Spatial Variation of Extreme Rainfall Observed from Two Century-long Datasets

H. Wang\textsuperscript{1} and Y. Xuan\textsuperscript{*}

\textsuperscript{1}College of Engineering, Swansea University Bay Campus, Swansea SA1 8EN, United Kingdom.

*Corresponding author: Y. Xuan (y.xuan@swansea.ac.uk)

Key Points:

- Rainfall series 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 modelled qualitatively and quantitatively showing impact from location, size and shape of regions.
- Methods and findings provide new perspective to understand heterogeneous nature of climate variability and climate change impact.
Abstract

This paper presents the spatial variation of annual maximum daily rainfall (AMDR), represented by the fitted generalized extreme value (GEV) distributions, from two century-long datasets of Great Britain (GB) and Australia with respect to three spatial properties: geographic locations, sizes and shapes of the region of interest (ROI). The results show that the GEV fits well the areal AMDR. The spatial variation of the GEV location-scale parameters, quantified by the generalized linear models, is dominated by geographic locations and area sizes with an eastward-decreasing-banded-pattern in GB and a concentrically-increasing-pattern from the middle to the coasts in Australia. Although the impact of the ROI shapes is insignificant, the round-shaped regions usually have higher-valued parameters than the elongated ones. The findings provide a new perspective to understanding the heterogeneity of extreme rainfall distribution in the space driven by the complex interactions among 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 amount with preferred exceedance probabilities. This procedure is always associated with a region of interest (ROI). Whilst 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. It is clear that 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, urbanisation, as well as the orientation and the size of the area can also affect the rainfall amount in question (Buytaert et al., 2006). Many studies, e.g., Buishand et al., 2008; Jung et al., 2017; Pedersen et al., 2010; Villarini et al., 2010; Zheng et al., 2016, have attempted to understand the spatial variation of rainfall extremes at different scales based on gauged records.

In addition, spatially disaggregated, grid-based hydro-climatic datasets, have become more accessible to the research community; processing these new datasets to support large-scale variation analysis of grid-by-grid extremes has become an important research topic to address (Peleg et al., 2018). Some studies have focused on spatial variation of grid-based hydroclimatic observations, e.g., UKCIP, Banwell et al., 2018; Kendon et al., 2019; Lowe et al., 2018; Prein et al., 2017; Others attempted to address the temporal variation, for example, 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. Further, 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) are analysed with an overall aim of gaining 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 and civil engineering design. Specifically, the study attempts to address the following questions:

1. How areal rainfall extremes change over space.
2. How other factors such as the size, shapes of the area in question may affect such spatial dependencies.
3. How the spatial patterns and variations are linked to the climate variability.
4. What is the implication of the spatial variation of the parameters to the applications (e.g. flood risk management).

The two datasets come with a duration over 100 years and have relatively high temporal and spatial resolutions (daily and 1–5 km respectively). In addition, 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. The toolbox can automatically generate arbitrary ROIs with predefined or randomised features, i.e., size, location and dominant orientation, from the supplied grid-based dataset. The sampled annual maximum daily rainfall (AMDR) at each ROI is fitted with the widely used and tested Generalised Extreme Value (GEV) distributions whose spatial variation is then analysed. The associated intensive computation demand is met by the high-performance computing (HPC) resources provided by Super Computing Wales (https://www.supercomputing.wales).

The remainder of this paper is organized as follows: section 2 describes the data and methods then shows the sampled ROI and the goodness-of-fit tests results. Both the qualitative and quantitative results of the spatial variation of the distribution parameters, as well as their linkage to the climate are discussed in sections 3, 4, and 5. Finally, 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 (1x1 km²) rainfall estimation that covers the mainland of Great Britain (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. The natural neighbour interpolation method with a normalisation step based on the average annual rainfall, was used to generate the daily estimates (Tanguy et al., 2016). The ADAM dataset is generated using a sophisticated analysis technique described in Jones et al. (2009), which is also grid-based (0.05°x0.05°, approx. 5x5 km²) rainfall from 01/01/1900 to 31/12/2018 over Australia (AU). The recorded rainfall values are provided as daily rainfall, i.e. the total rainfall amount over a predefined 24-hour (9AM-9AM) period which refers to the 24 hours prior to the reporting time for the ADAM dataset and the 24 hours after for the GEAR dataset.

2.2. Methodology

The geographical areas of the two data domains, i.e. GB and AU, are sampled into a series of ROIs using the SRS-GDA toolbox before the AMDR values are extracted from each
ROI. Three different types of predefined spatial features (geographical locations, sizes and shapes) are applied in this spatial sampling process to reduce the overall computing time while maintaining the representativeness of the samples. As a result, these ROIs are evenly distributed across the two study domains. The AMDR extracted from each ROI is then fitted with a probability distribution. In this study, the three-parameter GEV distribution is chosen as the candidate distribution. The goodness of fit (GOF) of the fitted distributions are further tested by two different methods: the Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests. The parameters ($\mu$ and $\sigma$) of the fitted distributions are then analysed with regards to their spatial distribution with reference to the climate variations.

1) ROI Generation and AMDR Extraction

The ROI sampling starts with an initial set of evenly distributed ROIs which comprise of 7 predefined shapes, parameterised by their spatial indexes (Wang & Xuan, 2020) reciprocally grouped as 0.2/5.0, 0.5/2.0, 0.8/1.25 and 1.0. The size of these ROIs is then gradually increased by 10 steps with 20% increment each, while maintaining the same shape and location (of the centroid). In the end, the largest sizes the ROIs are 1,050 km$^2$ for GB and 9,900 km$^2$ for AU respectively.

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 toolbox is configured to keep the centroid location unchanged while generating 10 ROIs only by varying their sizes. Table 1 also summarises all ROIs and their properties.

Table 1. ROIs for analysing 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             | Indicator              | G1                                            | G2                                            | G3                                            |
| ROI(s) of 1x1 km grid |                        | ![Image](image1.jpg)                           | ![Image](image2.jpg)                           | ![Image](image3.jpg)                           |
| Size (km$^2$)  | 500                    | 10, 43, 87, 164, 257, 366, 504, 660, 827, 1025 | 500 each ROI                                  |
| Total ROI number | 88                     | 81 x 10 = 810                                | 74 x 7 = 518                                  |
| Total meridional group | 10                     | 10                                            | 10                                            |
### Geographical location (marked as “×”)

| AU | Indicator  | A1 | A2 | A3 |
|----|------------|----|----|----|
| ROI(s) of 5x5 km grid | | | | |
| Size (km²) | 500 | 125, 400, 900, 1550, 2450, 3550, 4875, 6350, 8025, 9900 | 5000 each ROI |
| Total ROI number | 679 | 627 × 10 = 6270 | 378 × 7 = 2646 |
| Total meridional group number | 40 | 38 | 30 |

---

**Note:** This content has not been peer reviewed.
The extraction of the AMDR series from each ROI is then carried out using the super computers from Super Computing Wales. In total, there is around 642.3GB of data processed with 11,011 areal AMDR series produced.

2) Fitting the extracted AMDRs using GEV distribution

Derived from the extreme value theory, the generalised extreme value (GEV) distribution has become by far one of the most well-founded distributions for describing annual maximum rainfall. It has been applied to not only many gauged rainfall extreme studies (Feng et al., 2007; Martins & Stedinger, 2000; Westra et al., 2013) but also those using grid rainfall datasets (Overeem et al., 2010). A GEV distribution is controlled by three parameters, namely, the location $\mu$, the scale $\sigma$ and the shape $\xi$ parameter which defines the three limiting types: the Gumbel ($\xi = 0$), the Frechét ($\xi > 0$) and the Weibull ($\xi < 0$). The cumulative probability distribution function of GEV is given by:

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

where $F$ is defined for $\left\{ 1 + \frac{\xi(x-\mu)}{\sigma} > 0 \right\}$, $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$; and $x$ denotes the extracted AMDR. A maximum likelihood estimator (MLE, Hosting, 1985) is introduced to estimate the three parameters of the GEV distribution fitted to those AMDRs extracted from each ROI.

Although the GEV distribution generally fits well to the point rainfall extremes (e.g. gauge observation) as reported in many studies before (Schaefer, 1990; Yoon et al., 2013), very few have been done on the suitability of GEV distribution fitting the areal grid-based rainfall extremes. In this study, the GOF is tested using bootstrapped KS and AD tests (see supplementary text S1). Out of all the AMDR series from every ROI (1416 ROIs of GB and 9595 ROIs 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% for the AD test.

3) Analysing the spatial distribution of the location-scale parameters

The spatial variation of the location and scale parameters of the fitted GEV distributions are analysed 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. The procedure is briefly described below:

i. The chosen GEV parameter is aggregated meridionally, e.g. over all ROIs that have the same $x$-direction (easting or longitude) coordinate.

ii. The aggregated GEV parameter values are indexed by their $x$-direction only coordinate which is then used as an input variable to represent the geographical locations.

iii. The same procedure is also applied zonally, i.e., 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 (predictor). In AU, a concentric pattern is found where both the meridional average and the zonal average show a similar result; For the case of GB, only a strong west-east pattern exists. Therefore, for comparing two cases and convenience, the meridional average is taken for both cases.

Finally, a generalised linear model (GLM) is fitted to quantify the relationship between the GEV parameters and the associated spatial features, i.e., to explicitly model the spatial variation of the GEV parameters with respect to the locations, sizes, and shapes of the underlying ROIs.

3 Results and discussions

3.1. GEV parameter variation over geographical locations

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

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

(a)
(b) Number of ROIs, \( \xi \), \( \sigma \), and \( \mu \) for different ROIs sizes.

(c) Meridional average \( \mu \) and \( \sigma \) for different size of ROIs (km\(^2\)).
Figure 1. Histograms and spatial variations of the three GEV parameters in GB (a) and AU (b); the changes of meridional average location-scale parameters with the ROIs’ size in GB (c) and AU (d); and the ROIs’ shape sp in GB (e) and AU (f), where the x values shown in the legends refer to the geographic location (index in the x-direction) of the meridional groups.

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

The values of $\mu$ and $\sigma$ 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 notable that the rapid variations are close to the north-eastern coastal regions. For a matter of convenience, the meridional average is also taken for studying the west-east variation in AU.

3.2. Variation of GEV parameters with regards to the area size

Figures 1c and 1d show the changes of $\mu$ and $\sigma$ of all meridional groups in GB and AU, parameterised by the size of the ROI ($s$, in km$^2$). Generally, regardless of their locations ($x$), the parameter values are inversely proportional to the sizes of the ROIs, as reflected by the fitted trend lines.

The decreases in both $\mu$ and $\sigma$ with increased ROI sizes have an important implication: the most frequent AMDR (relating to $\mu$) becomes smaller for larger ROI alongside an overall
decreased extremity (relating 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., $x$-index from 300 to 360 km), and it increases near the coastal areas where large parameter values are also found. This feature can be explained by the fact that regions having more extreme rainfall (e.g. the outer coastal regions in AU) are not only manifested by the higher $\mu$ and $\sigma$; they also have more heterogenous rainfall than those with less extreme rainfall (lower $\mu$ and $\sigma$). Therefore, the changes of $\mu$ and $\sigma$ 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 change of ROI shape

Figures 1e and 1f present the changes of $\mu$ and $\sigma$ in GB and AU, parameterised by the ROI shape ($sp$). The variation of the shape starts from west-east orientated shapes ($sp = 0.2$), gradually growing into more rounded shapes ($sp = 1.0$) and then to more north-south orientated shapes ($sp = 5.0$). By the definition of $sp$, two shapes with reciprocal $sp$ values will have their major dimension swapped, i.e. east-west versus south-north and vice versa. The result is inspected and summarized as:

- For the majority of the meridional groups, there is little difference between the location-scale parameters of ROIs with reciprocal shapes, e.g. two shapes with $sp$ values of 0.2 and 5.0. This is regarded as a symmetric pattern around $sp = 1.0$;
- Generally, the values of $\mu$ and $\sigma$ 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 not as significant.

4 Quantify the spatial variation

The generalised linear models (GLM) are based on an extension to the classical linear regression model (McCullagh, 1989), and have found many applications in hydrology and meteorology (Coe & Stern, 1982; Stern & Coe, 1984). GLMs have been shown to be effective in incorporating complex structures (Segond et al., 2006). Chandler & Wheater (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 since followed, e.g. Yan et al., 2002; Yang et al., 2005; Rashid et al., 2013, with good performance reported.

In this study, the two parameters $\mu$ and $\sigma$ which reflect the property of rainfall extremes, also show a similar right-skewed distribution (Figure S1), therefore we broadly followed Chandler & Wheater (2002) and propose a GLM with a log-link to describe their spatial variation. Helped by the qualitative analysis in section 3, the three spatial properties of the underlying ROIs, i.e. the size ($s$), location ($x$-index: $x$), and shape ($sp$), as well as their interactions are chosen to be the candidates of the predictors.

The fitting of the GLM starts with a simplest form and then successively adds other predictors or their combinations (Chandler & Wheater, 2002; James, 2002). The significance of
the newly added predictor or the combination of each attempt is evaluated by calculating the
value of likelihood. The best fitted form of GLMs is obtained by considering both the likelihood
and the discrepancy (e.g. root mean squared error, RMSE). More details can be found in Text S2.
Finally, the optimum form of GLM models are identified as follows:

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

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

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

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

where $\beta$ is the estimated vector of coefficients of predictors and the subscripts GB and
AU refer to the study area in question. A maximum likelihood estimator (McCullagh, 1989) was
employed for obtaining $\beta$. These fitted GLMs help to reveal the following intriguing features
regarding the spatial variation of the two parameters:

1. In GB, both the meridionally averaged $\mu$ and $\sigma$ have a nonlinear dependency on the
geographical location, i.e., the easting index ($x$); and a linear dependency on the ROI size
($s$). However, they do not appear to be dependent on the ROI shape ($sp$).

2. In AU, the spatial changes of meridionally averaged $\mu$ and $\sigma$ are nonlinear with respect to
both the easting index and the ROI size. Further, the shape of ROI ($sp$) plays a more
significant role in contributing to the change of GEV parameters than it does in the case
of GB. The combined factor ($x \times s$) is significant in contributing to the variation of $\sigma$.

The GLMs are further visualised in Figures 2a and 2b where the previously
demonstrated, qualitative properties, are readily reproduced. For example, the spatial changes of
the two GEV parameters are “west high, east low” in GB whereas they are “centre low, outer
coastal regions high” in AU; the parameters get smaller as the size of ROI increases. However,
the RR, which can be interpreted as the vertical distance between curves, is more uneven in AU,
which means that the reduction on most frequent rainfall ($\mu$) and occurrence probability of
extremes ($\sigma$) is more spatially dependent and area-oriented comparing with GB. Moreover, ROI
shape is significant in the AU case where different $\mu$ or $\sigma$ values are observed in the east-west-
orientated elongated shapes ($sp = 0.2$) and the rounded ones ($sp = 1.0$) and the difference tends
to decrease for lager $\mu$ and $\sigma$. In comparison, the two GEV parameters in the north-south-
orientated ROIs ($sp = 5.0$) are also smaller than those in the east-west-orientated and rounded
ones, which can be explained as that in AU the north-south variation is in general smaller than
that of the east-west direction.
(b)
manuscript submitted to Geophysical Research Letters

(c)
Figure 2. Visualisation of the GLMs fitted to the meridional average GEV $\mu$ and $\sigma$ parameters as a colour-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 modelled GEV parameters.

The performance of the GLMs is evaluated by comparing the parameter values modelled by the GLMs and those from the originally fitted GEVs (Figure 2); as well as by conducting a residual analysis (McCullagh, 2018; Pierce & Schafer, 1986; Wang, 1987).

The GLMs for both cases perform well (Figure 2c). The GB case has slight underestimations for some large values that appear in the western coastal region; and for the AU case, some overestimation happens for the small values which are located in the middle-south dry zone. The GLM model probability structure is checked by the normal quantile plot (Figure 2d) of the residuals, where a theoretical normal distribution is shown on the x-axis compared with 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. It can be observed that 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.
5 Link between the spatial variations of GEV parameters and the climate variation

The GEV distribution parameters can reveal the characteristics of extreme rainfall in terms of both its amount and occurrence probability. The two parameters are shown to have strong spatial dependency as discussed in the previous sections. To help understand how such spatial variation of the extreme rainfall is related to the climate variability over space, Figure 3 is produced to display the spatial distributions of both the average annual rainfall and its standard deviation of GB and AU respectively, compared with the spatial distribution of the GEV parameters. A great deal of similarity exits in space between the corresponding quantities, i.e. the location parameter versus the annual mean, and the scale parameter versus the standard deviation of the annual rainfall. For example, regions with higher annual average rainfall are not only presented with higher standard deviation (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 values of the GEV parameters, and appear to be more heterogenous. This feature also exists in the 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). This findings is consistent with those published in the series of climate reports of GB (Kendon et al., 2015; Kendon et al., 2018, 2019) and AU (CSIRO & Australian Bureau of Meteorology, 2018).

![Image of average annual rainfall and standard deviation in GB](image)

### Average annual rainfall and its standard deviation (SD) in GB

| GEV $\mu$ and $\sigma$ of AMDR in GB |
|-------------------------------------|
| SD of average annual rainfall/mm    |
| - 585 (max)                         |
| - 500                               |
| - 400                               |
| - 300                               |
| - 200                               |
| - 100                               |
| - 88 (min)                          |

| GEV $\sigma$                        |
|-------------------------------------|
| - 21.0 (max)                        |
| - 16                                |
| - 12                                |
| - 8                                 |
| - 5.3 (min)                         |
Figure 3. Comparison between the climatic variables (the average annual rainfall and its standard deviation) and GEV parameters $\mu$ and $\sigma$ in GB (a) and AU (b) cases where the colour denotes the value of averaged annual rainfall or the GEV parameter $\mu$, and the size of the circles denotes the value of standard deviation of the annual rainfall or the GEV parameter $\sigma$.

6 Conclusions

This paper presents a study on the spatial variation of extreme rainfall using two-century long datasets covering Great Britain and Australia. The annual maximum daily rainfall (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 analysed over the space. Four generalised linear models (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 the GB and AU cases; more than 90% of the regions are better fitted with the Frechét type of distribution among the three GEV types.

2) The GEV location ($\mu$) and scale ($\sigma$) parameters present similar spatial patterns where a higher $\mu$ is usually accompanied by a higher $\sigma$ 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. Not only are they supported and consistent with many previous studies on rainfall distributions, the quantification of the extreme rainfall and its spatial dependencies are of great practical value in engineering design, e.g. designed rainfall/floods for constructions. The methods employed by 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. In addition, impact of the density of the underlying data observations, i.e., rain-gauges, and its variation over long term also need to be further studied.

Acknowledgments

The authors would like to thank the Centre of Hydrology and Ecology (CEH) and The Bureau of Meteorology, Australia for providing the datasets. The open-source toolbox of spatial random sampling for grid-based data analysis (SRS-GDA toolbox, doi: 10.5281/zenodo.4044626) developed by the authors used in this study. 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 Royal Academy of Engineering UK-China Urban Flooding Programme Grant (REF: UUFRI\10021), which are gratefully acknowledged.

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.
Buishand, T. A., De Haan, L., & Zhou, C. (2008). On spatial extremes: with application to a rainfall problem. *The Annals of Applied Statistics*, 2(2), 624-642.

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.

Chandler, R. E., & Wheater, H. S. (2002). Analysis of rainfall variability using generalized linear models: a case study from the west of Ireland. *Water Resources Research*, 38(10), 1010-1011.

Coe, R., & Stern, R. D. (1982). Fitting models to daily rainfall data. *Journal of Applied Meteorology*, 21(7), 1024-1031.

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-Clima-2018/Previous-State-of-the-Clima-reports.

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. Ser. II*, 85(5), 599-613.

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.

James, G. M. (2002). Generalized linear models with functional predictors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3), 411-432.

Jones, D. A., Wang, W., & Fawcett, R. (2009). High-quality spatial climate data-sets for Australia. *Australian Meteorological and Oceanographic Journal*, 58(4), 233.

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., Lock, A., Lowe, J., Martin, G., & Pirret, J. (2019). UKCP Convection-permitting model projections: Science report.

Kendon, M., MacCarthy, 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.

Kendon, M., McCarthy, M., Jevrejeva, S., Matthews, A., & Legg, T. (2019). State of the UK climate 2018. *International Journal of Climatology*, 39, 1-55.

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.

Lowe, J. A., Bernie, D., Bett, P., Bricheno, L., Brown, S., Calvert, D., Clark, R., Eagle, K., Edwards, & T., Fosser, G. (2018). UKCP18 science overview report. Met Office Hadley Centre: Exeter, UK.

Martins, E. S., & Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, 36(3), 737-744.
McCullagh, P. (1989). Generalized linear models: Routledge.

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*(4), 2682-2705.

Overeem, A., Buishand, T. A., Holleman, I., & Uijlenhoet, R. (2010). Extreme value modeling of areal rainfall from weather radar. *Water Resources Research, 46*(9).

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.

Peleg, N., Marra, F., Fatichi, S., Paschalis, A., Molnar, P., & Burlando, P. (2018). Spatial variability of extreme rainfall at radar subpixel scale. *Journal of hydrology, 556*, 922-933.

Pierce, D. A., & Schafer, D. W. (1986). Residuals in generalized linear models. *Journal of the American Statistical Association, 81*(396), 977-986.

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.

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.

Rogger, M., Agnoletti, M., Alaoui, A., Bathurst, J. C., Bodner, G., Borga, M., Chaplot, V., Gallart, F., Glatzel, G., & Hall, J. (2017). Land use change impacts on floods at the catchment scale: Challenges and opportunities for future research. *Water Resources Research, 53*(7), 5209-5219.

Schaefer, M. G. (1990). Regional analyses of precipitation annual maxima in Washington State. *Water Resources Research, 26*(1), 119-131.

Segond, M. L., Onof, C., & Wheater, H. S. (2006). Spatial–temporal disaggregation of daily rainfall from a generalized linear model. *Journal of hydrology, 331*(3-4), 674-689.

Stern, R. D., & Coe, R. (1984). A model fitting analysis of daily rainfall data. *Journal of the Royal Statistical Society: Series A (General), 147*(1), 1-18.

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.

Villarini, G., Smith, J. A., Baek, M. L., Sturdevant-Rees, P., & Krajewski, W. F. (2010). Radar analyses of extreme rainfall and flooding in urban drainage basins. *Journal of hydrology, 381*(3-4), 266-286.

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.
Wang, P. (1987). Residual plots for detecting nonlinearity in generalized linear models. *Technometrics, 29*(4), 435-438.

Westra, S., Alexander, L. V., & Zwiers, F. W. (2013). Global increasing trends in annual maximum daily precipitation. *Journal of Climate, 26*(11), 3904-3918.

Westra, S., Fowler, H. J., Evans, J. P., Alexander, L. V., Berg, P., Johnson, F., Kendon, E. J., Lenderink, G., & Roberts, N. M. (2014). Future changes to the intensity and frequency of short-duration extreme rainfall. *Reviews of Geophysics, 52*(3), 522-555.

Yan, Z., Bate, S., Chandler, R. E., Isham, V., & Wheater, H. (2002). An analysis of daily maximum wind speed in northwestern Europe using generalized linear models. *Journal of Climate, 15*(15), 2073-2088.

Yang, C., Chandler, R. E., Isham, V. S., & Wheater, H. S. (2005). Spatial-temporal rainfall simulation using generalized linear models. *Water Resources Research, 41*(11).

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.

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.