

Geophysical Research Letters



RESEARCH LETTER

10.1029/2020GL091933

Key Points:

- Rainfall series are extracted and processed from more than 11,000 regions of interest over the last 100 years in Great Britain and Australia.
- Spatial distribution of extreme rainfall modeled qualitatively and quantitatively shows impact from location, size, and shape of regions.
- Methods and findings provide new perspectives for understanding the heterogeneous nature of climate variability and climatic change impact.

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

Y. Xuan,
y.xuan@swansea.ac.uk

Citation:

Wang, H., & Xuan, Y. (2021). Spatial variation of extreme rainfall observed from two century-long datasets. *Geophysical Research Letters*, 48, e2020GL091933. <https://doi.org/10.1029/2020GL091933>

Received 3 DEC 2020
Accepted 10 MAR 2021

Spatial Variation of Extreme Rainfall Observed From Two Century-Long Datasets

H. Wang¹  and Y. Xuan¹ 

¹College of Engineering, Swansea University Bay Campus, Swansea, UK

Abstract This paper presents the spatial variation of area-orientated annual maximum daily rainfall (AMDR), represented by well-fitted generalized extreme value (GEV) distributions, over the last century in Great Britain (GB) and Australia (AU) with respect to three spatial properties: geographic locations, sizes, and shapes of the region-of-interest (ROI). The results show that the spatial variation of GEV location-scale parameters is dominated by geographic locations and area sizes. In GB, there is an eastward-decreasing banded pattern compared with a concentrically increasing pattern from the middle to coasts in AU. The parameters tend to decrease with increased area sizes in both studied regions. Although the impact of the ROI shapes is insignificant, the round-shaped regions usually have higher-valued parameters than the elongated ones. These findings provide a new perspective to understand the heterogeneity of extreme rainfall distribution over space driven by the complex interactions between climate, geographical features, and the practical sampling approaches.

1. Introduction

Intensive rainfall is considered to be one of the primary triggers for flooding alongside other factors, such as climate, topography, and soil type of different catchment patterns (Rogger et al., 2017; Westra et al., 2014). Evaluation of the flood risks caused by rainfall usually requires use of long-term observed data at one or more locations to derive flood-triggering rainfall amounts with preferred exceedance probabilities. This procedure is always associated with a region-of-interest (ROI). While the precipitation process is part of the global hydrological cycle and hence a (laterally) boundless phenomenon, its area-oriented variation is of the concern of the engineers and flood risk managers. The area-oriented rainfall variation and distribution are closely related to the climate at large scale (Millán et al., 2005); in the meantime, local features and processes, such as the topography, urbanization, as well as the orientation and the size of the area can also affect the rainfall amount (Buytaert et al., 2006). Many studies, for example, Buishand et al. (2008), Jung et al. (2017), Pedersen et al. (2010), and Zheng et al. (2016), have attempted to understand the spatial variation of rainfall extremes at different scales based on gauge records.

Moreover, spatially disaggregated, grid-based hydroclimatic datasets have become increasingly accessible, motivating the research community to process large-scale variation analysis of grid-by-grid extremes with these new datasets (Peleg et al., 2018), for example, analyzing the spatial variation of grid-based hydroclimatic observations (UKCIP, Banwell et al., 2018; E. J. Kendon, Fosser, et al., 2019; M. Kendon, McCarthy, et al., 2019; Lowe et al., 2018; Prein et al., 2017) and the temporal variation of hydroclimatic extremes, such as frequency analysis (Li et al., 2015; Overeem et al., 2010).

However, these efforts are often frustrated by the fact that the required data records with sufficient length are often scarce. It is unsurprising that very few studies have been produced so far. Furthermore, most aforementioned studies focus only on the spatio-temporal variation of averaged quantities of hydro-climatic variables instead of their extremes; for those indeed focusing on extremes, they tend to be limited by one or a few catchments or stations.

In this study, two century-long, grid-based rainfall datasets covering Great Britain (GB) and Australia (AU) with relatively high temporal and spatial resolutions (daily and 1–5 km, respectively) are analyzed. The overall aim is to gain insights into how area-orientated rainfall extremes vary with space with respect to the probability distribution parameters which are of concern of flood risk management (FRM) and civil engineering design. Specifically, this study attempts to address the following questions:

© 2021. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

1. How do areal rainfall extremes change over space?
2. How may other factors, such as the size, shapes of the area in question affect such spatial dependencies?
3. How are the spatial patterns and variations linked to the large-scale climatology of rainfall?
4. What is the implication of the spatial variation of the parameters to the applications (e.g., FRM)?

Additionally, a toolbox known as the spatial random sampling for grid-based data analysis (SRS-GDA, Wang & Xuan, 2020), is employed to assist the required spatial sampling with predefined or randomized features, for example, size, location, and dominant orientation of ROIs, from the grid-based datasets. The sampled annual maximum daily rainfall (AMDR) of each ROI is fitted with the widely used and tested Generalized Extreme Value (GEV) distributions whose spatial variation is then analyzed. The associated intensive computation demand is met by the high-performance computing (HPC) resources provided by Supercomputing Wales (<https://www.supercomputing.wales>).

The remainder of this paper is organized as follows: Section 2 describes the data and methods, then the sampled ROI and the goodness-of-fit (GOF) results. Both the qualitative and quantitative results of the spatial variation of the distribution parameters and further application are discussed in Sections 3 and 4, as well as their linkage to the climatology of rainfall are discussed in Section 5. The conclusions and recommendations of further study are given in Section 6.

2. Data and Methods

2.1. Datasets

This study makes use of two century-long datasets which are the “gridded estimates of daily areal rainfall” (GEAR) and the “Australian Data Archive for Meteorology” (ADAM). The GEAR dataset is a grid-based ($1 \times 1 \text{ km}^2$) rainfall estimation that covers the mainland of GB from 01/01/1898 to 31/12/2010. It is derived from the UK Met Office national database of observed precipitation from the UK rain gauge network, using the natural neighbor interpolation method. The coordinates are in the National Grid Reference (Ordnance Survey, 1946) which is a projected map coordinate system with the easting (x -) and northing (y -) expressed in linear kilometers (Tanguy et al., 2016). The ADAM data set is generated using a sophisticated analysis technique described in Jones et al. (2009), which is also grid-based ($0.05 \times 0.05^\circ$, $\sim 5 \times 5 \text{ km}^2$) rainfall from January 1, 1900 to December 31, 2018 over AU based on the Geocentric Datum of AU 1994 (GDA94; Collier, 2002) with the origin (44 S, 112 E), that is (0,0), and easting (x -) and northing (y -) transformed to kilometers. The recorded rainfall values are provided as daily rainfall, that is, the total rainfall amount over a predefined 24-h (9:00–9:00 a.m.) period which refers to the 24 h prior to the reporting time for the ADAM data set and the 24 h after for the GEAR data set.

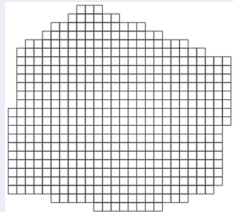
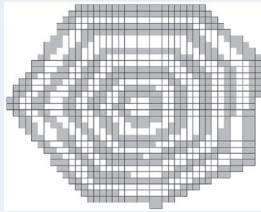
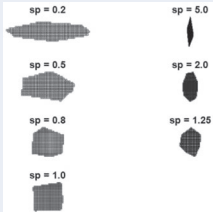
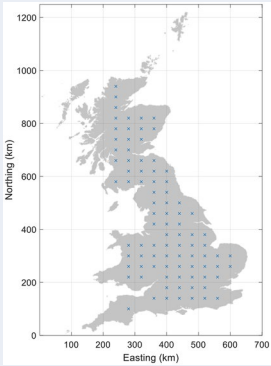
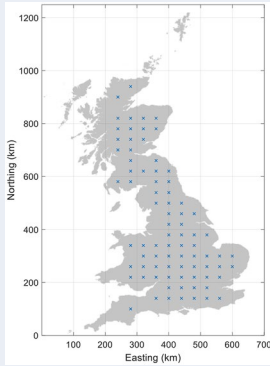
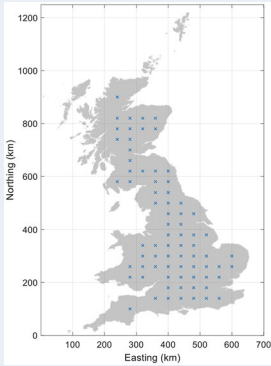
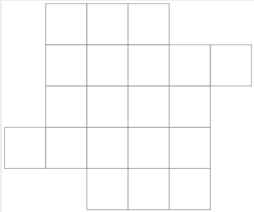
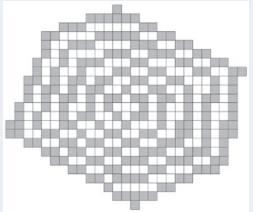
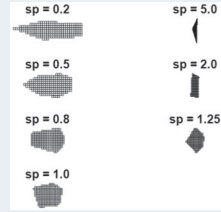
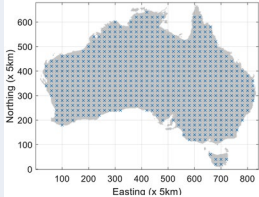
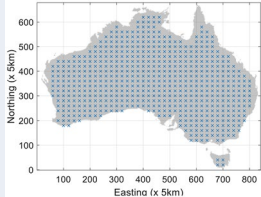
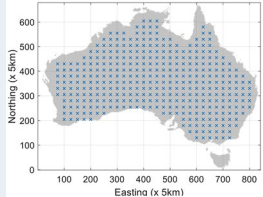
2.2. Methodology

The geographical areas of the two data domains, i.e., GB and AU, are sampled into a series of regions-of-interest (ROIs) using the SRS-GDA toolbox. Three predefined spatial features (i.e., geographical locations, sizes, and shapes) are applied during the sampling process to reduce the overall computing time while maintaining the representativeness of the samples. The block maxima series, that is, the AMDR of each sampled ROI is then extracted and further fitted by a proper probability distribution. In this study, the three-parameter GEV distribution is chosen as the candidate distribution whose GOF is evaluated. The parameters (μ and σ) of the fitted distributions are then analyzed with regards to their spatial distribution referring to the large-scale climatology of rainfall variations.

2.2.1. ROI Generation and AMDR Extraction

The sampling starts with an initial set of uniformly distributed ROIs whose locations are represented by the coordinates of their geometric centroids. At each location, there are seven ROI shapes produced, indicated by their distinctive *spatial indexes* (Wang & Xuan, 2020) and reciprocally grouped as 0.2/5.0, 0.5/2.0, 0.8/1.25, and 1.0, respectively. The size of these ROIs is then gradually increased by 10 steps with 20% increment each, while maintaining the same shape and location. In the end, the largest sizes the ROIs are 1,050 km^2 for GB and 9,900 km^2 for AU, respectively.

Table 1
ROIs for Analyzing the Spatial Variations in GB and AU

Sampling areas		Changing with location	Changing with size (each group includes 10 ROIs)	Changing with shape (each group includes 7 ROIs)
GB	Indicators	G1	G2	G3
	ROI(s) of $1 \times 1 \text{ km}^2$ grid			
	Size (km^2)	500	10, 43, 87, 164, 257, 366, 504, 660, 827, 1025	500 each ROI
	Total ROI number	88	$81 \times 10 = 810$	$74 \times 7 = 518$
	Total meridional group number	10	10	10
Geographical location (marked as "x")				
AU	Indicators	A1	A2	A3
	ROI(s) of $5 \times 5 \text{ km}^2$ grid			
	Size (km^2)	500	125, 400, 900, 1,550, 2,450, 3,550, 4,875, 6,350, 8,025, and 9,900	5,000 each ROI
	Total ROI number	679	$627 \times 10 = 6270$	$378 \times 7 = 2646$
	Total meridional group number	40	38	30
Geographical location (marked as "x")				

Abbreviations: AU, Australia; GB, Great Britain; ROI, region-of-interest.

The SRS-GDA toolbox used to generate the ROIs is set up in a way that only one spatial feature is allowed to vary at a time. For instance, to obtain ROI samples of G2 and A2 in Table 1, the centroid locations and shape are kept unchanged while generating 10 ROIs in different sizes. Table 1 summarizes all types of ROIs and their properties.

For each ROI, its daily areal rainfall is calculated by taking arithmetic average over the covered grids (e.g., spatial average); then the block annual maxima are picked up to generate a time series of AMDR. This work was carried out by the HPC, processing around 642.3 GB data with 11,011 areal AMDR series produced.

2.2.2. Fitting the Extracted AMDRs Using GEV Distribution

The GEV distribution is one of the most well-founded distributions for describing normalized maxima from a sequence of independent, identically distributed random variables, for example, the block maxima of annual rainfall series. It has been applied to characterize both gauged (Feng et al., 2007; Westra et al., 2013) and grid rainfall datasets (Overeem et al., 2010). A GEV distribution is controlled by three parameters, namely, the location μ , the scale σ , and the shape ξ which defines the three limiting types: the Gumbel ($\xi = 0$), the Frechét ($\xi > 0$), and the reversed Weibull ($\xi < 0$). AMDR (denoted as x) of each ROI is fitted by a GEV distribution whose cumulative probability distribution function is as follows:

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

where F is defined for $1 + \frac{\xi(x - \mu)}{\sigma} > 0$, $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$. A maximum likelihood estimator (Hosking, 1985) is employed to estimate the parameters of GEV.

However, the GEV distribution has been chiefly used to fit point rainfall extremes as reported in many studies (Schaefer, 1990; Yoon et al., 2013), very few studies have been done on the suitability of fitting the areal grid-based rainfall extremes with GEV. Thus, in this study, the GOF is tested using a bootstrapped version of Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests, and the L-moment ratio diagrams (Text S1). Out of the AMDRs from all ROIs (1,416 ROIs of GB and 9,595 of AU) tested, the results show that the GEV distribution fits well the AMDR series with a 100% pass of the KS test and more than 97% of the AD test. This is also supported by the L-moment ratio diagrams (Figure S1c) which demonstrates the fitted GEV distributions compared with the statistical characteristics of the AMDR itself.

2.2.3. Analyzing the Spatial Distribution of the Location-Scale Parameters

The spatial variation of the location and scale parameters of the fitted GEV distributions are analyzed both qualitatively and quantitatively. Instead of using full spatial coordinates to represent the geographical locations, a univariate spatial-location representation is adopted in this study. This procedure is briefly described below:

1. The chosen GEV parameter is aggregated meridionally, for example, over all ROIs that have the same x -direction (easting or longitude) coordinate.
2. The aggregated GEV parameter values are indexed by their x -direction only coordinates which are then used as an input variable to represent the geographical locations.
3. The same procedure is also applied zonally, that is, over the same y -direction coordinate.

With this arrangement, the meridional or zonal average of the GEV parameter in question is taken as the response variable. In AU, a concentric pattern is found where both the meridional and zonal average show similar results (Figure S6); For the GB case, only a strong west-east pattern exists. Therefore, for the convenience of comparison, the meridional average is chosen for both cases. The spatial variation of ξ is less significant over space thus is not considered (Figure 1 and S7).

Finally, a generalized linear model (GLM) is fitted to quantify the relationship between the GEV parameters and the associated spatial features of the ROIs, that is, to explicitly model the spatial variation of the GEV parameters with respect to the locations, sizes, and shapes of the underlying ROIs. A k -fold-cross-validation (Efron & Tibshirani, 1997) and variogram (Cressie & Hawkins, 1980) are employed to evaluate the performance of GLMs.

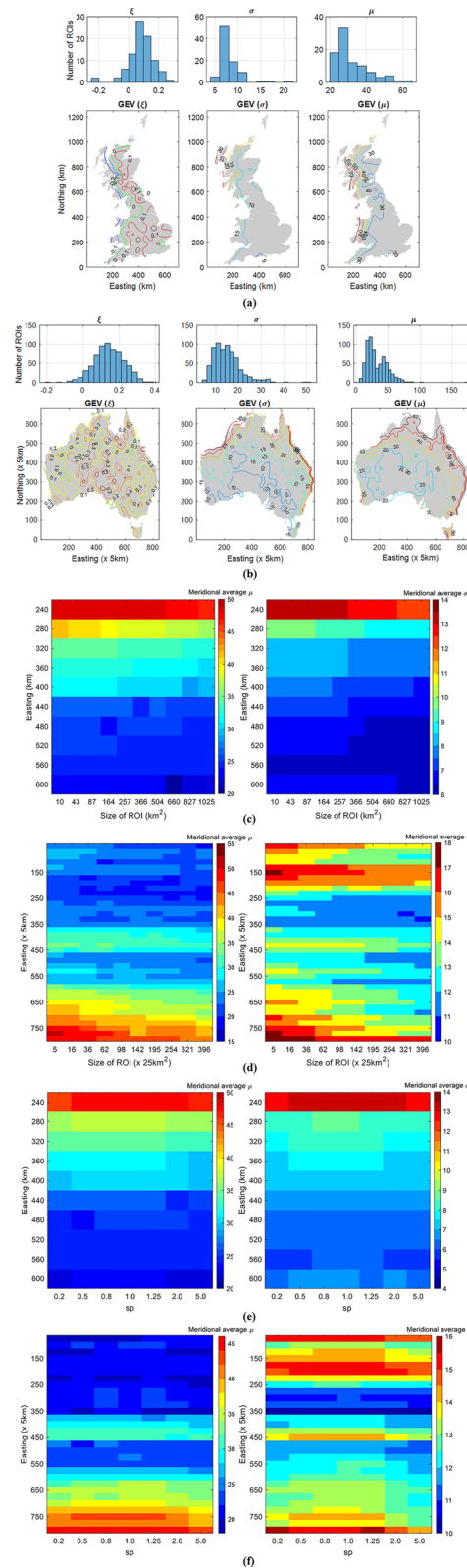


Figure 1. Histograms and spatial variations of the three GEV parameters in GB (a) and AU (b); the changes of meridional average μ and σ with the change of ROIs' size in GB (c) and AU (d) where the color scale indicates that majority of ROIs in both cases are decreased with the increase of ROI sizes; and the changes of meridional average μ and σ with the change of ROIs' shapes sp in GB (e) and AU (f) where a symmetric pattern in both cases is observed but the changes with respect to ROI shape is not significant. AU, Australia; GB, Great Britain; GEV, generalized extreme value; ROI, region-of-interest.

3. Results and Discussion

3.1. GEV Parameter Variation Over Geographical Locations

Figure 1a and 1b present the histograms and spatial variations of the three GEV parameters of all ROIs in both GB and AU where the following patterns can be clearly identified:

- Most ROIs are in favor of the Frechét type of distribution ($\xi > 0$).
- Both σ and μ present a similar spatial pattern where a higher μ is usually accompanied by a higher σ .

In GB, the values of μ and σ in the western region, especially those in the coastal area, are much larger than those in the east. Such west–east gradient is also strong in the west, indicated by much denser contours. However, there is no remarkable variation from south to north, even though both μ and σ in Scotland are higher. As such, the meridional average is thought to better reveal such eastward pattern which can be described as “west high, east low” with an apparently nonlinear variation.

Both μ and σ in AU have a clear increasing trend from the south-middle zone to the coastal regions. This spatial pattern can be seen as a series of concentric circles. It is also noted that the rapid variations are close to the north-eastern coastal regions.

3.2. Variation of GEV Parameters With Regards to the Area Size

Figure 1c and 1d show the changes of μ and σ of all meridional groups in GB and AU, parameterized by the size of the ROI (s , km²). Generally, regardless of their locations, the parameter values are inversely proportional to the ROI sizes.

The decreases in both μ and σ with increased ROI sizes have an important implication: the most frequent AMDR (relating to μ) becomes smaller for larger ROI alongside an overall decreased extremity (relating to both parameters). Another interesting measure is the rate of such reduction (RR) as the size of ROI increases, which has also shown a clear spatial dependency. In AU, the RR remains low in the central desert zone (e.g., from Easting 300 to 360 km), and increases near the coastal areas where large parameter values are also found. This feature can be explained by the fact that regions receiving more extreme rainfall (e.g., the coastal regions in AU) are not only manifested by the higher μ and σ ; but also have more heterogenous rainfall than regions with less extreme rainfall (lower μ and σ). Therefore, the changes of μ and σ are more sensitive to geographic locations, as revealed by the RR. GB also shows a similar pattern albeit not as remarkable.

3.3. Variation of GEV Parameters Due to the Area Shape

Figure 1e and 1f present the changes of μ and σ of all meridional groups in GB and AU, parameterized by the ROI shape (sp). The variation of the shape starts from west–east orientated ($sp = 0.2$), gradually growing into more rounded ($sp = 1.0$), then to north–south orientated ($sp = 5.0$) ones. Two shapes with reciprocal sp values will have their major dimension swapped, that is, east–west versus south–north and vice versa. The result is inspected and summarized as follows:

- For the majority of the meridional groups, there is little difference between the location-scale parameters of ROIs with reciprocal shapes, for example, two shapes with sp values of 0.2 and 5.0. This is regarded as a symmetric pattern around $sp = 1.0$.
- Generally, μ and σ of ROIs in an elongated shape are smaller than those of the ROIs in more rounded shapes. This indicates that the rounded-shape ROIs have a better chance to capture more rainfall extremes than the elongated ones. It also leads to that for the same area size, regions with more regular shape tend to have more extreme areal rainfall.
- Overall, the effects of ROI shape are insignificant.

4. Quantification of Spatial Variation

Generalized Linear Models (GLMs) are based on an extension to the classical linear regression model, having been widely applied in hydrometeorology (Coe & Stern, 1982; Stern & Coe, 1984) and shown to be effective in terms of incorporating complex structures (Segond et al., 2006). Since Chandler and Wheeler (2002) proposed a GLM-based framework for interpreting historical daily rainfall records and revealing the changes on rainfall occurrence and amount in western Ireland, many more applications have followed, for example, Yan et al. (2002), Yang et al. (2005), and Rashid et al. (2013) with good performance reported.

As the two meridionally averaged parameters μ and σ which reflect the property of rainfall extremes are shown to follow similar right-skewed gamma distributions (Figure S2a), we broadly followed Chandler and Wheeler (2002) and proposed a GLM with a log-link to quantify the spatial variation. The fitting starts from a simplest prediction form, and then adds other predictors or their combinations successively, for example, $x \times s$, $s \times sp$, or $x \times sp$ (James, 2002). To determine whether to keep the newly added predictor/combination, the significance of the new term is evaluated by calculating the log-likelihood at the significant level of 0.05 (detailed in Text S2). Finally, an optimal form of the GLMs is obtained by considering both the log-likelihood and the discrepancy (e.g., root-mean-squared-error) and identified as:

$$\text{For } \mu_{\text{GB}} : \left(1 + x + s + x^2 + x^3\right) \beta_{\mu_{\text{GB}}} \quad (2a)$$

$$\text{For } \sigma_{\text{GB}} : \left(1 + x + s + x^2 + x^3\right) \beta_{\sigma_{\text{GB}}} \quad (2b)$$

$$\text{For } \mu_{\text{AU}} : \left(1 + x + s + sp + x^2 + s^2\right) \beta_{\mu_{\text{AU}}} \quad (3a)$$

$$\text{For } \sigma_{\text{AU}} : \left(1 + x + s + sp + x^2 + s^2 + x \times s\right) \beta_{\sigma_{\text{AU}}} \quad (3b)$$

where the subscripts of β refer to the study case in question. A maximum likelihood estimator (McCullagh, 2018) was employed to obtain β (see the estimations in Table S1). These fitted GLMs are visualized in Figure 2a and 2b which show the following intriguing spatial features:

1. In GB, both the meridionally averaged μ and σ have a nonlinearly eastward-decreasing pattern. The RR is stronger in the west ($\Delta\mu_{\text{GB}} = -0.22 / \text{km}$; $\Delta\sigma_{\text{GB}} = -7.50 / 100\text{km}$) and gradually reduces to the east ($\Delta\mu_{\text{GB}} = -0.02 / \text{km}$; $\Delta\sigma_{\text{GB}} = -0.30 / 100\text{km}$). An almost-linearly decreasing pattern with increase of ROI size ($\Delta\mu_{\text{GB}} = -0.02 / 100\text{km}^2$ and $\Delta\sigma_{\text{GB}} = -0.08 / 100\text{km}^2$) is shown and such RRs are not affected by the geographic locations. However, they do not appear to be dependent on the ROI shape (sp)
2. In AU, the spatial changes of the meridionally averaged μ and σ are nonlinear with respect to both the ROI location and size. The parameter μ diminishes slowly from west coasts to south-middle zone ($\Delta\mu_{\text{AU}} = -0.05 / 500\text{km}$) then increases faster and faster to the east coasts (largest $\Delta\mu_{\text{AU}} = +6.70 / 500\text{km}$); The parameter σ shows a “center low, outer (coastal regions) high” pattern where the changing rate near the center ($|\Delta\sigma_{\text{AU}}| = 0.05 / 500\text{km}$) is lower than the coastal regions ($|\Delta\sigma_{\text{AU}}| = 1.40 / 500\text{km}$). Unlike in GB, sp plays a more significant role and diminishes both parameters ($\Delta\mu_{\text{AU}} = -0.9, \Delta\sigma_{\text{AU}} = -0.5$) when the ROI shape changes from the west–east-orientated ($sp = 0.2, 0.5$) to north–south-orientated ($sp = 5.0, 2.0$). It means that the north–south variation in AU is in general smaller than that of the east–west direction. Besides, the combined term ($x \times s$) is significant, which means that the RR of σ with respect to ROI size varies at different geographic locations. It is manifested by the uneven vertical gaps between contours in the right panel of Figure 2b

The performance of the GLMs is evaluated by comparing the parameter values modeled by the GLMs and those from the originally fitted GEVs (Figure 2c), where in the GB case there are slight underestimates for some large values that appear in the western coastal region; and in the AU case, some overestimates happen for the small values which are located in the middle-south dry zone. The GLM model probability structure is

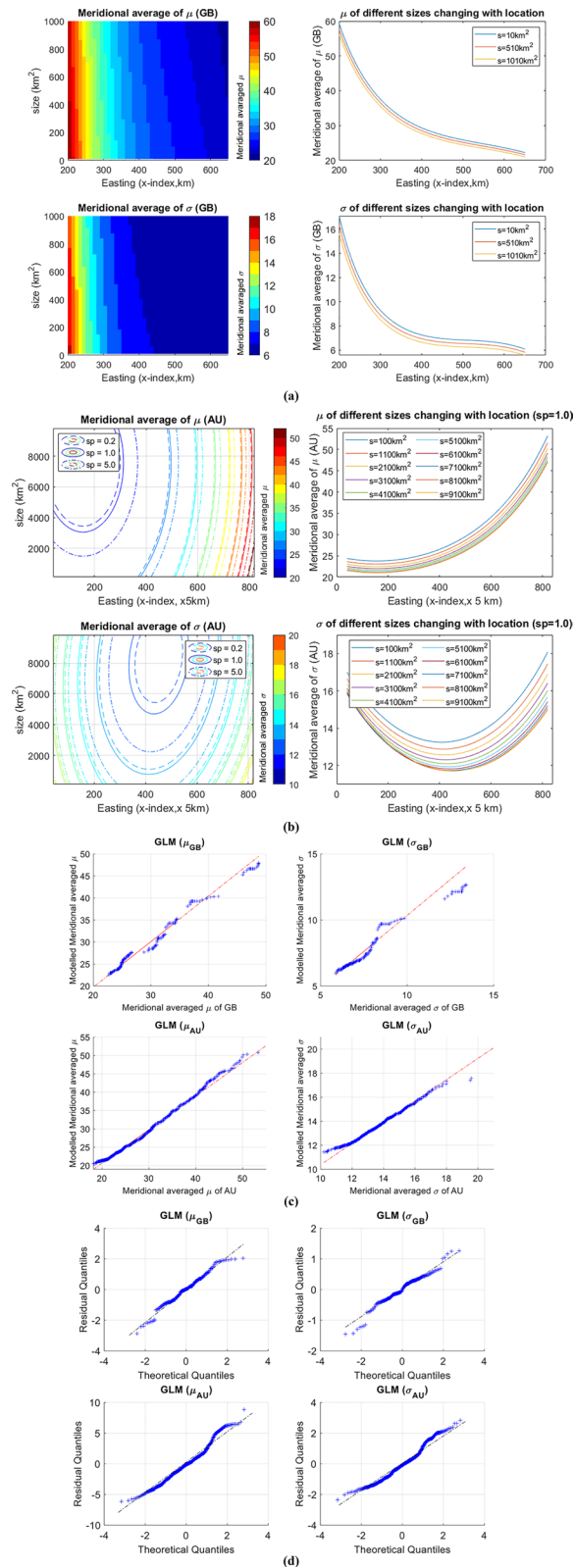


Figure 2. Visualization of the GLMs fitted to the meridional average GEV μ and σ parameters as a color-scale plot for GB (a) and a contour plot for AU (b) whose contours are picked up at the same stops of the values and their changes with respect to the geographic location and size; and a scatter plot (c) and a normal quantile plot (d) for revealing the difference between the actual GEV parameters and the modeled GEV parameters. AU, Australia; GB, Great Britain; GEV, generalized extreme value; GLM, generalized linear model.

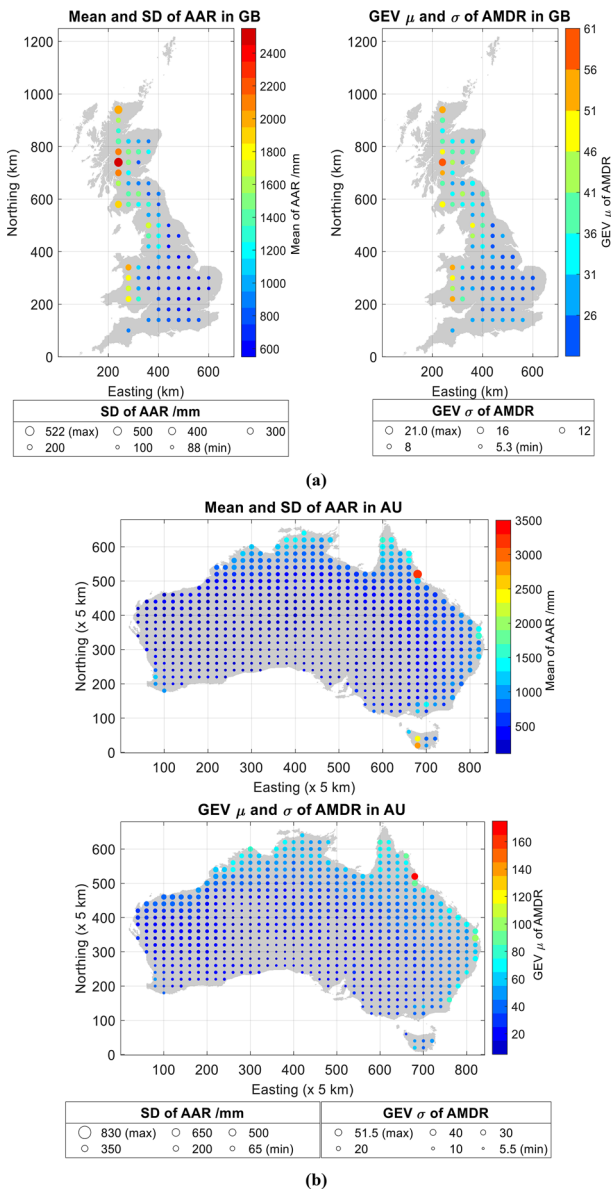


Figure 3. Comparison between the mean and standard deviation (SD) of areal annual rainfall (AAR) and GEV parameters μ and σ of annual maximum daily rainfall (AMDR) in GB (a) and AU (b) cases where the color denotes the values of mean of AAR or the GEV parameter μ of AMDR, and the size of the circles denotes the value of standard deviation of AAR or the GEV parameter σ of AMDR. AU, Australia; GB, Great Britain; GEV, generalized extreme value.

space between the daily maxima time series (i.e., the AMDR) and the cumulative annual rainfall (i.e., the ARR). For example, regions with higher mean of AAR are not only represented by a higher SD (e.g., the circles located in west Scotland and west Wales of GB and in north-eastern coastal regions of AU, appearing more reddish and larger), they are also associated with higher GEV parameters of AMDR, and appear to be more heterogenous. This feature also exists in regions with low and more even annual rainfall distribution, but works in an opposite way (e.g., circles located in middle and eastern England of GB and middle-north zone of AU are all more bluish and smaller). These findings are consistent with those published in the series of climate reports of GB (M. Kendon et al., 2015, 2018, E. J. Kendon, Fosser, et al., 2019, M.

checked by a residual analysis (McCullagh, 2018; Pierce & Schafer, 1986; Wang, 1987) whose results are shown in Figure 2d with a theoretical normal distribution shown on the x -axis and the residual quantiles on the y -axis. If the probability assumption (i.e., gamma assumption) is correct, all residuals would have the same distribution which is an approximate normal distribution. The distribution of the residuals of the four GLMs is symmetric with two flat sides. Generally, the approximation fits well except for the upper side which represents only 0.9% of the total data points. In view of the research aims, this is considered to be acceptable. The locative continuity in the adjacent ROIs was compared using variograms (see Text S4) which show highly similar spatial correlations in the parameters modeled by the GLM and those from the originally fitted GEVs for both the GB and AU cases. The prediction skill of the fitted GLMs is tested using a k -fold-cross-validation ($k = 10$) method which produces average Nash-Sutcliffe efficiency (NSE) coefficients across all random partitions of the four GLMs as 0.97, 0.89, 0.86, and 0.62 (see Text S3).

These quantitative findings regarding the spatial variation of the GEV parameters have two important implications to many downstream applications of the areal rainfall maxima, for example, FRM. For one aspect, the traditional approach in FRM makes use of point rainfall maxima to represent the areal one (catchment or a predefined area), where a scaling factor is involved. This simplistic treatment ignores the complexity in spatial distribution, nor can it account for the interplay of the size, location as well as shape of the area in question, as revealed above. For another, the overall quantification of the spatial variation of the GEV parameters (hence, the return values) makes it possible to study the FRM at country level as a single entity instead of looking at individual regions with isolations. It also helps to gain insights into how large scale hydroclimatic (rainfall) variation can affect the local FRM, which is very important for studying FRM under climate change impact.

5. Link to the Large-Scale Climatology of Rainfall Variation

The GEV distribution parameters that can reveal the characteristics of extreme rainfall in terms of both its amount and occurrence probability, are shown to have a strong spatial dependency as discussed in the previous sections. To understand how such spatial variation of the extreme rainfall is related to the climatology of rainfall variation, areal annual rainfall (AAR) time series from each ROI were obtained. The mean and standard deviation (SD) of the AAR series were then compared with the GEV parameters μ and σ of the AMDR series extracted from the same ROIs. To visualize the link, the spatial continuity of the corresponding parameters from both AAR and AMDR were represented by their variograms (see Figure S5) which show very little difference on locative continuity. Figure 3 demonstrates a great deal of similarity existing in

Kendon, McCarthy, et al., 2019) and AU (CSIRO & Australian Bureau of Meteorology, 2018; Frederiksen et al., 2014) and further discussed in Text S5.

6. Conclusions

This paper presents a study on the spatial variation of extreme rainfall using two-century long datasets covering GB and AU. The AMDR series extracted from regions-of-interest (ROI, 11,011 in total) with various spatial properties (location, size, and shape), are individually fitted with GEV distributions whose parameters are then analyzed over the space. Four GLMs are developed to quantify these variations by involving the effect from the geographical location, area size, and shapes. From the results discussed previously, the following conclusions can be drawn:

1. The GEV distributions are shown to be able to model well the grid-based areal-AMDR for both GB and AU; more than 90% of the regions are better fitted with the Fréchet distribution among the three GEV types.
2. The GEV location (μ) and scale (σ) parameters present similar spatial patterns where a higher μ is usually accompanied by a higher σ indicating those regions that have higher amount of most frequent rainfall often observe a higher occurrence probability of extremes.
3. Geographic location is the most significant factor affecting the two GEV parameters. The spatial pattern in GB is an eastward-decreasing-banded pattern with no significant difference along north–south direction. In AU, a concentrically increasing pattern from middle-south zone to north–east coasts is found.
4. Increasing the region size will decrease both parameters which means a decrease of the most frequent AMDR amount and the occurrence probability of extremes. However, in AU, the rate of such decrease varies with regions as the combined impact of ROI location and size is also detected to be significant.
5. Compared with other spatial properties, the shape of ROI is detected as insignificant, even though, a symmetric pattern is found for regions with reciprocal spatial indexes. Also, regions of more elongated shapes tend to have small parameter values in contrast with those having regular/rounded shapes.

These findings offer a new quantitative insight in understanding the spatial variation of large-scale climatology of rainfall. The quantification of the extreme rainfall and its spatial dependencies are of great practical value in engineering design, for example, designed rainfall/floods for constructions. The methods employed in this study are specifically designed for large grid-based datasets, and thus can be readily applied to climate projections for evaluating the spatial heterogeneity of climate change impact, such as flooding and droughts. It should be noted that the quality of the underlying datasets, which have undergone a series of quality control measures, may still bring in large amount of uncertainties and should be addressed in further work. Additionally, impact of the density of the underlying data observations, that is, rain-gauges, and its variation over long term also need to be further studied.

Acknowledgments

The authors would like to thank the Center of Hydrology and Ecology (CEH) and The Bureau of Meteorology, AU for providing the datasets, which are available in public domain online at <https://doi.org/10.5285/33604ea0-c238-4488-813d-0ad9ab7c51ca> for the GEAR data set and <http://www.bom.gov.au/jsp/awap/rain/index.jsp> for the ADAM data set. This research is supported by the Chinese Scholarship Council and the College of Engineering, Swansea University, UK via their PhD scholarships offered to the co-author Han Wang and the Academy of Medical Sciences GCRF Networking Grant (REF: GCRFNGR4_1165) which are gratefully acknowledged. We would also thank the three anonymous reviewers for their constructive comments that helped us to improve the quality of this manuscript.

Data Availability Statement

The open-source toolbox of spatial random sampling for grid-based data analysis (SRS-GDA toolbox, <https://doi.org/10.5281/zenodo.4044626>) developed by the authors used in this study.

References

- Banwell, N., Rutherford, S., Mackey, B., Street, R., & Chu, C. (2018). Commonalities between disaster and climate change risks for health: A theoretical framework. *International Journal of Environmental Research and Public Health*, 15(3), 538. <https://doi.org/10.3390/ijerph15030538>
- Buishand, T. A., De Haan, L., & Zhou, C. (2008). On spatial extremes: With application to a rainfall problem. *Annals of Applied Statistics*, 2(2), 624–642. <https://doi.org/10.1214/08-aos159>
- Buytaert, W., Celleri, R., Willems, P., De Bievre, B., & Wyseure, G. (2006). Spatial and temporal rainfall variability in mountainous areas: A case study from the south Ecuadorian Andes. *Journal of Hydrology*, 329(3–4), 413–421. <https://doi.org/10.1016/j.jhydrol.2006.02.031>
- Chandler, R. E., & Wheeler, H. S. (2002). Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland. *Water Resources Research*, 38(10), 10–11. <https://doi.org/10.1029/2001wr000906>
- Coe, R., & Stern, R. D. (1982). Fitting models to daily rainfall data. *Journal of Applied Meteorology*, 21(7), 1024–1031. [https://doi.org/10.1175/1520-0450\(1982\)021<1024:fmdtrd>2.0.co;2](https://doi.org/10.1175/1520-0450(1982)021<1024:fmdtrd>2.0.co;2)

- Collier, P. (2002). *Development of Australia's National GDA94 Transformation Grids. Consultant's Report to the Intergovernmental Committee on Surveying and Mapping*. Australia: University of Melbourne.
- Cressie, N., & Hawkins, D. M. (1980). Robust estimation of the variogram: I. *Mathematical Geology*, *12*(2), 115–125. <https://doi.org/10.1007/bf01035243>
- CSIRO & Australian Bureau of Meteorology. (2018). *Previous State of the Climate Reports*. Retrieved from <https://www.csiro.au/en/Research/OandA/Areas/Assessing-our-climate/State-of-the-Climite-2018/Previous-State-of-the-Climite-reports>
- Efron, B., & Tibshirani, R. (1997). Improvements on cross-validation: The 632+ bootstrap method. *Journal of the American Statistical Association*, *92*(438), 548–560. <https://doi.org/10.1080/01621459.1997.10474007>
- Feng, S., Nadarajah, S., & Hu, Q. (2007). Modeling annual extreme precipitation in China using the generalized extreme value distribution. *Journal of the Meteorological Society of Japan*, *85*(5), 599–613. <https://doi.org/10.2151/jmsj.85.599>
- Frederiksen, C. S., Zheng, X., & Grainger, S. (2014). Teleconnections and predictive characteristics of Australian seasonal rainfall. *Climate Dynamics*, *43*(5–6), 1381–1408. <https://doi.org/10.1007/s00382-013-1952-0>
- Hosking, J. R. M., Wallis, J. R., & Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, *27*(3), 251–261. <https://doi.org/10.1080/00401706.1985.10488049>
- James, G. M. (2002). Generalized linear models with functional predictors. *Journal of the Royal Statistical Society: Series B*, *64*(3), 411–432. <https://doi.org/10.1111/1467-9868.00342>
- Jones, D., Wang, W., & Fawcett, R. (2009). High-quality spatial climate data-sets for Australia. *Australian Meteorological and Oceanographic Journal*, *58*(4), 233. <https://doi.org/10.22499/2.5804.003>. Retrieved from <http://www.bom.gov.au/jsp/awap/rain/index.jsp>
- Jung, Y., Shin, J. Y., Ahn, H., & Heo, J. H. (2017). The spatial and temporal structure of extreme rainfall trends in South Korea. *Water*, *9*(10), 809.
- Kendon, E. J., Fosser, G., Murphy, J., Chan, S., Clark, R., Harris, G., et al. (2019). *UKCP convection-permitting model projections: Science Report*. Retrieved from <https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science?reports/UKCP?Convection?permitting?-model?projections?report.pdf>
- Kendon, M., McCarthy, M., & Jevrejeva, S. (2015). *State of the UK climate 2014*. Met Office.
- Kendon, M., McCarthy, M., Jevrejeva, S., Matthews, A., & Legg, T. (2018). State of the UK climate 2017. *International Journal of Climatology*, *38*, 1–35. <https://doi.org/10.1002/joc.5798>
- Kendon, M., McCarthy, M., Jevrejeva, S., Matthews, A., & Legg, T. (2019). State of the UK climate 2018. *International Journal of Climatology*, *39*, 1–55. <https://doi.org/10.1002/joc.6213>
- Li, Z., Li, Z., Zhao, W., & Wang, Y. (2015). Probability modeling of precipitation extremes over two river basins in northwest of China. *Advances in Meteorology*, *2015*. <https://doi.org/10.1155/2015/374127>
- Lowe, J. A., Bernie, D., Bett, P., Bricheno, L., Brown, S., Calvert, D., et al. (2018). *UKCP18 Science Overview Report*. Exeter, UK: Met Office Hadley Centre
- McCullagh, P. (2018). *Generalized Linear Models*. Routledge.
- Millán, M. M., Estrela, M. J., & Miró, J. (2005). Rainfall components: Variability and spatial distribution in a Mediterranean Area (Valencia Region). *Journal of Climate*, *18*(14), 2682–2705. <https://doi.org/10.1175/jcli3426.1>
- Ordnance Survey. (1946). *A Brief Description of the National Grid and Reference System London*. His Majesty's Stationery Office (HMSO).
- Overeem, A., Buishand, T. A., Holleman, I., & Uijlenhoet, R. (2010). Extreme value modeling of areal rainfall from weather radar. *Water Resources Research*, *46*(9). <https://doi.org/10.1029/2009wr008517>
- Pedersen, L., Jensen, N. E., Christensen, L. E., & Madsen, H. (2010). Quantification of the spatial variability of rainfall based on a dense network of rain gauges. *Atmospheric Research*, *95*(4), 441–454. <https://doi.org/10.1016/j.atmosres.2009.11.007>
- Peleg, N., Marra, F., Faticchi, S., Paschalis, A., Molnar, P., & Burlando, P. (2018). Spatial variability of extreme rainfall at radar subpixel scale. *Journal of Hydrology*, *556*, 922–933. <https://doi.org/10.1016/j.jhydrol.2016.05.033>
- Pierce, D. A., & Schafer, D. W. (1986). Residuals in generalized linear models. *Journal of the American Statistical Association*, *81*(396), 977–986. <https://doi.org/10.1080/01621459.1986.10478361>
- Prein, A. F., Rasmussen, R. M., Ikeda, K., Liu, C., Clark, M. P., & Holland, G. J. (2017). The future intensification of hourly precipitation extremes. *Nature Climate Change*, *7*(1), 48–52. <https://doi.org/10.1038/nclimate3168>
- Rashid, M., Beecham, S., Chowdhury, R., Piantadosi, J., Anderssen, R. S., & Boland, J. (2013). Simulation of extreme rainfall from CMIP5 in the Onkaparinga catchment using a generalized linear model. In *20th International Congress on Modelling and Simulation (MODSIM2013)*, 1–6 December 2013, Adelaide, Australia.
- Rogger, M., Agnoletti, M., Alaoui, A., Bathurst, J. C., Bodner, G., Borga, M., et al. (2017). Land use change impacts on floods at the catchment scale: Challenges and opportunities for future research. *Water Resources Research*, *53*(7), 5209–5219. <https://doi.org/10.1002/2017wr020723>
- Schaefer, M. G. (1990). Regional analyses of precipitation annual maxima in Washington State. *Water Resources Research*, *26*(1), 119–131. <https://doi.org/10.1029/wr026i001p00119>
- Segond, M. L., Onof, C., & Wheeler, H. S. (2006). Spatial-temporal disaggregation of daily rainfall from a generalized linear model. *Journal of Hydrology*, *331*(3–4), 674–689. <https://doi.org/10.1016/j.jhydrol.2006.06.019>
- Stern, R. D., & Coe, R. (1984). A model fitting analysis of daily rainfall data. *Journal of the Royal Statistical Society: Series A*, *147*(1), 1–18. <https://doi.org/10.2307/2981736>
- Tanguy, M., Dixon, H., Prosdocimi, I., Morris, D. G., & Keller, V. D. J. (2016). *Gridded Estimates of Daily and Monthly Areal Rainfall for the United Kingdom (1890-2015)[CEH-GEAR]*. NERC Environmental Information Data Centre. <https://doi.org/10.5285/33604ea0-c238-4488-813d-0ad9ab7c51ca>
- Wang, H., & Xuan, Y. (2020). SRS-GDA: A spatial random sampling toolbox for grid-based hydro-climatic data analysis in environmental change studies. *Environmental Modelling & Software*, *124*, 104598. <https://doi.org/10.1016/j.envsoft.2019.104598>
- Wang, P. C. (1987). Residual plots for detecting nonlinearity in generalized linear models. *Technometrics*, *29*(4), 435–438. <https://doi.org/10.1080/00401706.1987.10488271>
- Westra, S., Alexander, L. V., & Zwiers, F. W. (2013). Global increasing trends in annual maximum daily precipitation. *Journal of Climate*, *26*(11), 3904–3918. <https://doi.org/10.1175/jcli-d-12-00502.1>
- Westra, S., Fowler, H. J., Evans, J. P., Alexander, L. V., Berg, P., Johnson, F., et al. (2014). Future changes to the intensity and frequency of short-duration extreme rainfall. *Reviews of Geophysics*, *52*(3), 522–555. <https://doi.org/10.1002/2014rg000464>
- Yan, Z., Bate, S., Chandler, R. E., Isham, V., & Wheeler, H. (2002). An analysis of daily maximum wind speed in northwestern Europe using generalized linear models. *Journal of Climate*, *15*(15), 2073–2088. [https://doi.org/10.1175/1520-0442\(2002\)015<2073:aaodmw>2.0.co;2](https://doi.org/10.1175/1520-0442(2002)015<2073:aaodmw>2.0.co;2)

- Yang, C., Chandler, R. E., Isham, V. S., & Wheeler, H. S. (2005). Spatial-temporal rainfall simulation using generalized linear models. *Water Resources Research*, *41*(11). <https://doi.org/10.1029/2004wr003739>
- Yoon, P., Kim, T.-W., & Yoo, C. (2013). Rainfall frequency analysis using a mixed GEV distribution: a case study for annual maximum rainfalls in South Korea. *Stochastic Environmental Research and Risk Assessment*, *27*(5), 1143–1153. <https://doi.org/10.1007/s00477-012-0650-5>
- Zheng, Y., Xue, M., Li, B., Chen, J., & Tao, Z. (2016). Spatial characteristics of extreme rainfall over China with hourly through 24-hour accumulation periods based on national-level hourly rain gauge data. *Advances in Atmospheric Sciences*, *33*(11), 1218–1232. <https://doi.org/10.1007/s00376-016-6128-5>