Regional frequency analysis of annual daily rainfall maxima in Skåne, Sweden

Mahmood Ul Hassan1 | Zahra Noreen2 | Rashid Ahmed3

1Department of Statistics, Stockholm University, Stockholm, Sweden
2Division of Science and Technology, University of Education, Lahore, Pakistan
3Department of Statistics, The Islamia University of Bahawalpur, Bahawalpur, Pakistan

Abstract

Extreme daily rainfall events are critical for the urban drainage system, human life, agriculture and small catchments. The information about extreme rainfall magnitudes and frequencies is immensely important for civil engineers, city planners, scientists related to water management, rescue operations and flood control works. This study illustrates the results of regional frequency analysis (RFA) of annual maximum daily rainfall (AMDR) of Skåne County, Sweden. L-moments based heterogeneity measure (H) reveals that the Skåne County is a homogeneous region. Based on the L-moment ratio diagram and ZDist statistic results, the generalized normal (GNO) distribution is selected as the most suitable regional distribution. The accuracy measures used in K-fold cross validation indicate that support vector machine (SVM) model is an appropriate model to find the index rainfall at ungauged sites in the region. The sites characteristics, elevation and latitude are identified as the most important variables to explain the variation in mean annual maximum daily rainfall (MAMDR). Finally, spatial maps of predicted MAMDR for different return periods are constructed by using index rainfall combined with regional quantiles. Spatial maps offer an overall view of the expected MAMDR in the region that is helpful for multiple decision makers including infrastructure planners, city planners, emergency managers, engineers and many others.

KEYWORDS

K-fold cross validation, L-moments, Monte Carlo simulation, quantile estimates, regional frequency analysis, spatial map

1 | INTRODUCTION

Daily extreme rainfall is one of the utmost environmental events which cause sizeable damages throughout the world. The rainfall recorded in the last century was the highest over the past 8,500 years (Hong et al., 2014). A reliable estimate of extreme rainfall magnitude is important for flood management, irrigation, rescue operations, water management, construction of urban drainage systems and many other purposes.

Skåne is the tenth largest county of Sweden. It has the best agricultural land and is known as a breadbasket of Sweden. It stands for about half of Sweden’s food production (Skåne Region, 2014). Rainfall is an important factor for agriculture in Skåne. Skåne agriculture and food industries are highly dependent on this hydrological
event, and at the same time are quite vulnerable to natural disasters like extreme rainfall, floods and droughts. Therefore, it is critical to have precise estimates of rainfall quantiles in this region.

Skåne County experienced flash-floods on 31st August 2014. According to Swedish Meteorological and Hydrological Institute (SMHI), Malmö city received a month’s worth of rainfall (88 mm). Another town, Falsterbo, received 43 mm of rain during the night and early morning see Radio Sweden (2014) and The Local (2014). Flash floods may cause numerous deaths as they occur with a small warning which is not sufficient to warn and evacuate people. Therefore, the annual maximum daily rainfall (AMDR) is considered in this study, which is useful knowledge for water management, urban planning, drainage design, agricultural water management, planning of flood management, insurance studies, tourism industries, preventive measures for natural disasters and construction of hydraulic structures (bridges, spillways embankments, etc.).

There are mainly two approaches to frequency analysis in a practice known as at-site and regional analysis. We need a substantial rainfall data series for at-site extreme value extrapolation which is rarely available in practice. In this case, an efficient statistical method should be preferred for extreme value prediction. We can efficiently utilize regional frequency analysis (RFA) to reduce the prediction uncertainty if the region having rainfall-gauging sites can be established as a homogeneous region. One vital advantage of RFA over at-site is that the results of the regional quantile estimates can effectively be used to estimate the quantiles at ungauged sites. In RFA, using the L-moment method, the data from all sites in a region, regardless of different record lengths and missing values, can be combined to estimate the rainfall quantile for different return periods at any location in the region (Hosking and Wallis, 1997).

The estimation of rainfall magnitude at ungauged sites for different return periods is a key issue. Therefore, further modelling is required to estimate the index rainfall at ungauged sites in the region (Hosking and Wallis, 1997). The use of simple/multiple linear regression (MLR) model to estimate the extreme quantiles at gauges/ungauged sites is commonly traced in the literature (see e.g., Al Mamun et al. 2018; Hussain et al. 2017; Modarres 2008; Um et al. 2010; Naoum and Tsanis 2004b; Hallegoegris and Alfredsen 2017). To estimate the flood quantiles at ungauged sites, Desai and Ouarda (2021) have utilized random forest (RF) regression. Malekinzehad and Zare-Garizi (2014) and Forestieri et al. (2018) use mean annual precipitation as an explanatory variable in a simple linear regression model to predict index rainfall. Szolgyay et al. (2009) and Hussain et al. (2017) develop a linear regression model to predict index rainfall using elevation as an explanatory variable.

Four machine learning models are considered in the present study along with one geostatistical interpolation model for estimating rainfall quantiles at ungauged sites. Machine learning models include MLR, RF, Gaussian process (GP) and support vector machine (SVM) model which require some characteristics of gauging sites (e.g., elevation and average annual temperature) to predict average rainfall at ungauged sites. Among geostatistical models, the inverse distance weighting (IDW) model is considered which does not require site characteristics. To evaluate the model prediction performance, a K-fold cross-validation (CV) algorithm is used. The K-fold CV is a robust method to estimate the accuracy of a model (James et al., 2013). The best predictive model is chosen based on accuracy measures defined in Section 3.7.

This is the first study of regional rainfall frequency analysis (RRFA) in the Skåne region. We use the methodology proposed by Hosking and Wallis (1997). The main objectives of this study include: the identification of homogeneous regions/region, selecting the best suitable regional frequency distribution and choosing the most appropriate interpolation model for the estimation of mean annual maximum daily rainfall (MAMDR) at ungauged sites in the region. Lastly, we estimate the extreme quantiles for different return periods with non-exceedance probability at gauged/ungauged sites in the Skåne region using regional quantiles and predicted average rainfall magnitude with the suitable model. The spatial map of the expected MAMDR in the region for different return periods is also presented.

2 STUDY REGION AND DATA

County Skåne (English: Scania) of Sweden is located in the southeast of the country having a total area of 11,303 km². The left panel of Figure 1 indicates the location of Skåne in Sweden. It shares the border with the counties of Halland, Kronoberg and Blekinge. It is also connected with the capital of Danmark through the Öresund bridge. Skåne is one of the warmest regions of Sweden with the average daily high temperature of 12°C centigrade. The daily rainfall data for each site are collected from the Swedish Meteorological and Hydrological Institute (SMHI) (www.smhi.se). The maximum value out of the daily rainfall record of each year for the complete span of data at each site is used in this study. A map of rainfall-gauged sites used in this study is presented in Figure 1. Ystad is the only site where 6 years (1984–1989) data are missing. The period of the data series varies from 22 to 70 years. The average record length of gauging sites is ~48 years. The summary of rainfall gauging sites characteristics is presented in Table 1. The average elevation of the region is 42 m while the maximum elevation in the region is 226 m.
FIGURE 1  Geographical locations of 24 sites and the study region [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1  Summary of basic information about gauging sites in Skåne

| Serial no. | Site names | Site no. | Time period | Latitude | Longitude | Elevation |
|------------|------------|----------|-------------|----------|-----------|-----------|
| 1          | Bromölla   | 64,070   | 1961–2019   | 56.06    | 14.48     | 5.57      |
| 2          | Brösarp    | 54,410   | 1996–2019   | 55.72    | 14.10     | 49.57     |
| 3          | Falsterbo  | 52,230   | 1950–2019   | 55.38    | 12.82     | 1.54      |
| 4          | Gillastig  | 63,010   | 1962–2019   | 56.01    | 13.23     | 149.19    |
| 5          | Hallands Väderö | 62,260 | 1995–2019   | 56.45    | 12.55     | 9.17      |
| 6          | Helsingborg | 62,020  | 1992–2019   | 56.04    | 12.75     | 46.03     |
| 7          | Höoby      | 53,530   | 1996–2019   | 55.86    | 13.67     | 113.39    |
| 8          | Klippan    | 63,070   | 1958–2019   | 56.14    | 13.25     | 55.75     |
| 9          | Kristianstad | 64,030  | 1997–2019   | 56.01    | 14.29     | 10.00     |
| 10         | Landskrona | 52,520   | 1960–2019   | 55.87    | 12.85     | 9.00      |
| 11         | Louisefred | 62,150   | 1965–2019   | 56.24    | 12.58     | 12.00     |
| 12         | Love Stad  | 53,400   | 1979–2019   | 55.64    | 13.89     | 76.56     |
| 13         | Lund       | 53,430   | 1961–2019   | 55.69    | 13.23     | 26.45     |
| 14         | Malmö      | 53,360   | 1950–2019   | 55.57    | 13.08     | 20.00     |
| 15         | Norra Ströö | 64,080  | 1967–2019   | 56.10    | 14.00     | 15.00     |
| 16         | örkelljunga | 63,170  | 1976–2019   | 56.28    | 13.30     | 76.94     |
| 17         | Skillinge  | 54,290   | 1995–2019   | 55.49    | 14.31     | 4.02      |
| 18         | Stehag     | 53,540   | 1958–2019   | 55.88    | 13.44     | 89.36     |
| 19         | Stiby      | 54,340   | 1998–2019   | 55.56    | 14.19     | 71.16     |
| 20         | Tångå      | 62,120   | 1980–2019   | 56.20    | 12.78     | 13.00     |
| 21         | Tomelilla  | 53,320   | 1958–2019   | 55.55    | 13.95     | 66.05     |
| 22         | Trelleborg | 53,230   | 1950–2019   | 55.38    | 13.13     | 5.00      |
| 23         | Vomb       | 53,410   | 1953–2019   | 55.66    | 13.53     | 25.00     |
| 24         | Ystad      | 53,260   | 1958–2019   | 55.43    | 13.81     | 20.00     |
3 | METHODOLOGY

3.1 | Theory of L-moments method

RFA, which based on L-moments, is used in this study. The L-moments can be defined for any random variable \( z \) if its mean exists (Hosking, 1990). The probability weighted moment (PWM) of \( r \)th order \( (\alpha_r) \) is defined as

\[
\alpha_r = \frac{1}{r} \int_0^r z(F(z))F'(z)dz, \quad r=0,1,2,...
\]

The first four L-moments in terms of PWMs can be computed by using the following relationship

\[
\begin{align*}
\lambda_1 &= \alpha_0 \\
\lambda_2 &= 2\alpha_1 - \alpha_0 \\
\lambda_3 &= 6\alpha_2 - 6\alpha_1 + \alpha_0 \\
\lambda_4 &= 20\alpha_3 - 30\alpha_2 + 12\alpha_1 - \alpha_0
\end{align*}
\]

where \( \lambda_1 \) represents the location and \( \lambda_2 \) is a dispersion. The L-moments ratio can be defined as

\[
\begin{align*}
\text{L-CV} &= \text{L-Coefficient of variation (r)} = \frac{\lambda_2}{\lambda_1} \\
\text{L-Skewness (r)} &= \frac{\lambda_3}{\lambda_2} \\
\text{L-Kurtosis (r)} &= \frac{\lambda_4}{\lambda_2}
\end{align*}
\]

The first four unbiased estimates of PWM \( (\alpha_r) \) for any probability distribution can be computed as

\[
\begin{align*}
b_0 &= n^{-1} \sum_{p=1}^{n} z_{p,n} \\
b_1 &= n^{-1} \sum_{p=2}^{n} \frac{(p-1)}{(n-1)} z_{p,n} \\
b_2 &= n^{-1} \sum_{p=3}^{n} \frac{(p-1)(p-2)}{(n-1)(n-2)} z_{p,n} \\
b_3 &= n^{-1} \sum_{p=4}^{n} \frac{(p-1)(p-2)(p-3)}{(n-1)(n-2)(n-3)} z_{p,n}
\end{align*}
\]

where \( z_{p,n} \) is the data set in ascending order. Hosking (1990) defines the sample L-moments denoted by \( l_1, l_2, l_3 \) and \( l_4 \). The sample L-moment ratios are defined by Hosking and Wallis (1997) as

\[
\begin{align*}
t &= \frac{l_2}{l_1}, \quad t_3 = \frac{l_3}{l_2}, \quad t_4 = \frac{l_4}{l_2}, \quad t_r = \frac{l_r}{l_2}, \quad r=3,4,...
\end{align*}
\]

where \( t \) is a sample L-coefficient of variation (L-CV), \( t_3 \) represents a sample L-coefficient of skewness (L-Skewness) and \( t_4 \) is a sample L-coefficient of kurtosis (L-Kurtosis).

3.2 | Discordancy measure

The discordancy measure \( D_i \) is used to identify a discordant site in the region. It is based on L-moments. The discordancy measure is defined by

\[
D_i = \frac{1}{3}N(\mu_i - \bar{\mu})^T A^{-1}(\mu_i - \bar{\mu}), \quad i=1,2,...,N,
\]

where \( N \) is the number of sites in the region, \( \mu_i = [t_1, t_3, t_4]^T \), \( \bar{\mu} = [\bar{t}_1, \bar{t}_3, \bar{t}_4]^T \) and \( A = \sum_{i=1}^{N} (\mu_i - \bar{\mu})(\mu_i - \bar{\mu})^T \).

Site \( i \) in the region is considered as a discordant site if \( D_i \) is greater than the critical value. For \( N \geq 15 \) the critical value for \( D_i \) is 3. Hosking and Wallis (1997) provide the table of critical values of \( D_i \) for \( N < 15 \).

3.3 | Heterogeneity measures

Hosking and Wallis (1997) propose the heterogeneity measures \( (H) \) for identification of homogeneous region. \( H \) is computed as

\[
H = \frac{V - \mu_V}{\sigma_V}, \quad \text{where} \quad V = \left( \frac{\sum_{i=1}^{N} n_i (t_i^R - \bar{t}^R)^2}{\sum_{i=1}^{N} n_i} \right)^{\frac{1}{2}}
\]

where \( \mu_V \) and \( \sigma_V \) are the mean and \( SD \) of \( V \) respectively. The values of \( V \) are obtained through simulation. \( N \) is the number of sites in the region, \( n_i \) is the record length for site \( i \), \( t = L - CV \) and \( \bar{t}^R \) is a weighted average of \( L - CV \).

To perform the simulation, we first fit a Kappa distribution using the regional average L-moment ratios \( \bar{t}_1, \bar{t}_3, \bar{t}_4 \) and use this distribution to simulate the region. In a simulated region, sites have same record length as in the original data. We calculate the value of \( V \) for each simulated region.

For sufficient large value of \( H \), we declare the region heterogeneous. The region under study is acceptably homogeneous for \( H < 1 \), for \( 1 \leq H < 2 \) the region is possibly heterogeneous and for \( H \geq 2 \) the region is definitely heterogeneous.
### 3.4 Goodness-of-fit methods

Goodness-of-fit (GOF) methods are used to identify the best suitable regional frequency distribution. In this study, we consider the L-moment ratio diagram and $Z_{\text{Dist}}$ criterion to select the most appropriate regional frequency distribution.

The L-moment ratio diagram, which is gaining much attention in the literature, is a means to find the suitable distribution for RFA (Peel et al., 2001). The L-moment ratio diagram is a scatter plot of L-skewness ($t_3$) versus L-kurtosis ($t_4$) of each site in the region along with a point of their regional average. We can assess the appropriateness of a particular probability distribution using the relationship of sample L-moment ratios ($t_3$, $t_4$) with L-moment ratios of various probability distributions. These relationships are described by Hosking and Wallis (1997). The probability distribution is appropriate for the region if its growth curve is close to average sample L-moment ratio.

The $Z_{\text{Dist}}$ criterion is another method proposed by Hosking and Wallis (1997) to select the appropriate distribution for the region. It is defined as

$$ Z_{\text{Dist}} = \frac{(r_{4}^{\text{Dist}} - t_{4}^{R} + B_{R})}{\sigma_{4}}, $$

where $r_{4}^{\text{Dist}}$ = L-kurtosis of the fitted distribution, $B_{R}$= bias of $t_{4}^{R}$, $t_{4}^{R}$ = weighted regional average L-kurtosis, $\sigma_{4}$ = SD of $t_{4}^{R}$ are obtained by simulation.

The candidate distribution is accepted as a suitable distribution for the region if $|Z_{\text{Dist}}| < 1.64$. It is possible that many probability distributions are accepted as appropriate distributions for the region. In that case, the most appropriate distribution among the accepted distributions is the one that has the smallest value of $|Z_{\text{Dist}}|$.

### 3.5 Quantiles estimate at gauged/ungauged sites

After the selection of the most appropriate regional frequency distribution, the main goal of the RRFA is to estimate precise quantiles at gauged/ungauged sites in the region. The quantile estimate for the $i$th site ($\hat{Q}_i(F)$) in the region with non-exceedance probability $F$ can be computed as

$$ \hat{Q}_i(F) = l_i^{(i)} \hat{q}_e(F), $$

where $\hat{q}_e(F)$ represents the regional quantile estimate with non-exceedance probability $F$.

To find the quantile at an ungauged site, we need the value of MAMDR on that site. We cannot compute the MAMDR at ungauged sites due to unavailability of data on these sites. In this situation, we can model the relationship between the MAMDR at gauged sites ($l_i$) and some geographical or climate characteristics of gauged sites.

To estimate the MAMDR at ungauged sites in the Skåne region, we have used inverse distance weighted (IDW), MLR, RF, GP and SVM models. We have to identify suitable explanatory variables (site characteristics) that explain variation in $l_i$ for machine learning models (MLR, RF, GP and SVM). The details of IDW, MLR, GP, RF and SVM are provided in Section 3.7. The K-fold cross validation is used to evaluate the prediction performance of these models. The best model is selected based on the accuracy measures in K-fold CV. Finally, the most suitable model is used to estimate the MAMDR ($\hat{l}_i$) at ungauged sites in the region. To find the quantiles at ungauged sites for different return periods with non-exceedance probability $F$, we can replace $l_i$ in Equation (1) by $\hat{l}_i$.

### 3.6 Accuracy measures

For accuracy measures, we have used the mean absolute error (MAE), root mean square error (RMSE), prediction error rate (PER) and correlation coefficient ($R^2$) in order to evaluate the model prediction performance. These accuracy measures are defined as

$$ MAE = \frac{1}{n} \sum_{j=1}^{n} |z_j - \hat{z}_j|, \quad RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^{n} (z_j - \hat{z}_j)^2}, $$

$$ PER = \frac{RMSE}{\bar{z}} \quad \text{and} \quad R^2 = \frac{\sum_{j=1}^{n} (z_j - \bar{z})^2}{\sum_{j=1}^{n} (z_j - \hat{z}_j)^2}, $$

where $z_j$ are observed values, $\hat{z}_j$ are expected values from the model and $\bar{z}$ is a sample mean of $z_j$.

### 3.7 Interpolations models

#### 3.7.1 IDW

The IDW method is one of the oldest, least complicated and most commonly used interpolation methods. The
IDW is a deterministic model. It gives more weight to the observations which are closer to known values when estimating the values at unobserved points. The IDW formula is defined as

\[ \hat{Z}(s_0) = \frac{\sum_{i=1}^{n} W_i(s_0)Z(s_i)}{\sum_{i=1}^{n} W_i(s_0)} \]

where \( \hat{Z}(s_0) \) is the estimated average rainfall value at ungauged position \( x_0 \), \( Z(s_i) \) is the observed rainfall value at gauged location \( s_i \), \( d(s_0,s_i) \) is a distance measure from ungauged position \( s_0 \) to known gauging point \( s_i \) and \( m \) is a coefficient that adjusts the weight according to distance. For more detail of the IDW method see Johnston et al. (2001).

3.7.2  |  MLR model

MLR describes the linear relationship between a response variable and two or more explanatory variables by fitting a linear equation to the observed data. The model for MLR can be defined as

\[ z_i = f(x_{i1}, x_{i2}, ..., x_{ip}) + \xi_i, \ i = 1, 2, 3, ..., n. \]  or  
\[ z_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \xi_i, \ i = 1, 2, 3, ..., n. \]

Where \( z \) is a response (dependent) variable whereas \( x_1, x_2, ..., x_p \) are independent or explanatory variables. To estimate the parameters with the least squares method, we minimize the squared difference between observed and expected response that is, minimize \( \sum_{i=1}^{n} (z_i - \hat{z}_i)^2 \) where \( \hat{z}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip} \).

3.7.3  |  GP

The GP regression is a nonparametric Bayesian approach to regression. It is a kernel based probabilistic model and can be viewed as a Bayesian version of SVM models. The GP regression models are robust against the model overfitting problem (Thapa et al., 2020). The model can be summarized as follows.

The observational model:

\[ z_i | f, \varphi \sim \prod_{i=1}^{n} p(z_i | f, \varphi) \]

We assume a GP prior:

\[ f(x) | \theta \sim GP(m(x), k(x, x' | \theta)), \]

and hyper prior:

\[ \theta, \varphi \sim p(\theta)p(\varphi). \]

Where \( m(x) \) is a mean function, \( k(X, X' | \theta) \) indicates the covariance function of a process \( f(x) \), \( \theta \) represents kernel parameters and \( \varphi \) indicates the parameters of the observational model. The joint distribution of test output \((\hat{f})\) and training output \((f)\) based on the prior is expressed as

\[ f_{\hat{f}} | \varphi, \theta \sim N \left( 0, \begin{bmatrix} k(x,x | \theta) & k(x,\hat{x} | \theta) \\ k(\hat{x},x | \theta) & k(\hat{x},\hat{x} | \theta) \end{bmatrix} \right). \]

The marginal distribution of \( \hat{f} \) is \( P(\hat{f}|\hat{x},\theta) = N(\hat{f}|0, K(\hat{x},\hat{x}|\theta)) \). The conditional distribution of \( \hat{f} \) (test output) is defined as

\[ \hat{f} | f, \hat{x}, x, \theta \sim N(\mu(\hat{f}), COV(\hat{f})), \]

where

\[ \mu(\hat{f}) = K(\hat{x}, x | \theta)K(x, x | \theta)^{-1}f \]
\[ COV(\hat{f}) = K(\hat{x}, \hat{x} | \theta) - K(\hat{x}, x | \theta)K(x, x | \theta)^{-1}K(x, \hat{x} | \theta) \]

In the predictive equation of \( \hat{f} \), one could use a different Kernel function, for example, Gaussian, linear, Bessel and polynomial. In this study, we have used a polynomial kernel.

3.7.4  |  SVM

The SVM is a well-known supervised machine learning method which is utilized in regression, pattern recognition and classification problems. It is an efficient and robust machine learning algorithm for flood prediction (Dehghan et al., 2014). The SVM algorithm uses the structural risk minimization theory and reduces the overfitting problem (Yu et al., 2006). The objective of SVM for regression is to find a regression function that best approximates the response observations \((z)\) with an error tolerance value \(\varepsilon\). The decision function can be defined by the following equation
\[ f(x) = w \cdot \phi(x) + \delta, \]

where \( w \) and \( \delta \) are vectors of parameters of the functions and \( \phi(x) \) is a non-linear function. The quality of estimation is measured by the loss function \( L(z_i) \). The SVM regression uses Vapnik’s \( \varepsilon \)-insensitive loss function:

\[
L_{\varepsilon}(z_i) = \begin{cases} 
0 & \text{for } |z_i - [w \cdot \phi(x_i) + \delta]| < \varepsilon \\
|z_i - [w \cdot \phi(x_i) + \delta]| - \varepsilon & \text{for } |z_i - [w \cdot \phi(x_i) + \delta]| \geq \varepsilon 
\end{cases}
\]

The SVM problem can be expressed as the following convex optimization problem

\[
\min_{w, b, \xi, \xi^*} \frac{1}{2} w^2 + C \sum_{i=1}^{l} (\xi_i + \xi_i^*) \\
\text{Subject to } z_i - [w \cdot \phi(x_i) + \delta] - \varepsilon + \xi_i - \xi_i^* \leq 0, \\
\quad w \cdot \phi(x_i) + \delta - z_i \leq \varepsilon + \xi_i^* \\
\quad \xi_i, \xi_i^* \geq 0, \quad i = 1, 2, ..., l,
\]

where \( \xi_i \) and \( \xi_i^* \) are slack variables indicating the upper and lower training error subject to error tolerance \( \varepsilon \) and \( C \) (positive constant) that determines degree of penalized loss when a training error occurs. The optimization problem can be transformed as a dual problem using a Lagrangian function and this leads to the following dual function

\[
f(x) = \sum_{i=1}^{m} (\gamma_i - \gamma_i^*) K(x_i, x_j) + \delta \quad \text{Subject to } 0 \leq \gamma_i \leq C, 0 \leq \gamma_i^* \leq C.
\]

Where \( \gamma, \gamma^* \) are Lagrange multiplier factor, \( m \) represents the number of support vectors and \( K(x_i, x_j) \) is a kernel function. One could select different Kernel functions, for example, linear, Gaussian, sigmoid and polynomial. In this study, we have used a polynomial kernel. For further details of the SVM method see Vapnik (1995).

3.7.5 RF

The RF is the most popular, oldest and an efficient decision tree-based machine learning predictive model. It is an industrial workhorse of machine learning.

The algorithm of RF is summarized by following steps

a. Select \( n \) data points and \( p \) features (explanatory variables) randomly from the training data set.

b. Construct decision tree based on the dataset in step (a)

c. Select the number of trees (\( N \)) that we want to build and repeat step (a) and (b)

d. To predict a value \( (z) \) for a new data point, we use each of our \( N \) trees to predict the value of \( y \) and take an average across all of the predicted \( z \) values. This average serves as a predicted value for the new data point.

The RF is competitive with VSM and Neural Networks in many practical tasks of prediction. However, it is much faster to train as it has fewer parameters. There is less risk of overfitting in RF because it reduces the variance by training on randomly selected different samples from the training data. For further detail of the RF method see Hastie et al. (2009).

3.8 K-fold CV

K-fold CV is a robust method for estimating the accuracy of the model. K-fold CV is better than leave-one-out cross validation (LOOCV). It often provides more accurate estimates of test error rates (James et al., 2013). In this study, K-fold CV is used to examine the model’s prediction performance. The K-fold CV algorithm has the following steps:

- In the first step, we randomly divide the data into \( K \) groups/subsets/folds.
- Treat one group of data as a test or validation data and train the model on another group of data (called training data)
- Test the model on a validation subset and compute some accuracy measures, for example, RMSE, \( R^2 \), MAE and PER as evaluation metrics.
- Repeat this procedure \( K \) times. Each time a different group of observations serves as validation data.
- Take the average of the accuracy measures in \( K \) records. This average serves as a model performance measure.

3.9 Spatial map

Spatial mapping of rainfall is a helpful tool to depict the rainfall frequency for different return periods throughout the study region. It is a map of the region where colours indicate the intensity of expected rainfall in different areas of the region.

To construct the map, we use gridded data of the Skåne region having 23,231 grid-cells. We estimate the MAMDR value \( \hat{l}_1 \) for each grid-cell using one of the best models mentioned in Section 3.7. The quantile estimate for each grid-cell can be computed using Equation (1) by replacing \( l_1 \) with \( \hat{l}_1 \). Finally, the values of grid-cells lie in different numerical ranges are represented by different colours.
RESULT AND DISCUSSION

The RRFA is performed using annual maximum daily rainfall (AMDR) series of 24 sites located in Skåne County, Sweden. The flowchart (Figure 2) describes the main steps involved in analysing the data. It is always crucial to review the appropriateness of data before applying any statistical technique.

The principal statistical assumptions of at-site and RFA include independence, stationarity and randomness of data series at each gauging site (see e.g., Ul Hassan et al. 2019; Hussain et al. 2017; Shahzadi et al. 2013 and Kousar et al. 2020). To investigate the serial correlation of the observed data series at each gauging site, we use the lag-1 correlation coefficient ($r$). The corresponding $p$ values of lag-1 correlation coefficients ($r$) indicate no serial correlation of data series at any given site, see Table 2. We use run test and Mann-Kendall test to examine the stationarity and randomness of the observed data series at each gauging site, respectively. The results of these tests are presented in Table 2. According to $p$ values of these tests, the data series at each gauging site holds the assumption of stationarity and randomness.

To examine the discordant site in the region, the discordancy measure $D_i$ is used which is based on L-moments. The summary statistics along with a discordancy measure value of each gauging site are presented in Table 3. All the sites’ $D_i$ values are less than the critical value three suggested by Hosking and Wallis (1997). This indicates that there is no discordant site in the region. The results of various tests at all sites of the Skåne region demonstrate that the data at each site is suitable for RFA.

The next step is to establish that these 24 study sites of Skåne can be considered as a single region. The heterogeneity measure (H) is calculated using the procedure in Section 3.3. To compute H, 500 simulations are carried out using four-parameter Kappa distribution. The estimated parameters of Kappa distribution used in the simulation are $\xi = 0.801$, $\alpha = 0.265$, $k = -0.090$ and $h = 0.165$. The value of H is $-0.10$ which is less than 1. So, the entire Skåne region based on these 24 sites can be seen as a homogeneous region. This suggests that the Skåne region is suitable for RFA. We expected this because of similar climatological and topography characteristics of the Skåne region.

The terrain of Skåne County is almost flat. The gauged sites in the region are within a small range of MAMDR, which is 30.083–40.789. This one climate characteristic is enough to constitute the boundary of the region. Many studies have formed homogeneous regions based on a small range of mean annual precipitations.

FIGURE 2 Flowchart summarized the main steps of the methodology [Colour figure can be viewed at wileyonlinelibrary.com]
If we look at the seasonality of the extremes AMDR in the region, overall 81% and at least 67% at each site extremes AMDR are in the period of July to October. A previous study about the regional analysis of short-duration rainfall extremes finds the south-west counties of Sweden as one homogeneous region, which includes Skåne (Olsson et al., 2019).

After identification of the homogeneous region, the next step in the RFA is to choose the best suitable frequency distribution for the region. In RFA, a single frequency distribution is fitted for all the sites in the region that provides the most accurate quantile estimates at each site in the region. To determine the most appropriate distribution for the region, we use the L-moment ratio diagram and $Z_{\text{Dist}}$ criterion. In the L-moment ratio diagram, the regional average L-skewness ($t^3$) and L-kurtosis ($t^4$) are plotted in Figure 3, represented with a red dot. This dot lies closest to the theoretical curve of the generalized normal (GNO) distribution. Therefore, the GNO distribution seems to be a well fitted distribution for the Skåne region. Table 4 provides $|Z_{\text{Dist}}|$ values of various three parameter distributions for the region. The value of $|Z_{\text{Dist}}|$ is less than 1.64 for the GNO and GEV distributions. So, these distributions are accepted as suitable candidate distributions for the region. Furthermore, we have a smaller value of $|Z_{\text{Dist}}|$ for GNO than GEV. Therefore, GNO is the most favourable and appropriate distribution for the region. It also validates the conclusion from the L-moment ratio diagram. We have used the R package lmomRFA developed by Hosking (2019) for L-moment analysis.

After a successful selection of the most appropriate regional frequency distribution, the next step is to estimate the regional quantiles for different return periods using the parameter estimates of GNO distribution in Table 4. The regional quantiles for different values of the non-exceedance probability ($F$) along with 90% error bounds and RMSE are presented in Table 5. The RMSE and 90% error bounds are obtained by using Monte Carlo simulation (Hosking and Wallis, 1997). The error bounds

| Table 2 Statistical assumptions of data series at each gauging site |
|---|
| S. No. | Station names | Station no. | $n$ | $r$ | $p$ value | Mann-Kendall test | Run test |
|---|---|---|---|---|---|---|---|
| 1 | Bromölla | 64,070 | 59 | −.111 | .045 | −.052 | .565 | 26 | .238 |
| 2 | Brösarp | 54,410 | 24 | −.250 | .250 | .007 | .980 | 15 | .404 |
| 3 | Falsterbo | 52,230 | 70 | −.043 | .726 | .049 | .550 | 37 | .810 |
| 4 | Gillastig | 63,010 | 58 | .245 | .066 | .077 | .398 | 24 | .127 |
| 5 | Hallands Väderö | 62,260 | 25 | .053 | .806 | −.104 | .483 | 15 | .534 |
| 6 | Helsingborg | 62,020 | 28 | −.005 | .982 | .024 | .874 | 15 | 1.000 |
| 7 | Hörby | 53,530 | 24 | .010 | .964 | .044 | .785 | 10 | .220 |
| 8 | Klippan | 63,070 | 62 | .095 | .467 | .120 | .172 | 28 | .306 |
| 9 | Kristianstad | 64,030 | 23 | .107 | .637 | −.020 | .916 | 14 | .515 |
| 10 | Landskrona | 52,520 | 60 | .141 | .288 | −.012 | .898 | 29 | .602 |
| 11 | Louisefred | 62,150 | 55 | .130 | .350 | .053 | .576 | 23 | .135 |
| 12 | Love Stad | 53,400 | 41 | .190 | .241 | −.104 | .345 | 21 | .877 |
| 13 | Lund | 53,430 | 59 | −.158 | .236 | .146 | .105 | 37 | .087 |
| 14 | Malmö | 53,360 | 70 | .048 | .692 | .073 | .378 | 33 | .470 |
| 15 | Norra Ströö | 64,080 | 53 | .021 | .882 | .058 | .544 | 32 | .185 |
| 16 | örkelljunga | 63,170 | 44 | .251 | .104 | .085 | .424 | 18 | .127 |
| 17 | Skillinge | 54,290 | 25 | .153 | .476 | −.003 | 1.000 | 15 | .534 |
| 18 | Stehag | 53,540 | 62 | .120 | .358 | .096 | .277 | 29 | .447 |
| 19 | Stiby | 54,340 | 22 | .036 | .875 | .039 | .822 | 11 | .662 |
| 20 | Tånga | 62,120 | 40 | .138 | .403 | .168 | .130 | 19 | .522 |
| 21 | Tomelilla | 53,320 | 62 | −.011 | .932 | −.014 | .874 | 30 | .608 |
| 22 | Trelleborg | 53,230 | 70 | .027 | .825 | .115 | .163 | 33 | .470 |
| 23 | Vomb | 53,410 | 67 | .125 | .317 | .129 | .124 | 34 | .903 |
| 24 | Ystad | 53,260 | 56 | .050 | .718 | .106 | .252 | 31 | .590 |

Note: $r$ represents the Pearson correlation coefficient and $n$ indicates the sample size of data series.
and RMSE indicates that there is more uncertainty around the regional quantiles estimate at high return periods. At-site quantiles can be estimated using Equation (1) in Section 3.5.

The quantile estimates at ungauged sites in the region are paramount to engineers, city planners, and scientist related to water management. To estimate the quantiles at ungauged sites, the MAMDR is needed on these sites since MAMDR cannot be computed at ungauged sites.
sites because of unavailability of data. Thus further modelling to estimate the MAMDR is required. For this purpose, we can use the relationship between the MAMDR at gauged sites \( l_1 \) in the region and their characteristics. In this study, three characteristics (longitude, latitude and elevation) of gauged sites are used. Unfortunately, these are the only characteristics available for each ungauged site in the region. This is also a limitation of the research. Therefore, only the best explanatory variables for the models among these characteristics to estimate the rainfall index for each ungauged site in the region are selected.

The correlation matrix of the site characteristics is

\[
M_{cor} = \begin{bmatrix}
\text{long} & \text{lat} & \text{Elevation} & l_1 \\
\text{long} & 1.00 & -0.32 & 0.17 & -0.11 \\
\text{lat} & -0.32 & 1.00 & 0.04 & 0.67 \\
\text{Elevation} & 0.17 & 0.04 & 1.00 & 0.37
\end{bmatrix}
\]

The correlation matrix exhibits that \( l_1 \) has a high correlation with latitude and elevation. Therefore, latitude and elevation are important variables to explain the variation in \( l_1 \). Figure 4 manifests that there is more rain in elevated areas. The MAMDR is increasing along the latitude. Figure 4 also validates the results from the correlation matrix. Therefore, these variables are used as explanatory variables for the models among these characteristics to estimate the rainfall index for each ungauged site in the region.

The machine learning models are data hungry and their predictive performance heavily relies on the size of the training data set. Many research studies demonstrate that it is hard to have a training data set with enough size. This study have used 20 observations to train the models. It is a limitation of this study. The problem of overfitting of models arises due to small sample sizes. To avoid overfitting and produce a model that have good predictive ability, simple models like (IDW and MLR), and models that resist overfitting (RF, SVM, GP) are used. Moreover, K-fold CV and relevant features of the models are used.

The 10-fold CV algorithm as described in Section 3.8 is used to identify the best model which has an adequate predictive performance. For 10-fold CV, we consider 80% of the observed data as training data and 20% as test data. The model that possesses the lowest RMSE, MAE and
PER value and the highest $R^2$ value is considered as the best predictive model. Accuracy measures in Table 6 indicate that the IDW method which does not use any explanatory variable performs about equally well as the MLR model. Based on the accuracy measures, the SVM model is chosen as the best predictive model. Therefore, this model is used to estimate the MAMDR at any ungauged sites ($l_j$) in the region. The range of estimated MAMDR for all ungauged sites is 30.731–40.984 which is close to the range of MAMDR for gauged sites. This shows that all ungauged sites should be part of this region. To construct the spatial map of the region the SVM model is used to estimate the rainfall index all over the Skåne region. Lastly, by using Equation (1), isopluvial precipitation maps of Skåne County (Figure 5) are generated for different return periods with non-exceedance probability $F$.

Isopluvial precipitation maps of Skåne County (Figure 5) for different return periods exhibit that the

**TABLE 6** 10-fold cross-validation for accuracy of models

| Accuracy measures | Interpolation models | IDW | MLR | RF | GPR | SVM |
|-------------------|----------------------|-----|-----|----|-----|-----|
| RMSE              |                      | 2.077 | 2.017 | 2.577 | 1.836 | 1.450 |
| SD(RMSE)          |                      | 1.079 | 0.935 | 0.752 | 0.747 | 0.886 |
| $R^2$             |                      | .693 | .699 | .633 | .722 | .866 |
| SD($R^2$)         |                      | 0.376 | 0.307 | 0.281 | 0.226 | 0.167 |
| MAE               |                      | 1.639 | 1.647 | 1.939 | 1.508 | 1.203 |
| SD(AMSE)          |                      | 0.790 | 0.621 | 0.577 | 0.551 | 0.709 |
| PER               |                      | 0.060 | 0.058 | 0.074 | 0.052 | 0.042 |
| SD(PER)           |                      | 0.031 | 0.026 | 0.021 | 0.021 | 0.025 |

**Figure 5** Spatial map of expected rain at different ungauged sites of Skåne for different return period. Solid dots indicate the location of gauging sites [Colour figure can be viewed at wileyonlinelibrary.com]
areas close to the south coast are expected to receive less rain while the northern parts of Skåne are expected to receive more rain. It is also observed that highly elevated areas receive more rain. Overall, there is an increasing trend of expected rain with increasing latitude.

5 | CONCLUSION

Daily rainfall data are prime information for urban drainage systems and management of flash flood. In this study, RFA is carried out by using annual maximum daily rainfall (AMDR) data for 24 gauging sites in the Skåne County, Sweden. None of the sites are discordant based on a discordancy measure. All the assumptions of RFA are satisfied for all rainfall gauging sites. The result of a L-moment based homogeneity measure reveals that the Skåne County is a homogeneous region. This result was expected due to similar climate and topography of the study area. The GOF methods indicate that GNO is the most suitable regional frequency distribution for this region.

The results manifest that latitude and elevation are important variables among the available set of site characteristics to explain the variation in the MAMDR at gauged site ($l_i$). Therefore, these variables are used in machine learning models as explanatory variables to predict the MAMDR at ungauged sites. To detect the best predicted model, K-fold cross validation is carried out. Accuracy measures indicate that the SVM model offers the best prediction performance. Therefore, the SVM model is used to estimate index rainfall (at-site MAMDR) efficiently in the Skåne County. Index rainfall combined with regional quantiles is used effectively to estimate the rainfall quantiles at ungauged sites for different return periods in the region. Finally, a spatial map of the expected AMDR for different return period is constructed. The map depicts that areas of high elevation receive more rainfall. There is an escalating trend of rainfall magnitude with increasing latitude.

The results found in this research study can be useful for management of flash flood, water resource planning, designing the urban drainage system, crop insurance and many other purposes related to water management activities in this region. The findings of this research could be used as guidelines for RFA of other regions in Sweden similar to Skåne County. The predictive model (SVM) for index rainfall is based on only two explanatory variables. Therefore, the prediction results of MAMDR can be improved by adding more sites characteristics, if available. This study is limited to daily rainfall data. One may consider sub-daily, monthly and seasonal data for further research in this region.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

ORCID

Mahmood Ul Hassan © https://orcid.org/0000-0003-2889-0263

REFERENCES

Al Mamoon, A., Joergensen, N.E., Rahman, A. and Qasem, H. (2014) Derivation of new design rainfall in Qatar using L-moment based index frequency approach. International Journal of Sustainable Built Environment, 3(1), 111–118.

Al Mamun, A., bin Salleh, M.N. and Noor, H.M. (2018) Estimation of short-duration rainfall intensity from daily rainfall values in Klang Valley, Malaysia. Applied Water Science, 8(7), 203.

Becerra, P.I. (2016) Relationship between climate and geographical variation of local woody species richness within the Mediterranean-type region of Chile. Revista Chilena de Historia Natural, 89(1), 12.

Chen, T., Ren, L., Yuan, F., Yang, X., Jiang, S., Tang, T., Liu, Y., Zhao, C. and Zhang, L. (2017) Comparison of spatial interpolation schemes for rainfall data and application in hydrological modeling. Water, 9(5), 342.

Dehghani, M., Saghaﬁan, B., Nasiri Saleh, F., Farokhnia, A. and Noori, R. (2014) Uncertainty analysis of streamflow drought forecast using artificial neural networks and Monte-Carlo simulation. International Journal of Climatology, 34(4), 1169–1180.

Desai, S. and Ouarda, T.B.M.J (2021) Regional hydrological frequency analysis at Ungauged sites with random Forest regression. Journal of Hydrology, 594, 125861.

Forestieri, A., Lo Conti, F., Blenkinsop, S., Cannarozzo, M., Fowler, H.J. and Noto, L.V. (2018) Regional frequency analysis of extreme rainfall in Sicily (Italy). International Journal of Climatology, 38, e696–e716.

Gentilucci, M., Bisci, C., Burt, P., Fazzini, M., & Vaccaro, C. (2018). Interpolation of Rainfall through Polynomial Regression in the Marche Region (Central Italy). Paper presented at the The Annual International Conference on Geographic Information Science.

Hailegeorgis, T.T. and Alfredsen, K. (2017) Regional flood frequency analysis and prediction in ungauged basins including estimation of major uncertainties for mid-Norway. Journal of Hydrology: Regional Studies, 9, 104–126.

Hastie, T., Tibshirani, R. and Friedman, J. (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer Science & Business Media.

Hong, B., Gasse, F., Uchida, M., Hong, Y., Leng, X., Shibata, Y., An, N., Zhu, Y. and Wang, Y. (2014) Increasing summer rainfall in arid eastern-Central Asia over the past 8500 years. Scientific Reports, 4, 5279.

Hosking, J.R.M. (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society: Series B, 52(1), 105–124.

Hosking, J. R. M. (2019). Regional frequency analysis using L-moments, lmomRFA R package, version 3.3. Available at https://cran.r-project.org/web/packages/lmomRFA/

Hosking, J.R.M. and Wallis, J.R. (1997) Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge, England: Cambridge university press.
Hussain, Z., Shahzad, M.N. and Abbas, K. (2017) Application of regional rainfall frequency analysis on seven sites of Sindh, Pakistan. *KSCE Journal of Civil Engineering*, 21(5), 1812–1819.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013) *An introduction to statistical learning* (Vol. 112). New York: Springer.

Johnston, K., Ver Hoef, J.M., Krivoruchko, K. and Lucas, N. (2001) Using ArcGIS geostatistical analyst. In: *Redlands*. CA, USA: ArcGIS Manual by ESRI.

Kousar, S., Khan, A.R., Ul Hassan, M., Noreen, Z. and Bhatti, S.H. (2020) Some best-fit probability distributions for at-site flood frequency analysis of the Ume River. *Journal of Flood Risk Management*, 13(3), e12640.

Lv, A. and Zhou, L. (2016) A rainfall model based on a geographically weighted regression algorithm for rainfall estimations over the arid Qaidam Basin in China. *Remote Sensing*, 8(4), 311.

Malekinezhad, H. and Zare-Garizi, A. (2014) Regional frequency analysis of daily rainfall extremes using L-moments approach. *Atmosfera*, 27(4), 411–427.

Modarres, R. (2008) Regional frequency distribution type of low flow in north of Iran by L-moments. *Water Resources Management*, 22(7), 823–841.

Naoum, S. and Tsanis, I.K. (2004a) A multiple linear regression GIS module using spatial variables to model orographic rainfall. *Journal of Hydroinformatics*, 6(1), 39–56.

Naoum, S. and Tsanis, I.K. (2004b) Orographic precipitation modeling with multiple linear regression. *Journal of Hydrologic Engineering*, 9(2), 79–102.

Olsson, J., Södling, J., Berg, P., Wern, L. and Ernon, A. (2019) Short-duration rainfall extremes in Sweden: a regional analysis. *Hydrology Research*, 50(3), 945–960.

Peel, M.C., Wang, Q.I., Vogel, R.M. and McMahon, T.A. (2001) The utility of L-moment ratio diagrams for selecting a regional probability distribution. *Hydrological Sciences Journal*, 46(1), 147–155.

Skåne Region. (2014). The open Skåne 2030-Skåne’s regional development strategy. Retrieved from https://utveckling.skane.se/siteassets/publikationer_dokument/rus_slutdokument_210x275_eng.pdf

Schaefer, M. G., Barker, B. L., Taylor, G. H., & Wallis, J. R. (2006). *Regional Precipitation-Frequency Analysis and Spatial Mapping of Precipitation for 23-Hour and 2-Hour Durations in Eastern Washington*. Retrieved from https://www.wsdot.wa.gov/research/reports/fullreports/640-1.pdf

Shahzadi, A., Akhter, A.S. and Saf, B. (2013) Regional frequency analysis of annual maximum rainfall in monsoon region of Pakistan using L-moments. *Pakistan Journal of Statistics and Operation Research*, 9, 111–136.

Radio Sweden. (2014). *Skåne hit by flash-floods*. Retrieved from https://sverigesradio.se/sida/artikel.aspx?programid=2054&artikelid=5951748

Szolgy, J., Parajka, J., Kohnová, S. and Hlavčová, K. (2009) Comparison of mapping approaches of design annual maximum daily precipitation. *Atmospheric Research*, 92(3), 289–307.

Teodoro, P.E., da Cunha, E.R., Corrêa, C.C.G., Ribeiro, L.P., Torres, F.E., de Oliveira-Junior, J.F., Gois, G. and Bacani, V.M. (2016) Altitude and geographic coordinates to estimate monthly rainfall in the state of Mato Grosso do Sul. *Bioscience Journal*, 32(1), 41–47.

Thapa, S., Zhao, Z., Li, B., Lu, L., Fu, D., Shi, X., Tang, B. and Qi, H. (2020) Snowmelt-driven Streamflow prediction using machine learning techniques (LSTM, NARX, GPR, and SVR). *Water*, 12(6), 1734.

The Local. (2014). Flash floods cause chaos in southern Sweden. Retrieved from https://www.thelocal.se/20140831/flash-floods-cause-chaos-in-southern-sweden

Ul Hassan, M., Hayat, O. and Noreen, Z. (2019) Selecting the best probability distribution for at-site flood frequency analysis; a study of Torne River. *SN Applied Sciences*, 1(12), 1629.

Urn, M.-J., Yun, H., Cho, W. and Heo, J.-H. (2010) Analysis of orographic precipitation on Jeju-Island using regional frequency analysis and regression. *Water Resources Management*, 24(7), 1461–1487.

Vapnik, V. (1995). *The nature of statistical learning theory*. New York: Springer.

Wallis, J.R., Schaefer, M.G., Barker, B.L. and Taylor, G.H. (2007) Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington state. *Hydrology and Earth System Sciences Discussions*, 11(1), 415–442. Available at https://hal.archives-ouvertes.fr/hal-00305628.

Water Encyclopedia. (2021). Precipitation, Global Distribution of. Retrieved from http://www.waterencyclopedia.com/Po-Re/Precipitation-Global-Distribution-of.html

Yu, P.-S., Chen, S.-T. and Chang, I.-F. (2006) Support vector regression for real-time flood stage forecasting. *Journal of Hydrology*, 328(3–4), 704–716.

**How to cite this article:** Ul Hassan M, Noreen Z, Ahmed R. Regional frequency analysis of annual daily rainfall maxima in Skåne, Sweden. *Int J Climatol.* 2021;41:4307–4320. [https://doi.org/10.1002/joc.7074](https://doi.org/10.1002/joc.7074)