\xi = 0.3782$	0.0189	(0.34, 0.42)
0670281RF	$\mu = 63.1554$	3.1578	(56.96, 69.34)
	$\sigma = 75.4944$	3.7747	(68.09, 82.89)
	$\xi = 0.3782$	0.0189	(0.34, 0.42)
0030021RF	$\mu = 61.1868$	3.0593	(55.19, 67.18)
	$\sigma = 73.6385$	3.6819	(66.42, 80.85)
	$\xi = 0.3782$	0.0189	(0.34, 0.42)
0290101RF	$\mu = 64.6912$	3.2346	(58.35, 71.03)
	$\sigma = 75.5181$	3.7759	(68.11, 82.91)
	$\xi = 0.3782$	0.0189	(0.34, 0.42)
0180381RF	$\mu = 62.1863$	3.1093	(56.09, 68.28)
	$\sigma = 74.1485$	3.7074	(66.88, 81.41)
	$\xi = 0.3782$	0.0189	(0.34, 0.42)
0551421RF	$\mu = 63.2359$	3.1618	(57.03, 69.43)
	$\sigma = 75.0153$	3.7508	(67.66, 82.36)
	$\xi = 0.3782$	0.0189	(0.34, 0.42)
0210461RF	$\mu = 63.6306$	3.1815	(57.39, 69.86)
	$\sigma = 74.9055$	3.7453	(67.56, 82.24)
	$\xi = 0.3782$	0.0189	(0.34, 0.42)
0010061RF	$\mu = 60.4542$	3.0227	(54.52, 66.37)
	$\sigma = 73.3274$	3.6664	(66.14, 80.51)
	$\xi = 0.3782$	0.0189	(0.34, 0.42)
0730501RF	$\mu = 62.7490$	3.1375	(56.59, 68.89)
	$\sigma = 74.7894$	3.7395	(67.46, 82.11)
	$\xi = 0.3782$	0.0189	(0.34, 0.42)

Table 8: Schlather model dependency parameters

Correlation Function	**Nugget**	**Range**	**Smooth**	**TIC**
Whittle-Matrn	0	300.0007	0.5224	2007.9936
Non-stationary Whittle-Matrn	0	$\lambda_0 = 10$, $\lambda_1 = 1$	2	2010.4178
Powered Exponential	0	1	0.8458	2007.9936
Non-stationary Powered Exponential	0	$\lambda_0 = -10$, $\lambda_1 = -1$	0.2079	2043.4178
Bessel	0	1	0.5978	2086.6031
Non-stationary Bessel	0	$\lambda_0 = -10$, $\lambda_1 = -1$	0.3628	2043.4178
Cauchy	0	1	2	2184.2544
Non-stationary Cauchy	0	$\lambda_0 = -10$, $\lambda_1 = -1$	0.7121	2043.4178

The next step is to determine the best model using the RMSE value. However, in order to be able to compare the return level in prediction results with the rainfall testing data in mm/month, it is necessary to first transform the return level prediction data in Table 10 to initial data units

Table 9: Schlather model parameter estimations

Station ID	Parameters	Standard Error	95% C.I.
0500551RF	$\mu = 0.1444$	0.0072	(0.130, 0.158)
	$\sigma = 208.8866$	10.4443	(188.41, 229.35)
	$\xi = 0.5$	0.025	(0.451, 0.549)
0270121RF	$\mu = 0.1434$	0.0072	(0.129, 0.157)
	$\sigma = 205.8953$	10.2947	(185.71, 226.07)
	$\xi = 0.5$	0.025	(0.451, 0.549)
1910051RF	$\mu = 0.1366$	0.0068	(0.123, 0.149)
	$\sigma = 202.2170$	10.1108	(182.39, 222.03)
	$\xi = 0.5$	0.025	(0.451, 0.549)
0670281RF	$\mu = 0.1372$	0.0069	(0.123, 0.150)
	$\sigma = 207.2757$	10.3637	(186.96, 227.58)
	$\xi = 0.5$	0.025	(0.451, 0.549)
0030021RF	$\mu = 0.1357$	0.0068	(0.122, 0.148)
	$\sigma = 203.6385$	10.1276	(183.70, 223.40)
	$\xi = 0.5$	0.025	(0.451, 0.549)
0290101RF	$\mu = 64.6912$	0.0072	(0.129, 0.158)
	$\sigma = 75.5181$	10.3034	(185.87, 226.26)
	$\xi = 0.5$	0.025	(0.451, 0.549)
0180381RF	$\mu = 62.1863$	0.0069	(0.124, 0.151)
	$\sigma = 74.1485$	10.1733	(183.52, 223.40)
	$\xi = 0.5$	0.025	(0.451, 0.549)
0551421RF	$\mu = 63.2359$	0.0070	(0.125, 0.153)
	$\sigma = 75.0153$	10.2781	(185.41, 225.70)
	$\xi = 0.5$	0.025	(0.451, 0.549)
0210461RF	$\mu = 63.6306$	0.0071	(0.127, 0.155)
	$\sigma = 74.9055$	10.2427	(184.77, 224.93)
	$\xi = 0.5$	0.025	(0.451, 0.549)
0010061RF	$\mu = 60.4542$	0.0067	(0.120, 0.146)
	$\sigma = 73.3274$	10.1050	(182.29, 221.90)
	$\xi = 0.5$	0.025	(0.451, 0.549)
0730501RF	$\mu = 62.7490$	0.0069	(0.124, 0.151)
	$\sigma = 74.7894$	10.2597	(185.08, 225.30)
	$\xi = 0.5$	0.025	(0.451, 0.549)

with GEV distribution which standardizes data through parameter estimation. Results of the return level transformation are presented in Table 11. These values indicate that the expected maximum rainfall is likely to occur within the different return periods (5 years, 10 years, 20 years, and 30 years). The return level represents the value of rainfall expected to be exceeded, on average, once every specified number of years. For example, for station 0210461RF, the Smith Model predicts that the maximum rainfall in a 30-year period is likely to be around 126.1762 mm, while the Schlather Model predicts a higher value of 206.0148. The Schlather Model generally predicts higher return levels than the Smith Model, indicating that it assumes a heavier tail for the distribution. After calculating the predicted return level from each existing station, it is then compared with the actual data. Comparing transformed return levels with actual data is important to check if the GEV model predictions are reliable. If the transformed return levels closely macth actual data for past extreme events, this indicates the model is well fit for this study. To see the goodness of the prediction results, the RMSE value can be calculated for each existing model.

Table 11 represents insights into the accuracy of return level predictions for the Smith and

Table 10: Return levels of Smith and Schlather models

Smith Model				
Station	**5 years**	**10 years**	**20 years**	**30 years**
0500551RF	235.3855	346.9372	491.9232	596.1214
0270121RF	231.5650	341.4115	484.1812	586.7865
1910051RF	224.3676	331.5403	470.8346	570.9423
0670281RF	230.4369	340.4029	483.3278	586.0447
0030021RF	224.3560	331.6186	471.0299	571.2218
0290101RF	232.0252	342.0257	484.9955	587.7446
0180381RF	226.4856	334.4911	474.8679	575.7536
0551421RF	229.4558	338.7239	480.7418	582.8068
0210461RF	229.6072	338.7154	480.5254	582.4411
0010061RF	222.9340	329.7435	468.5659	568.3344
0730501RF	228.4684	337.4074	478.9976	580.7553
Schlather Model				
0500551RF	516.5405	903.4861	1450.7099	1870.6095
0270121RF	509.1447	890.5493	1429.9368	1843.8234
1910051RF	500.0445	874.6353	1404.3866	1810.8791
0670281RF	512.5510	896.5127	1439.5164	1856.1778
0030021RF	500.8764	876.0913	1406.7251	1813.8948
0290101RF	509.5739	891.2996	1431.1413	1845.3764
0180381RF	503.1360	880.0421	1413.0678	1822.0728
0551421RF	508.3192	889.1082	1427.6251	1840.8438
0210461RF	506.1021	885.1000	1419.9309	1829.5100
0010061RF	497.3329	868.7439	1395.2264	1798.1487
0730501RF	506.1244	885.1556	1419.9437	1829.6970

Schlather models compared to actual data over different return periods (5, 10, 20, and 30 years). For longer return periods, 20 and 30 years,the Schlather model generally predicts higher return levels compared to Smith model. This indicates that the Schlather model exhibit more extreme tail distribution which indicates stronger spatial dependence in extreme rainfall events. Based on results in Table 11, most stations show an increasing trend in return levels with longer return periods which means that more extreme rainfall events are expected to occur in the future. However, for station 0670281RF, it shows a slight decrese in return levels over time for Smith model indicating that extreme rainfall does not significantly increase throug certain treshold. For stations 0670281RF and 0030021RF, shows consistently high return levels in both model. This indicates these areas are prone to extreme rainfall events. For station 0180381RF, in Smith model,shows a stable and relatively low return levels across all return periods indicating less variability in extreme rainfall. Therefore, the Schlather model predicts more extreme events than Smith model which indicates stronger spatial dependence in rainfall patterns.

Table 12 shows the RMSE calculations for the return levels of the Smith and Schlather models. The calculation of RMSE is important in this study because it helps evaluate which model provides more accurate estimates of extreme rainfall return levels. RMSE indicates the potential impact of extreme rainfall on infrastructure, agriculture, and disaster management by measuring prediction errors. Using RMSE ensures that the chosen model minimizes errors and improves prediction reliability, which is crucial in extreme rainfall modeling where underestimation or overestimation can have significant consequences. The model with the lower RMSE value is generally considered more reliable due to a better fit. Based on the results in Table 12, the Schlather model consistently has lower RMSE values across all return periods compared to the Smith model. These lower RMSE values indicate a better fit and suggest that the Schlather model has superior predictive accuracy for extreme rainfall return levels. However, the differ-

Table 11: Result of return level transformation

Station	5 years	10 years	20 years	30 years
Smith Model				
0500551RF	137.0675	144.3581	152.4935	157.6358
0270121RF	30.4337	79.9864	79.9864	79.9864
1910051RF	123.8218	128.3163	133.4374	136.7226
0670281RF	182.9637	180.5906	177.4179	175.0729
0030021RF	151.1794	151.9219	152.8715	152.8715
0290101RF	97.2461	104.3173	111.7278	116.2228
0180381RF	72.7675	72.7675	72.7675	72.7675
0551421RF	61.4690	72.8758	82.8219	82.8220
0210461RF	84.1263	95.2442	104.1576	126.1762
0010061RF	75.7373	78.6457	82.0220	84.2190
0730501RF	93.4979	97.8114	102.5846	105.5808
Schlather Model				
0500551RF	153.7552	170.4003	187.5831	197.9018
0270121RF	79.9864	99.7856	121.9344	154.9674
1910051RF	134.4265	145.2150	156.5833	163.4737
0670281RF	176.7564	167.6248	153.0692	140.2369
0030021RF	153.0725	155.5361	158.8324	161.2275
0290101RF	100.1042	107.1856	114.2061	118.3100
0180381RF	72.7675	97.8647	103.4576	129.8789
0551421RF	82.8219	122.3445	163.7658	204.3673
0210461RF	106.0435	132.4654	163.0156	206.0148
0010061RF	94.3741	113.6409	131.6495	146.9179
0730501RF	111.0789	118.6737	126.7033	134.4287

ence in RMSE values between the two models is relatively small, implying that the Schlather model performs only slightly better.

Figure 6 presents a spatial heatmap of 11 rainfall stations across Peninsular Malaysia, illustrating the spatial distribution of extreme rainfall intensity or frequency based on the Schlather model. Warmer colors on the heatmap indicate higher intensity or more frequent extreme rainfall events. Each station is plotted according to its latitude and longitude coordinates. The overlaid heat layer reflects the magnitude or frequency of extreme rainfall, with the color gradient ranging from yellow to orange whereby darker shades represent areas expected to experience higher return levels of extreme rainfall. This visualization allows for easy identification of regions more vulnerable to extreme precipitation events by highlighting spatial clustering and geographic hotspots of extremes. As a preliminary step, the heatmap supports further advanced spatial modeling approaches, such as the Smith and Schlather models, by indicating where extreme rainfall tends to concentrate geographically. This study links the shape of the rainfall distributions to flood risk. Stations with higher skewness and kurtosis have heavier tails, which means a higher chance of very large events. Results are organised by broad regions rather than by individual stations. East-coast/northeast-influenced and northern-corridor stations show the largest extremes, while central-west/inland and southern stations are generally lower and less variable. This pattern is consistent with monsoonal forcing and terrain effects. Uncertainty varies by site. Wider confidence intervals occur where there are data gaps, strong convective activity, or complex topography; tighter intervals are seen at lower-variance sites. Goodness-of-fit is supported by the diagnostic plots and KolmogorovSmirnov tests.

The spatial extremes analysis shows that allowing a time-varying range $\lambda(t) = \exp(\lambda_0 + \lambda_1 t)$ in the Schlather model captures changes in spatial dependence better than stationary alternatives. Extremal-coefficient behaviour and information criteria indicate improved dependence

Table 12: RMSE for the return levels (rotated and split side-by-side)

Station ID	Actual	Smith Prediction	Schlather Prediction	Smith Error	Schlather Error
5 years return level (1994-1998)					
0500551RF	763.5	137.0675	153.7552	626.4325	609.7448
0270121RF	463.46	30.4337	79.9864	433.0263	383.4736
1910051RF	811.6	123.8218	134.4265	687.7782	677.1735
0670281RF	737.77	182.9637	176.7564	554.8063	561.0136
0030021RF	816.9	151.1794	153.0725	665.7206	663.8275
0290101RF	520.0	92.6168	100.1042	427.3832	419.8958
0180381RF	352.4	72.7675	72.7675	279.6325	279.6325
0551421RF	459.8	61.4690	82.8219	398.3310	376.9781
0210461RF	363.31	84.1263	84.1263	279.1837	279.1837
0010061RF	285.59	75.7373	82.7274	209.8527	202.8626
0730501RF	331.62	93.4979	103.4531	238.1221	228.1669
RMSE				404.6679	387.6982
10 years return level (1994-2003)					
0500551RF	1486.30	144.3581	170.4003	1341.9419	1315.8997
0270121RF	934.46	79.9864	99.7856	854.4736	834.6744
1910051RF	1368.30	128.3163	145.2150	1239.9837	1223.0850
0670281RF	1549.97	180.5906	167.6248	1369.3794	1382.3452
0030021RF	1923.80	151.9219	155.5361	1771.8781	1768.2639
0290101RF	931.50	95.9411	107.1856	835.5589	824.3144
0180381RF	657.20	72.7675	97.8647	584.4325	559.3353
0551421RF	982.80	72.8758	122.3445	909.9242	860.4555
0210461RF	714.64	95.2442	98.6789	619.3958	615.9611
0010061RF	703.31	78.6457	90.0479	624.6643	613.2621
0730501RF	918.12	97.8114	113.0951	820.3086	805.0249
RMSE				947.3259	922.1997

Station ID	Actual	Smith Prediction	Schlather Prediction	Smith Error	Schlather Error
20 years return level (1994-2013)					
0500551RF	3088.40	152.4935	187.5831	2935.9065	2900.8169
0270121RF	1927.96	79.9864	121.9344	1847.9736	1806.0256
1910051RF	2607.80	133.4374	156.5833	2474.3626	2451.2167
0670281RF	4488.17	177.4179	153.0692	4310.7521	4335.1008
0030021RF	3531.10	152.8715	158.8324	3378.2285	3372.2676
0290101RF	1800.66	99.5468	114.2061	1701.1132	1686.4539
0180381RF	1495.10	72.7675	103.4576	1422.3325	1391.6424
0551421RF	1821.83	82.8219	163.7658	1739.0081	1658.0642
0210461RF	1760.70	104.1576	239.4578	1656.5424	1521.2422
0010061RF	1766.61	82.0220	97.9547	1684.5880	1668.6553
0730501RF	1808.57	102.5846	122.8882	1705.9854	1685.6818
RMSE				2081.6880	2045.0778
30 years return level (1994-2023)					
0500551RF	4329.20	157.6358	197.9018	4171.5642	4131.2982
0270121RF	2934.76	79.9864	154.9674	2854.7736	2779.7926
1910051RF	4014.00	136.7226	163.4737	3877.2774	3877.2774
0670281RF	7524.87	175.0729	140.2369	7349.7971	7384.6331
0030021RF	5235.00	152.8715	161.2275	5082.1285	5073.7725
0290101RF	2821.36	101.7775	118.3100	2719.5825	2703.0500
0180381RF	2460.80	72.7675	129.8789	2388.0325	2330.9211
0551421RF	2808.31	82.8219	204.3673	2725.4880	2603.9427
0210461RF	2869.60	126.1762	288.3421	2743.4238	2581.2579
0010061RF	2714.21	84.2190	102.8159	2629.9910	2611.3941
0730501RF	2899.86	105.5808	128.6968	2794.2792	2771.1632
RMSE				3498.3025	3446.21

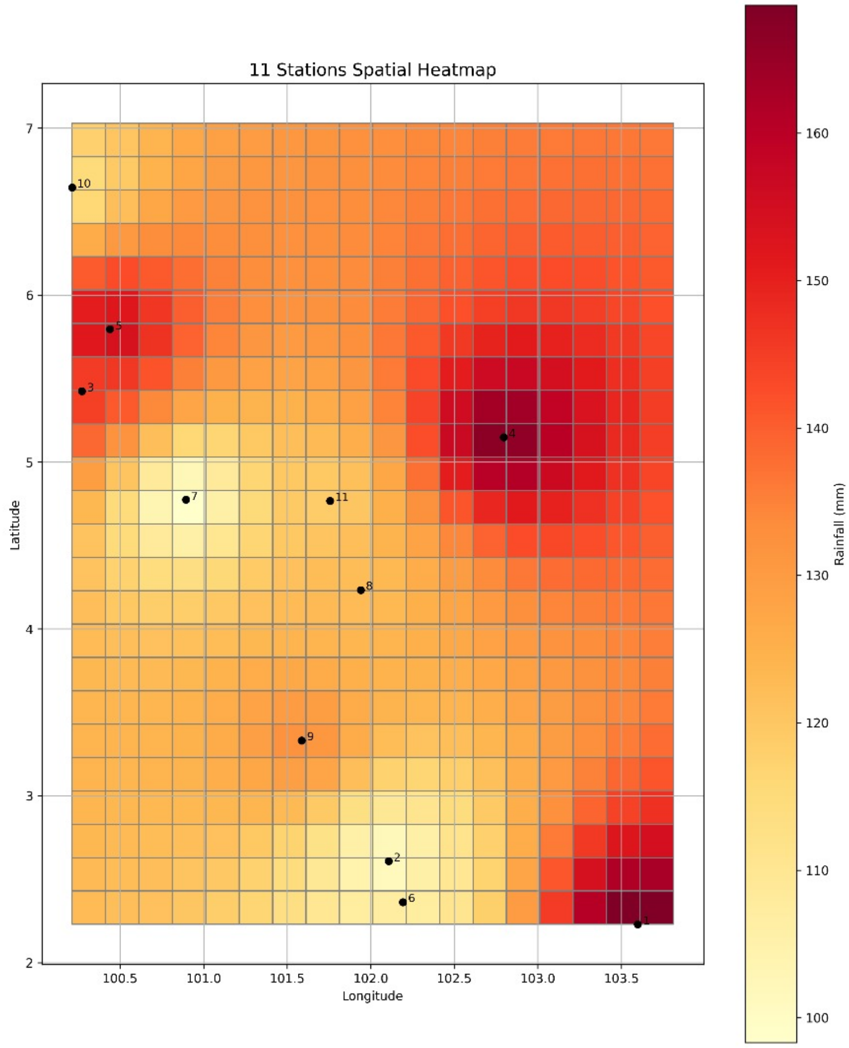

Figure 6: Heatmap of stations across Peninsular Malaysia

fit for the Schlather specifications. Across 530-year horizons, Schlather gives higher (more conservative) return levels than Smith and generally lower RMSE, indicating better predictive performance and a more suitable description of spatial dependence for extreme rainfall in Peninsular Malaysia. These results have practical use. Schlather-based return levels can be used as upper bounds for drainage sizing, flood-zone delineation, and dam-safety checks, especially in east-coast and northern sub-catchments where tails are heavy and uncertainty is larger. Sensitivity checks against Smith-based values help reduce the risk of under-design. Where long-horizon Schlather levels are much higher than Smith, more conservative zoning, use of floodplains or green buffers, and review of reservoir operating rules are recommended. The framework can be extended by adding climate covariates (e.g., El Nio-Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD)) to represent non-stationarity in both the marginals and the dependence, supporting scenario-based planning and early-warning thresholds that reflect tail risk.

4. Conclusion

This study shows that the GEV distribution is appropriate for modelling extreme rainfall in Peninsular Malaysia. Based on the extremal-coefficient analysis and model comparisons, the Schlather max-stable model provides a more accurate representation of spatial dependence than the Smith model, with stronger dependence persisting over longer distances. Return levels for 5-, 10-, 20- and 30-year periods were transformed to standard Frchet margins to enable spatial modelling on a common scale while preserving tail dependence. Across horizons, the Schlather model yields higher (more conservative) return-level estimates and lower RMSE than the Smith model, indicating better predictive performance and a closer match to the spatial dependence structure of extreme rainfall in the region. Allowing a time-varying range parameter $\lambda(t) = \exp(\lambda_0 + \lambda_1 t)$ in the Schlather framework further improves the fit where dependence changes over time. These findings are useful for improving extreme-event prediction, flood-risk assessment, and the planning and design of drainage systems, flood zones, and dam safety measures in Peninsular Malaysia. Future work could consider alternative spatial extreme-value approaches, incorporate additional covariates (e.g., ENSO, IOD) to represent non-stationarity more explicitly, and extend the analysis across different spatial scales. Such extensions would refine uncertainty estimates and deepen understanding of extreme rainfall behaviour in Peninsular Malaysia.

Acknowledgments

I would like to express my deepest gratitude to my supervisor, Dr Syafrina Abdul Halim, for her invaluable guidance, patience, and encouragement. This work was supported by Ministry of Higher Education under Fundamental Research Grant Scheme (FRGS/1/2024/STG06/UPM/02/6).

References

Abdelmoaty H.M. & Papalexiou S.M. 2023. Changes of extreme precipitation in CMIP6 projections: Should we use stationary or nonstationary models? *Journal of Climate* **36**(9): 2999–3014.

Agilan V. & Umamahesh N.V. 2017. What are the best covariates for developing non-stationary rainfall intensity-duration-frequency relationship? *Advances in Water Resources* **101**: 11–22.

Coles S., Bawa J., Trenner L. & Dorazio P. 2001. *An Introduction to Statistical Modeling of Extreme Values*. London, UK: Springer.

Deni S.M., Suhaila J., Wan Zin W.Z. & Jemain A.A. 2010. Spatial trends of dry spells over peninsular malaysia during monsoon seasons. *Theoretical and Applied Climatology* **99**(3): 357–371.

Fisher R.A. & Tippett L.H.C. 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society* **24**(2): 180–190.

Gilli M. & Këllezi E. 2006. An application of extreme value theory for measuring financial risk. *Computational Economics* **27**(2): 207–228.

Huser R. & Genton M.G. 2016. Non-stationary dependence structures for spatial extremes. *Journal of Agricultural, Biological, and Environmental Statistics* **21**(3): 470–491.

Huser R., Opitz T. & Wadsworth J.L. 2025. Modeling of spatial extremes in environmental data science: Time to move away from max-stable processes. *Environmental Data Science* **4**: e3.

Huser R. & Wadsworth J.L. 2022. Advances in statistical modeling of spatial extremes. *Wiley Interdisciplinary Reviews: Computational Statistics* **14**(1): e1537.

Karuna E.E., Nugroho S. & Rizal J. 2023. Extreme rainfall modeling using spatial extreme value with max-stable processes models of Smith, Brown-Resnick, and Schlather. *International Journal of Scientific and Technical Research in Engineering* **8**(2): 1–14.

Love C.A., Skahill B.E., Russell B.T., Baggett J.S. & AghaKouchak A. 2022. An effective trend surface fitting framework for spatial analysis of extreme events. *Geophysical Research Letters* **49**(11): e2022GL098132.

Lu G.Y. & Wong D.W. 2008. An adaptive inverse-distance weighting spatial interpolation technique. *Computers & Geosciences* **34**(9): 1044–1055.

McNeil A.J. 1999. Extreme value theory for risk managers. *Departement Mathematik ETH Zentrum* **12**(5): 217–237.

Padoan S.A., Ribatet M. & Sisson S.A. 2010. Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association* **105**(489): 263–277.

Ribatet M., Singleton R. & R Core Team 2018. Spatialextremes: Modelling spatial extremes, R package version 2.0-7. https://CRAN.R-project.org/package=SpatialExtremes.

Saimi F.M., Hamzah F.M., Toriman M.E., Jaafar O. & Tajudin H. 2020. Trend and linearity analysis of meteorological parameters in peninsular malaysia. *Sustainability* **12**(22): 9533.

Zeng L., Bi H., Li Y., Liu X., Li S. & Chen J. 2022. Nonstationary annual maximum flood frequency analysis using a conceptual hydrologic model with time-varying parameters. *Water* **14**(23): 3959.

Department of Mathematics and Statistics
Faculty of Science
Universiti Putra Malaysia
43400 UPM Serdang
Selangor, MALAYSIA
E-mail: 211136@student.upm.edu.my

Institute of Mathematical Research
Universiti Putra Malaysia
43400 UPM Serdang
Selangor, MALAYSIA
*E-mail: syafrina@upm.edu.my**

Received: 9 June 2025
Accepted: 5 October 2025

*Corresponding author