

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

2 **H. Wang¹ and Y. Xuan^{1*}**

3 ¹College of Engineering, Swansea University Bay Campus, Swansea SA1 8EN, United
4 Kingdom.

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

6 **Key Points:**

- 7 • Rainfall series extracted and processed from more than 11,000 regions of interest over the
8 last 100 years in Great Britain and Australia.
- 9 • Spatial distribution of extreme rainfall modelled qualitatively and quantitatively showing
10 impact from location, size and shape of regions.
- 11 • Methods and findings provide new perspective to understand heterogenous nature of
12 climate variability and climate change impact.

13 Abstract

14 This paper presents the spatial variation of annual maximum daily rainfall (AMDR), represented
15 by the fitted generalized extreme value (GEV) distributions, from two century-long datasets of
16 Great Britain (GB) and Australia with respect to three spatial properties: geographic locations,
17 sizes and shapes of the region of interest (ROI). The results show that the GEV fits well the areal
18 AMDR. The spatial variation of the GEV location-scale parameters, quantified by the
19 generalized linear models, is dominated by geographic locations and area sizes with an eastward-
20 decreasing-banded-pattern in GB and a concentrically-increasing-pattern from the middle to the
21 coasts in Australia. Although the impact of the ROI shapes is insignificant, the round-shaped
22 regions usually have higher-valued parameters than the elongated ones. The findings provide a
23 new perspective to understanding the heterogeneity of extreme rainfall distribution in the space
24 driven by the complex interactions among climate, geographical features and the practical
25 sampling approaches.

26 1 Introduction

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

41 In addition, spatially disaggregated, grid-based hydro-climatic datasets, have become
42 more accessible to the research community; processing these new datasets to support large-scale
43 variation analysis of grid-by-grid extremes has become an important research topic to address
44 (Peleg et al., 2018). Some studies have focused on spatial variation of grid-based hydroclimatic
45 observations, e.g., UKCIP, Banwell et al., 2018; Kendon et al., 2019; Lowe et al., 2018; Prein et
46 al., 2017; Others attempted to address the temporal variation, for example, frequency analysis
47 (Li et al., 2015; Overeem et al., 2010).

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

53 In this study, two century-long, grid-based rainfall datasets covering Great Britain (GB)
54 and Australia (AU) are analysed with an overall aim of gaining insights into how area-orientated
55 rainfall extremes vary with space with respect to the probability distribution parameters which

56 are of concern of flood risk management and civil engineering design. Specifically, the study
57 attempts to address the following questions:

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

64 The two datasets come with a duration over 100 years and have relatively high temporal
65 and spatial resolutions (daily and 1–5 km respectively). In addition, a toolbox known as the
66 Spatial Random Sampling for Grid-based Data Analysis (SRS-GDA, Wang & Xuan, 2020), is
67 employed to assist the required spatial sampling. The toolbox can automatically generate
68 arbitrary ROIs with predefined or randomised features, i.e., size, location and dominant
69 orientation, from the supplied grid-based dataset. The sampled annual maximum daily rainfall
70 (AMDR) at each ROI is fitted with the widely used and tested Generalised Extreme Value
71 (GEV) distributions whose spatial variation is then analysed. The associated intensive
72 computation demand is met by the high-performance computing (HPC) resources provided by
73 Super Computing Wales (<https://www.supercomputing.wales>).

74 The remainder of this paper is organized as follows: section 2 describes the data and
75 methods then shows the sampled ROI and the goodness-of-fit tests results. Both the qualitative
76 and quantitative results of the spatial variation of the distribution parameters, as well as their
77 linkage to the climate are discussed in sections 3, 4, and 5. Finally, the conclusions and
78 recommendations of further study are given in section 6.

79 **2 Data and Methods**

80 **2.1. Datasets**

81 This study makes use of two century-long datasets which are the ‘Gridded Estimates of
82 daily Areal Rainfall’ (GEAR) and the ‘Australian Data Archive for Meteorology’ (ADAM). The
83 GEAR dataset is a grid-based ($1 \times 1 \text{ km}^2$) rainfall estimation that covers the mainland of Great
84 Britain (GB) from 01/01/1898 to 31/12/2010. It is derived from the UK Met Office national
85 database of observed precipitation from the UK rain gauge network. The natural neighbour
86 interpolation method with a normalisation step based on the average annual rainfall, was used to
87 generate the daily estimates (Tanguy et al., 2016). The ADAM dataset is generated using a
88 sophisticated analysis technique described in Jones et al. (2009), which is also grid-based
89 ($0.05^\circ \times 0.05^\circ$, approx. $5 \times 5 \text{ km}^2$) rainfall from 01/01/1900 to 31/12/2018 over Australia (AU). The
90 recorded rainfall values are provided as daily rainfall, i.e. the total rainfall amount over a
91 predefined 24-hour (9AM-9AM) period which refers to the 24 hours prior to the reporting time
92 for the ADAM dataset and the 24 hours after for the GEAR dataset.

93 **2.2. Methodology**

94 The geographical areas of the two data domains, i.e. GB and AU, are sampled into a
95 series of ROIs using the SRS-GDA toolbox before the AMDR values are extracted from each

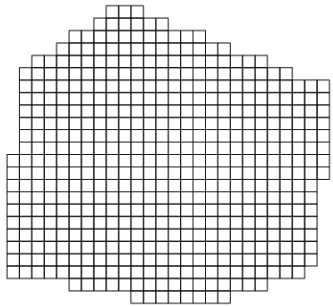
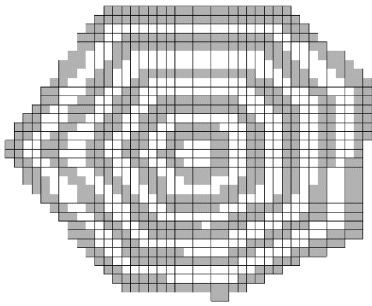
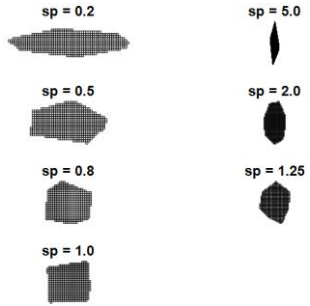
96 ROI. Three different types of predefined spatial features (geographical locations, sizes and
 97 shapes) are applied in this spatial sampling process to reduce the overall computing time while
 98 maintaining the representativeness of the samples. As a result, these ROIs are evenly distributed
 99 across the two study domains. The AMDR extracted from each ROI is then fitted with a
 100 probability distribution. In this study, the three-parameter GEV distribution is chosen as the
 101 candidate distribution. The goodness of fit (GOF) of the fitted distributions are further tested by
 102 two different methods: the Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests. The
 103 parameters (μ and σ) of the fitted distributions are then analysed with regards to their spatial
 104 distribution with reference to the climate variations.

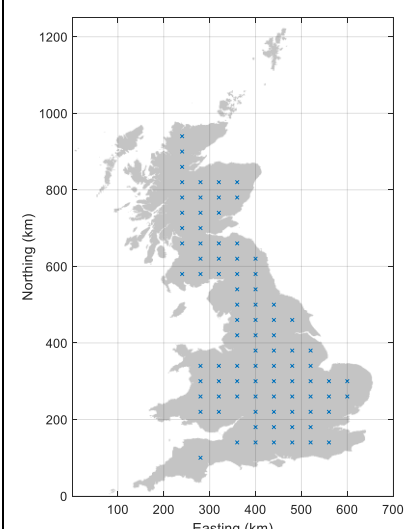
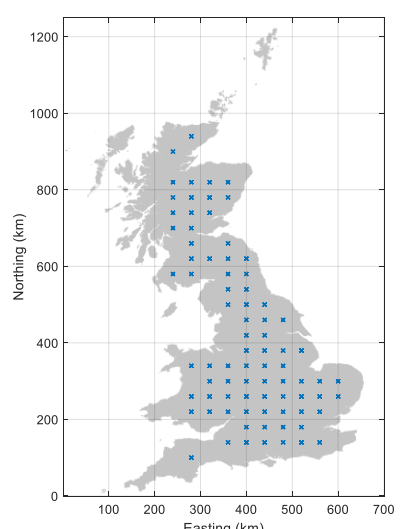
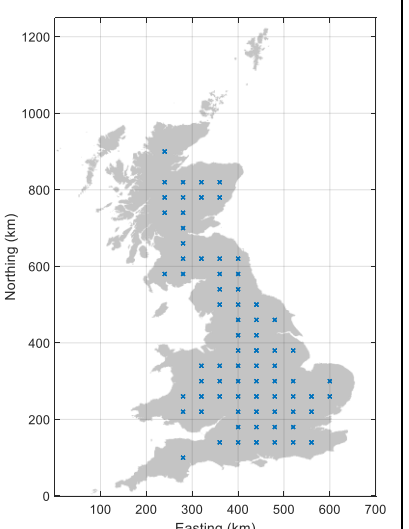
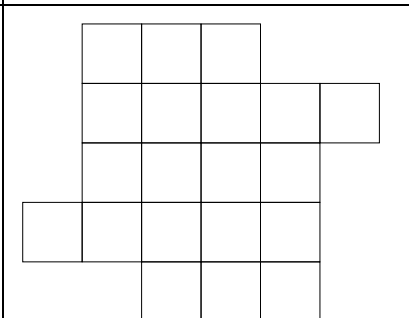
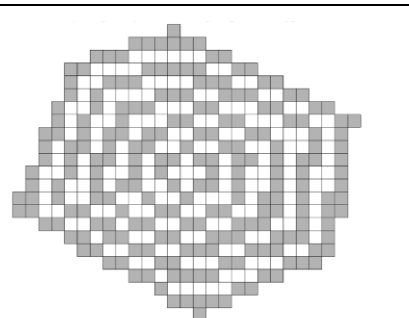
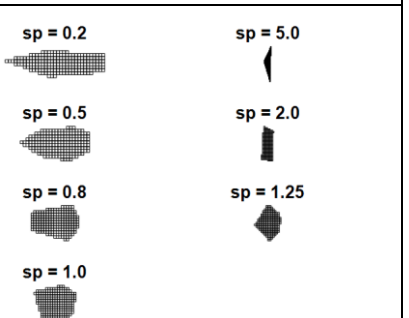
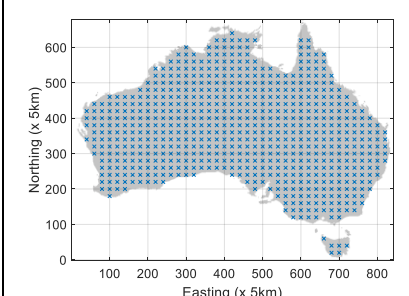
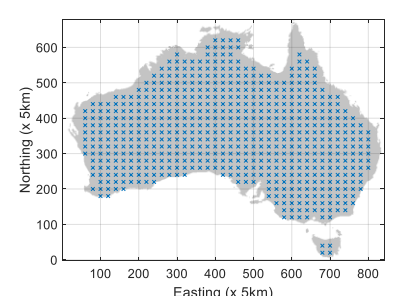
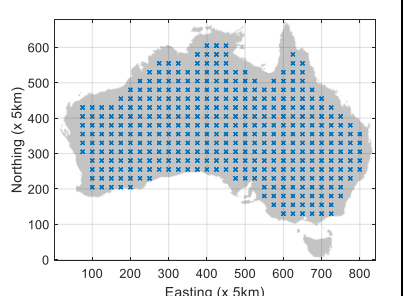
105 **1) ROI Generation and AMDR Extraction**

106 The ROI sampling starts with an initial set of evenly distributed ROIs which comprise of
 107 7 predefined shapes, parameterised by their *spatial indexes* (Wang & Xuan, 2020) reciprocally
 108 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
 109 by 10 steps with 20% increment each, while maintaining the same shape and location (of the
 110 centroid). In the end, the largest sizes the ROIs are 1,050 km² for GB and 9,900 km² for AU
 111 respectively.

112 The SRS-GDA toolbox used to generate the ROIs is set up in a way that only one spatial
 113 feature is allowed to vary at a time. For instance, to obtain ROI samples of G2 and A2 in Table
 114 1, the toolbox is configured to keep the centroid location unchanged while generating 10 ROIs
 115 only by varying their sizes. Table 1 also summarises all ROIs and their properties.

116 **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			
	Size (km ²)	500	10, 43, 87, 164, 257, 366, 504, 660, 827, 1025	500 each ROI
	Total ROI number	88	81 × 10 = 810	74 × 7 = 518
Total meridional group	10	10	10	

	number			
	Geographical location (marked as "X")			
AU	Indicator	A1	A2	A3
	ROI(s) of 5x5 km grid			
	Size (km ²)	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
	Geographical location (marked as "X")			

117

118 The extraction of the AMDR series from each ROI is then carried out using the super
 119 computers from Super Computing Wales. In total, there is around 642.3GB of data processed
 120 with 11,011 areal AMDR series produced.

121 **2) Fitting the extracted AMDRs using GEV distribution**

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

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

130 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
 131 denotes the extracted AMDR. A maximum likelihood estimator (MLE, Hosting, 1985) is
 132 introduced to estimate the three parameters of the GEV distribution fitted to those AMDRs
 133 extracted from each ROI.

134 Although the GEV distribution generally fits well to the point rainfall extremes (e.g.
 135 gauge observation) as reported in many studies before (Schaefer, 1990; Yoon et al., 2013), very
 136 few have been done on the suitability of GEV distribution fitting the areal grid-based rainfall
 137 extremes. In this study, the GOF is tested using bootstrapped KS and AD tests (see
 138 supplementary text S1). Out of all the AMDR series from every ROI (1416 ROIs of GB and
 139 9595 ROIs of AU) tested, the results show that the GEV distribution fits well the AMDR series
 140 with a 100% pass of the KS test and more than 97% for the AD test.

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

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

- 146 i. The chosen GEV parameter is aggregated meridionally, e.g. over all ROIs that have the
 147 same x -direction (easting or longitude) coordinate.
- 148 ii. The aggregated GEV parameter values are indexed by their x -direction only coordinate
 149 which is then used as an input variable to represent the geographical locations.
- 150 iii. The same procedure is also applied zonally, i.e., over the same y -direction coordinate.

151 With this arrangement, the meridional or zonal average of the GEV parameter in question
 152 is taken as the response variable (predictor). In AU, a concentric pattern is found where both the
 153 meridional average and the zonal average show a similar result; For the case of GB, only a strong
 154 west-east pattern exists. Therefore, for comparing two cases and convenience, the meridional
 155 average is taken for both cases.

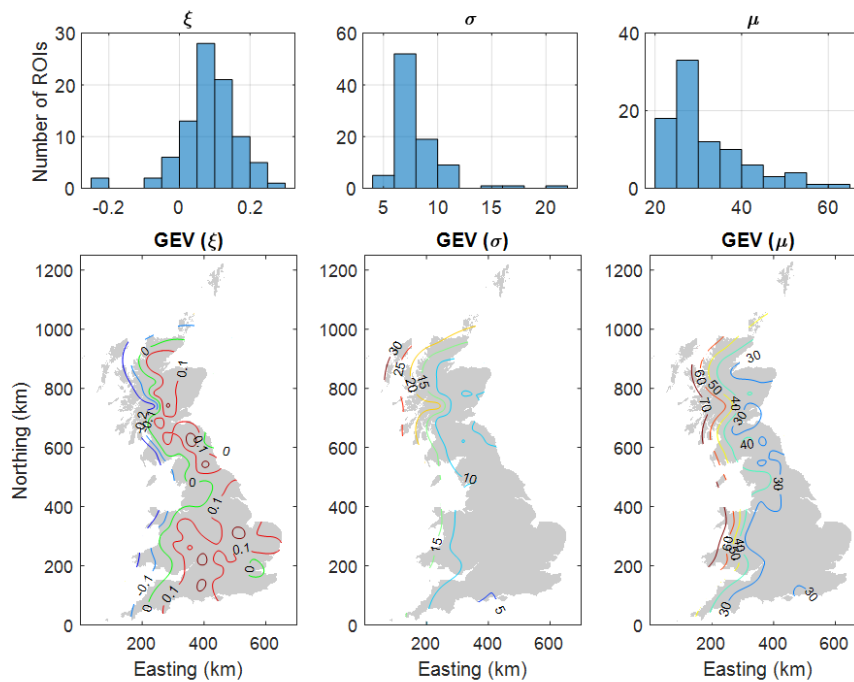
156 Finally, a generalised linear model (GLM) is fitted to quantify the relationship between
 157 the GEV parameters and the associated spatial features, i.e., to explicitly model the spatial
 158 variation of the GEV parameters with respect to the locations, sizes, and shapes of the underlying
 159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

160 3 Results and discussions

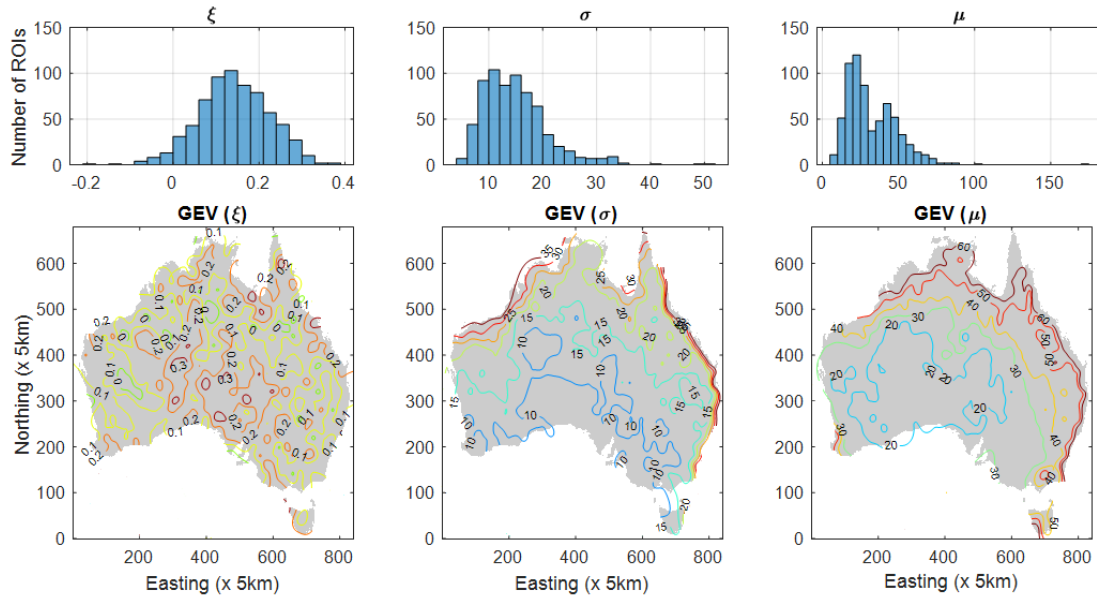
161 3.1. GEV parameter variation over geographical locations

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

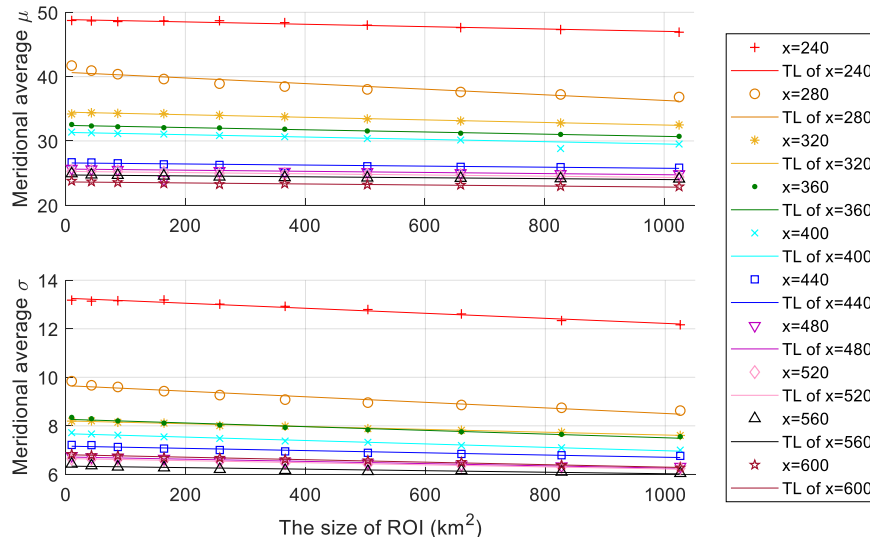
- 164 • Most ROIs are in favour of the Fréchet type of distribution ($\xi > 0$).
- 165 • Both σ and μ present a similar spatial pattern where a higher μ is usually accompanied by a
 166 higher σ .



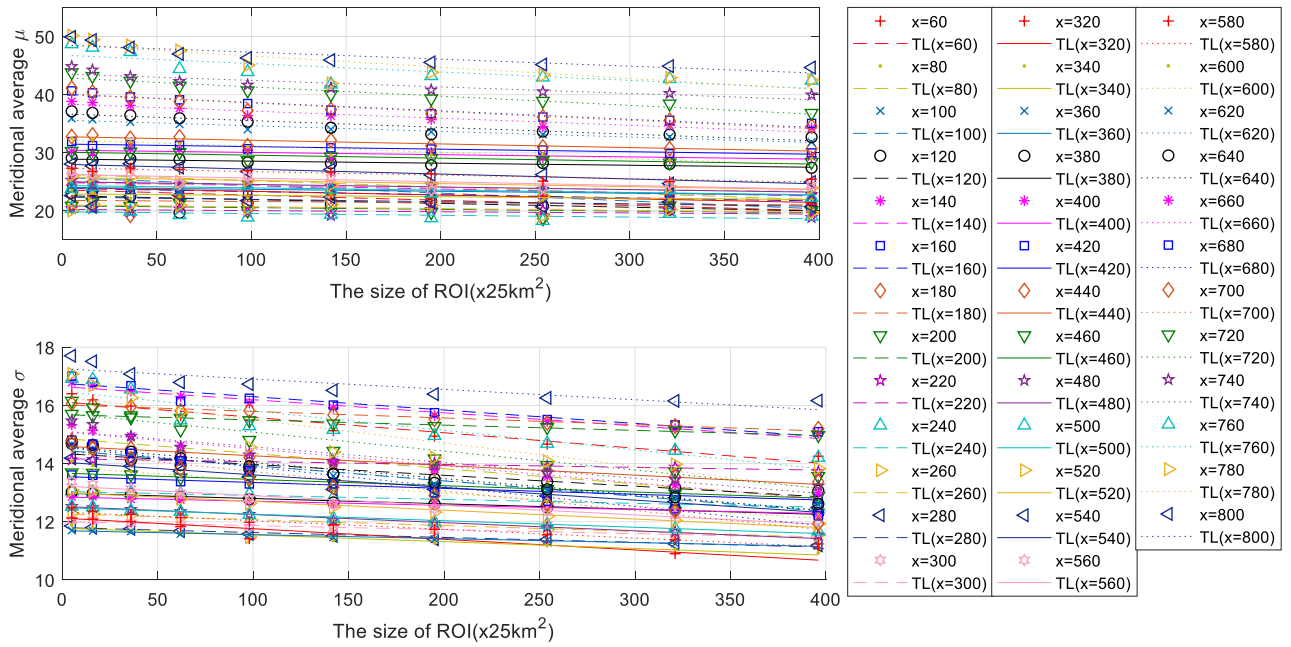
(a)



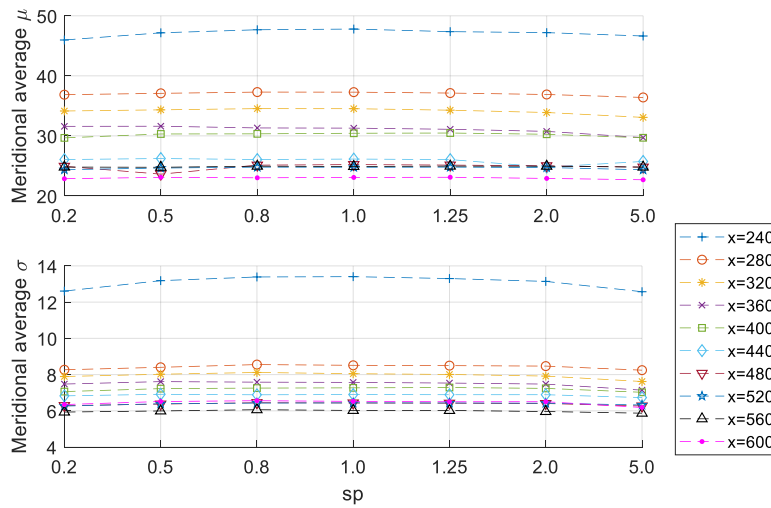
(b)



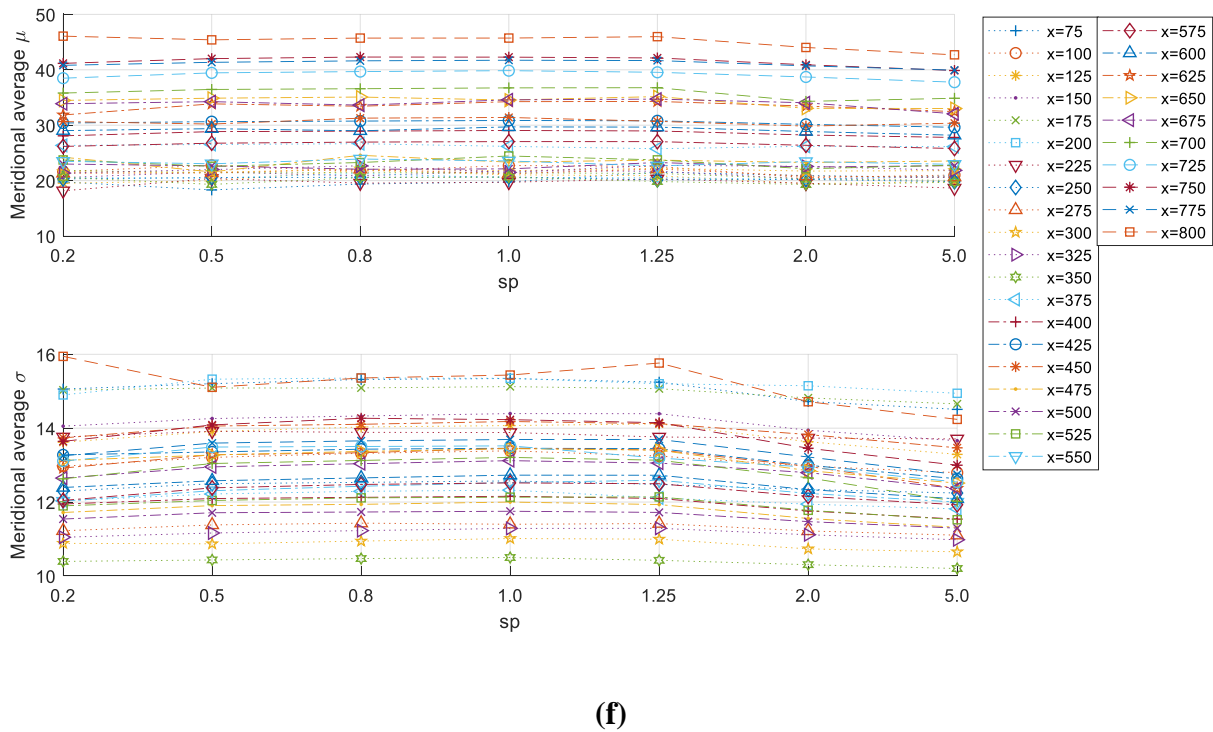
(c)



(d)



(e)



167 **Figure 1.** Histograms and spatial variations of the three GEV parameters in GB (a) and AU (b);
 168 the changes of meridional average location-scale parameters with the ROIs' size in GB (c) and
 169 AU (d); and the ROIs' shape sp in GB (e) and AU (f), where the x values shown in the legends
 170 refer to the geographic location (index in the x -direction) of the meridional groups.

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

177 The values of μ and σ in AU have a clear increasing trend from the south-middle zone to
 178 the coastal regions. This spatial pattern can be seen as a series of concentric circles. It is also
 179 notable that the rapid variations are close to the north-eastern coastal regions. For a matter of
 180 convenience, the meridional average is also taken for studying the west-east variation in AU.

181 3.2. Variation of GEV parameters with regards to the area size

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

186 The decreases in both μ and σ with increased ROI sizes have an important implication:
 187 the most frequent AMDR (relating to μ) becomes smaller for larger ROI alongside an overall

188 decreased extremity (relating both parameters). Another interesting measure is the rate of such
 189 reduction (RR) as the size of ROI increases, which has also shown a clear spatial dependency. In
 190 AU, the RR remains low in the central desert zone (e.g., x -index from 300 to 360 km), and it
 191 increases near the coastal areas where large parameter values are also found. This feature can be
 192 explained by the fact that regions having more extreme rainfall (e.g. the outer coastal regions in
 193 AU) are not only manifested by the higher μ and σ ; they also have more heterogenous rainfall
 194 than those with less extreme rainfall (lower μ and σ). Therefore, the changes of μ and σ are more
 195 sensitive to geographic locations, as revealed by the RR. GB also shows a similar pattern albeit
 196 not as remarkable.

197 3.3. Variation of GEV parameters due to change of ROI shape

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

- 204 • For the majority of the meridional groups, there is little difference between the location-scale
 205 parameters of ROIs with reciprocal shapes, e.g. two shapes with sp values of 0.2 and 5.0.
 206 This is regarded as a symmetric pattern around $sp = 1.0$;
- 207 • Generally, the values of μ and σ of ROIs in an elongated shape are smaller than those of the
 208 ROIs in more rounded shapes. This indicates that the rounded-shape ROIs have a better
 209 chance to capture more rainfall extremes than the elongated ones. It also leads to that for the
 210 same area size, regions with more regular shape tend to have more extreme areal rainfall.
- 211 • Overall, the effects of ROI shape are not as significant.

212 4 Quantify the spatial variation

213 The generalised linear models (GLM) are based on an extension to the classical linear
 214 regression model (McCullagh, 1989), and have found many applications in hydrology and
 215 meteorology (Coe & Stern, 1982; Stern & Coe, 1984). GLMs have been shown to be effective in
 216 incorporating complex structures (Segond et al., 2006). Chandler & Wheater (2002) proposed a
 217 GLM-based framework for interpreting historical daily rainfall records and revealing the changes
 218 on rainfall occurrence and amount in western Ireland. Many more applications have since
 219 followed, e.g. Yan et al., 2002; Yang et al., 2005; Rashid et al., 2013, with good performance
 220 reported.

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

227 The fitting of the GLM starts with a simplest form and then successively adds other
 228 predictors or their combinations (Chandler & Wheater, 2002; James, 2002). The significance of

229 the newly added predictor or the combination of each attempt is evaluated by calculating the
 230 value of likelihood. The best fitted form of GLMs is obtained by considering both the likelihood
 231 and the discrepancy (e.g. root mean squared error, RMSE). More details can be found in Text S2.
 232 Finally, the optimum form of GLM models are identified as follows:

$$\text{For } \mu_{GB}: (1 + x + s + x^2 + x^3)\boldsymbol{\beta}_{\mu_{GB}} \quad 2a)$$

$$\text{For } \sigma_{GB}: (1 + x + s + x^2 + x^3)\boldsymbol{\beta}_{\sigma_{GB}} \quad 2b)$$

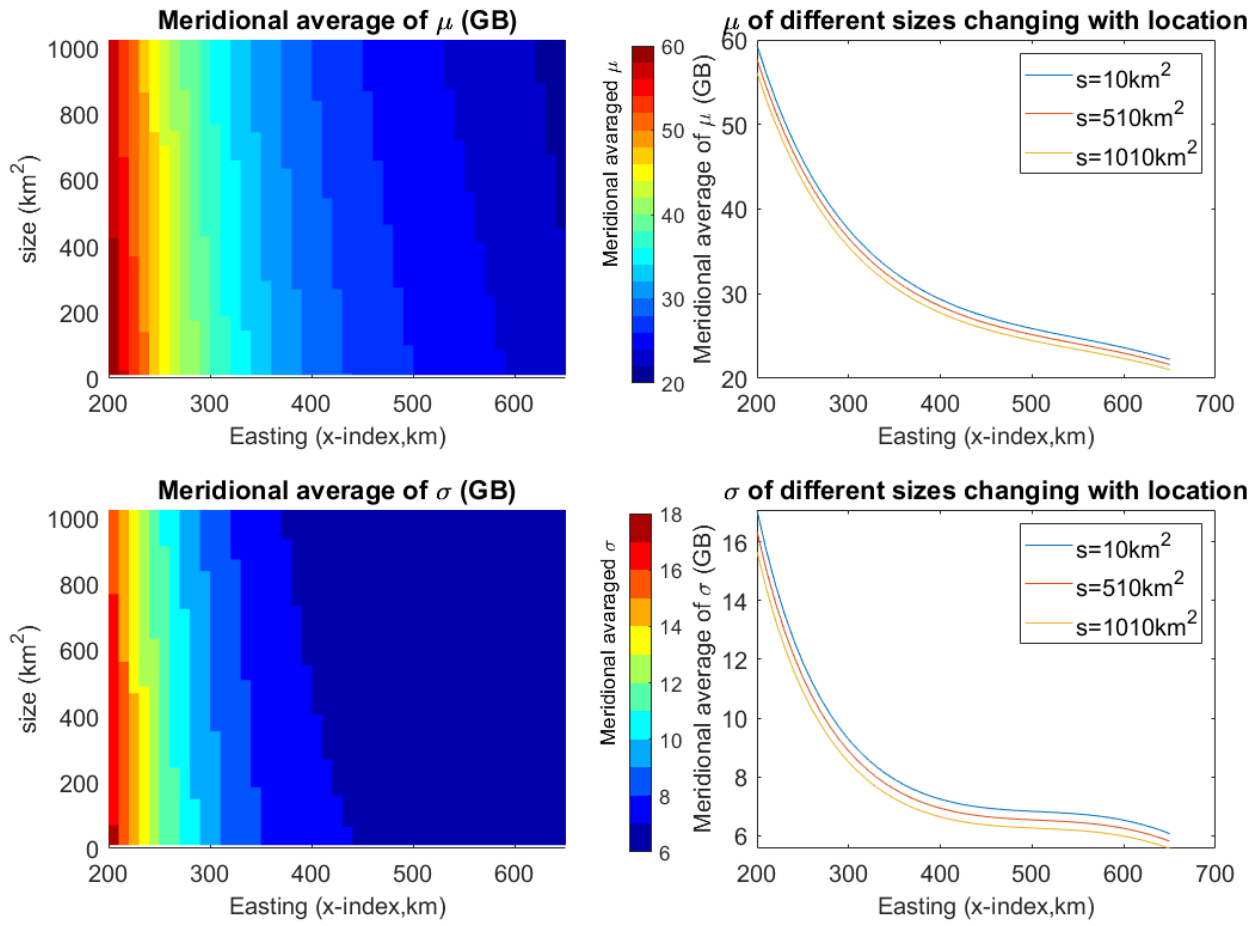
$$\text{For } \mu_{AU}: (1 + x + s + sp + x^2 + s^2)\boldsymbol{\beta}_{\mu_{AU}} \quad 3a)$$

$$\text{For } \sigma_{AU}: (1 + x + s + sp + x^2 + s^2 + x \times s)\boldsymbol{\beta}_{\sigma_{AU}} \quad 3b)$$

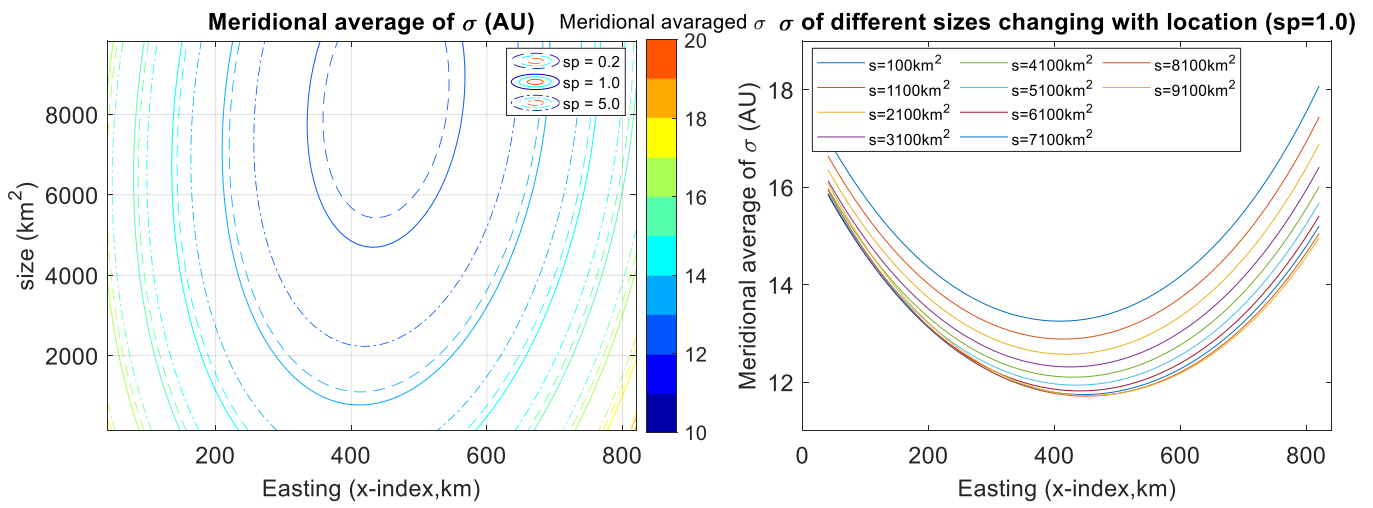
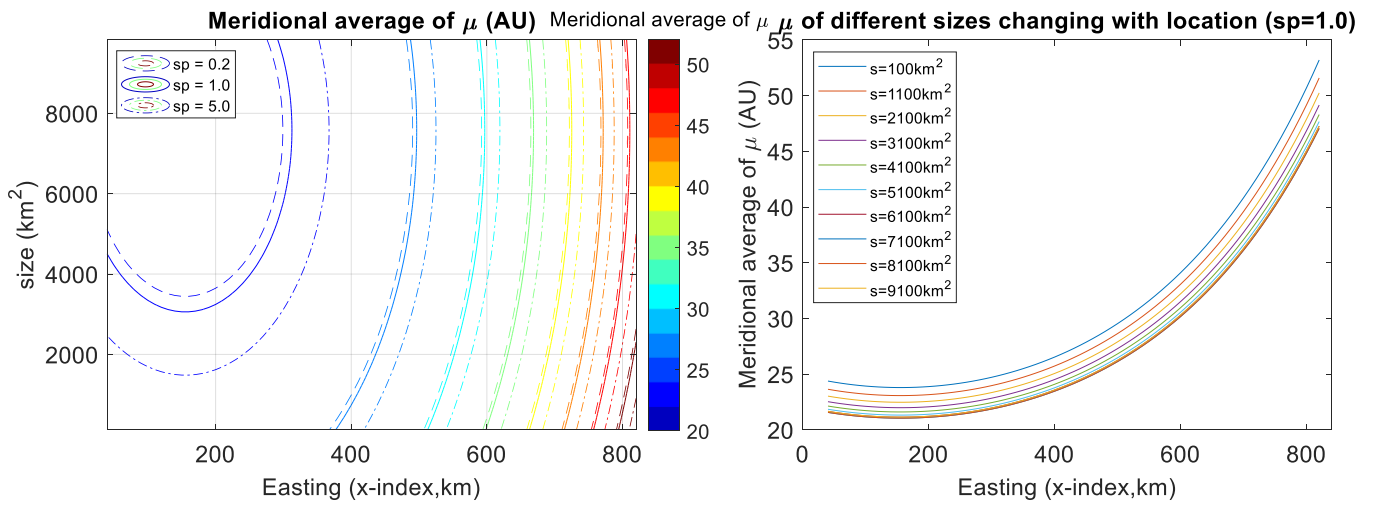
233 where $\boldsymbol{\beta}$ is the estimated vector of coefficients of predictors and the subscripts GB and
 234 AU refer to the study area in question. A maximum likelihood estimator (McCullagh, 1989) was
 235 employed for obtaining $\boldsymbol{\beta}$. These fitted GLMs help to reveal the following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

- 237 1. In GB, both the meridionally averaged μ and σ have a nonlinear dependency on the
 238 geographical location, i.e., the easting index (x); and a linear dependency on the ROI size
 239 (s). However, they do not appear to be dependent on the ROI shape (sp).
- 240 2. In AU, the spatial changes of meridionally averaged μ and σ are nonlinear with respect to
 241 both the easting index and the ROI size. Further, the shape of ROI (sp) plays a more
 242 significant role in contributing to the change of GEV parameters than it does in the case
 243 of GB. The combined factor ($x \times s$) is significant in contributing to the variation of σ .

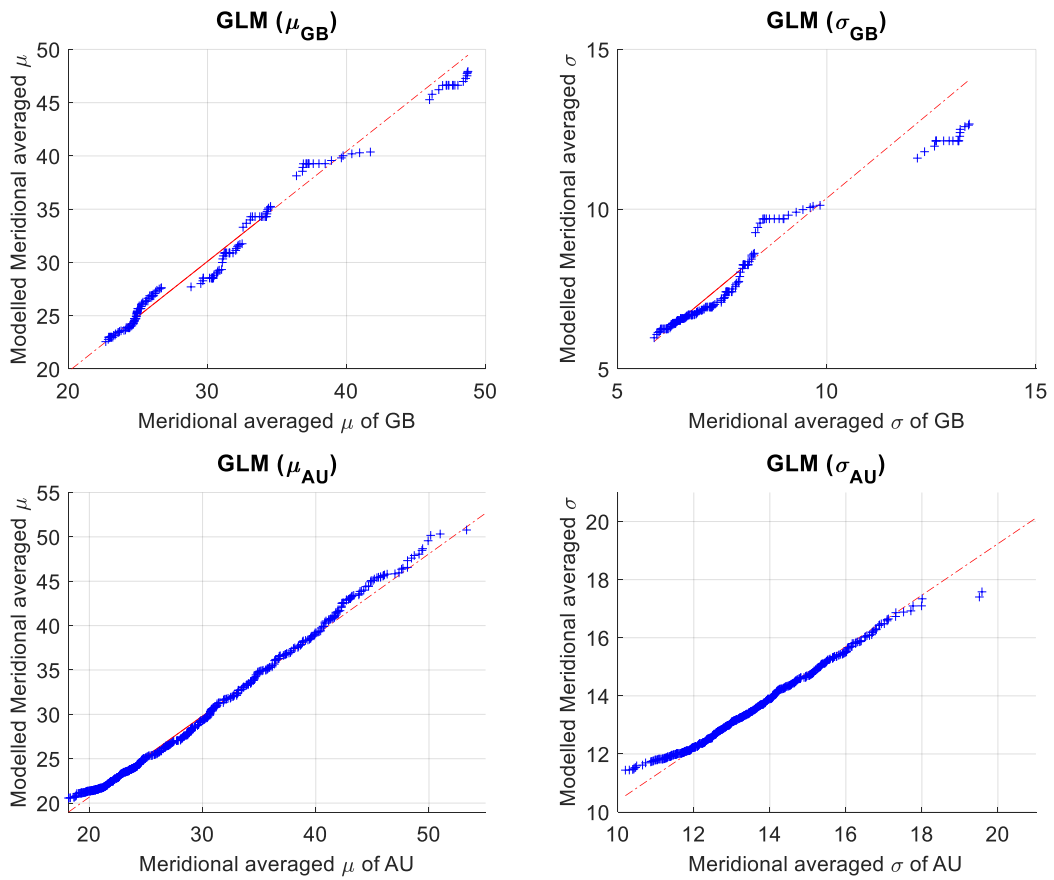
244 The GLMs are further visualised in Figures 2a and 2b where the previously
 245 demonstrated, qualitative properties, are readily reproduced. For example, the spatial changes of
 246 the two GEV parameters are “west high, east low” in GB whereas they are “centre low, outer
 247 coastal regions high”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



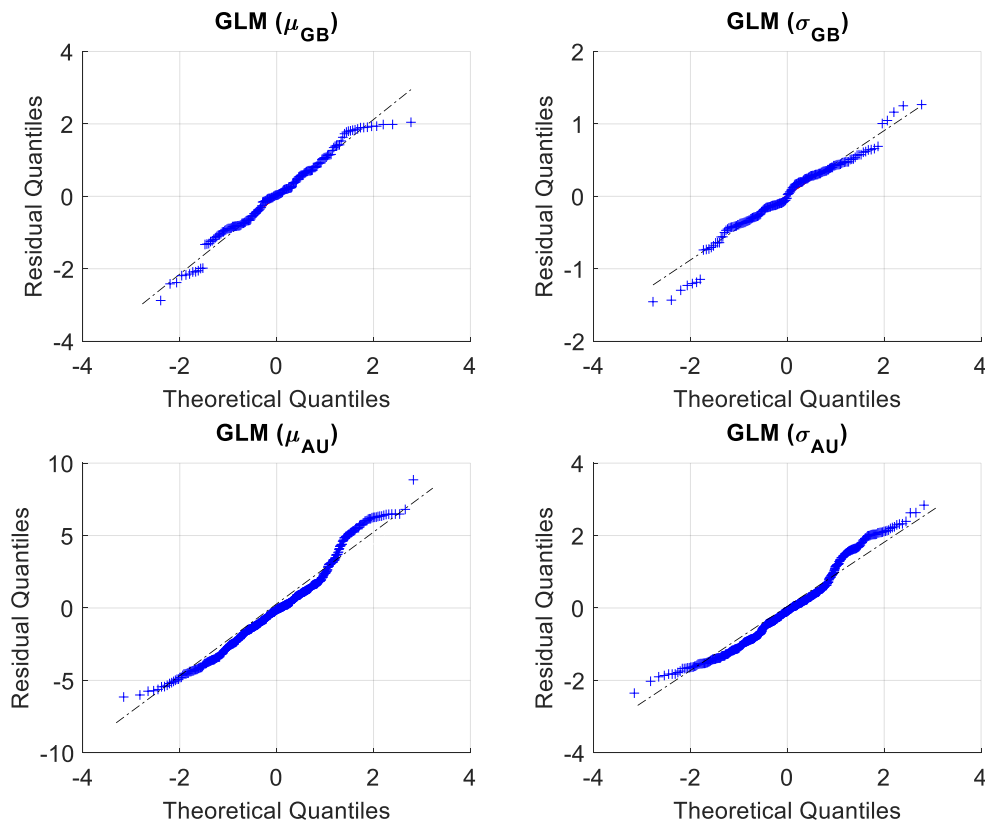
(a)



(b)



(c)



(d)

257 **Figure 2.** Visualisation of the GLMs fitted to the meridional average GEV μ and σ parameters as
 258 a colour-scale plot for GB (a) and a contour plot for AU (b) whose contours are picked up at the
 259 same stops of the values and their changes with respect to the geographic location and size; and a
 260 scatter plot (c) and a normal quantile plot (d) for revealing the difference between the actual
 261 GEV parameters and the modelled GEV parameters.

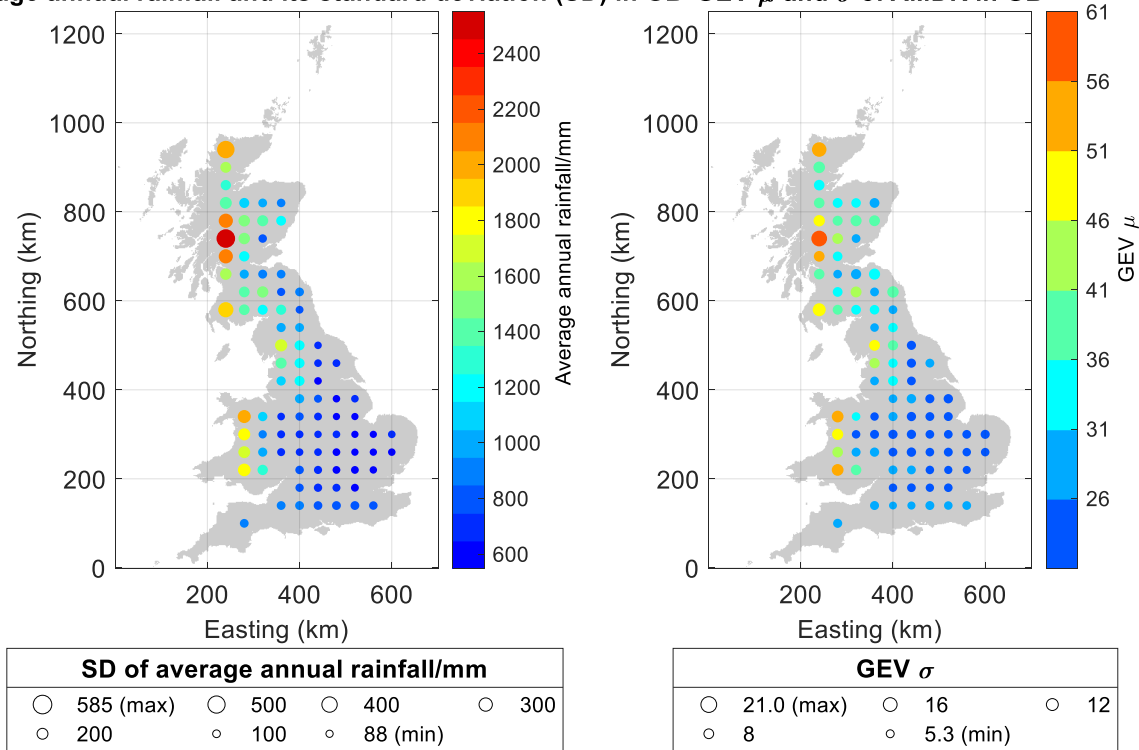
262 The performance of the GLMs is evaluated by comparing the parameter values modelled
 263 by the GLMs and those from the originally fitted GEVs (Figure 2); as well as by conducting a
 264 residual analysis (McCullagh, 2018; Pierce & Schafer, 1986; Wang, 1987).

265 The GLMs for both cases perform well (Figure 2c). The GB case has slight
 266 underestimations for some large values that appear in the western coastal region; and for the AU
 267 case, some overestimation happens for the small values which are located in the middle-south
 268 dry zone. The GLM model probability structure is checked by the normal quantile plot (Figure
 269 2d) of the residuals, where a theoretical normal distribution is shown on the x -axis compared
 270 with the residual quantiles on the y -axis. If the probability assumption (i.e. gamma assumption)
 271 is correct, all residuals would have the same distribution which is an approximate normal
 272 distribution. It can be observed that the distribution of the residuals of the four GLMs is
 273 symmetric with two flat sides. Generally, the approximation fits well except for the upper side
 274 which represents only 0.9% of the total data points. In view of the research aims, this is
 275 considered to be acceptable.

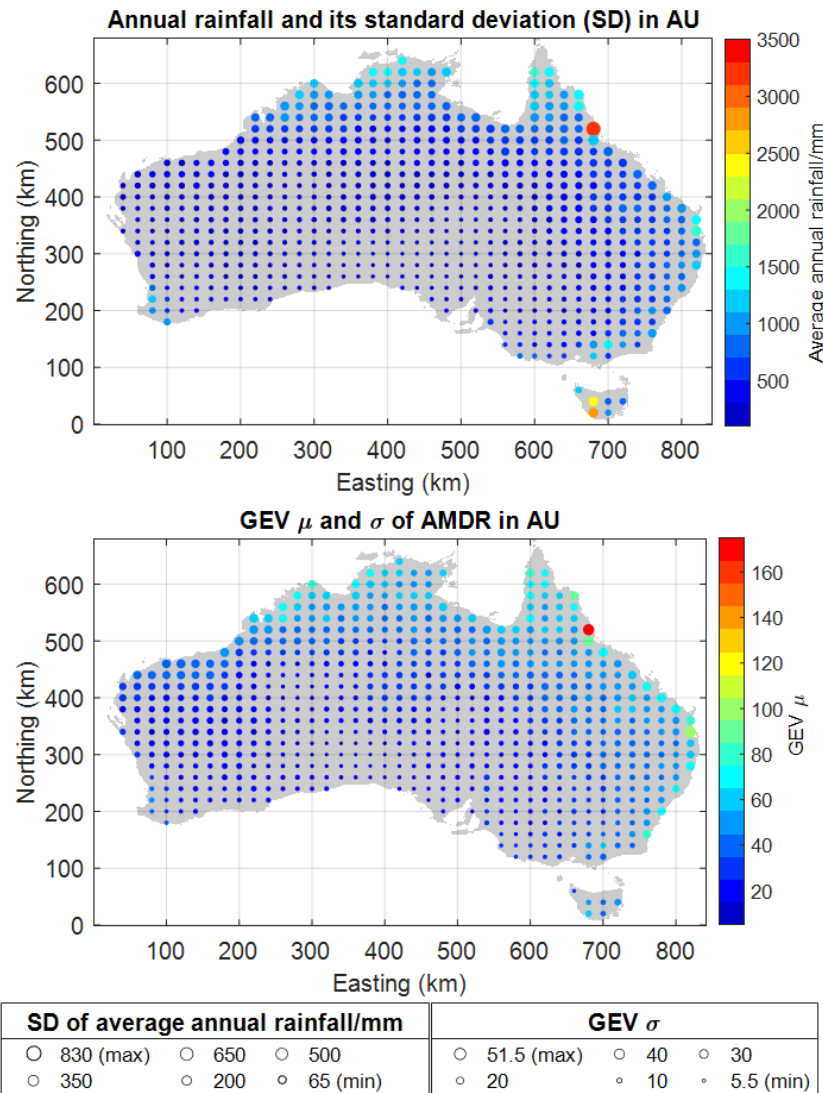
276 **5 Link between the spatial variations of GEV parameters and the climate variation**

277 The GEV distribution parameters can reveal the characteristics of extreme rainfall in
 278 terms of both its amount and occurrence probability. The two parameters are shown to have
 279 strong spatial dependency as discussed in the previous sections. To help understand how such
 280 spatial variation of the extreme rainfall is related to the climate variability over space, Figure 3 is
 281 produced to display the spatial distributions of both the average annual rainfall and its standard
 282 deviation of GB and AU respectively, compared with the spatial distribution of the GEV
 283 parameters. A great deal of similarity exists in space between the corresponding quantities, i.e.
 284 the location parameter versus the annual mean, and the scale parameter versus the standard
 285 deviation of the annual rainfall. For example, regions with higher annual average rainfall are not
 286 only presented with higher standard deviation (e.g. the circles located in west Scotland and west
 287 Wales of GB and in north-eastern coastal regions of AU, appearing more reddish and larger),
 288 they are also associated with higher values of the GEV parameters, and appear to be more
 289 heterogenous. This feature also exists in the regions with low and more even annual rainfall
 290 distribution, but works in an opposite way (e.g., circles located in middle and eastern England of
 291 GB and middle-north zone of AU are all more bluish and smaller). This findings is consistent
 292 with those published in the series of climate reports of GB (Kendon et al., 2015; Kendon et al.,
 293 2018, 2019) and AU (CSIRO & Australian Bureau of Meteorology, 2018).

Average annual rainfall and its standard deviation (SD) in GB GEV μ and σ of AMDR in GB



(a)



(b)

294 **Figure 3.** Comparison between the climatic variables (the average annual rainfall and its
 295 standard deviation) and GEV parameters μ and σ in GB (a) and AU (b) cases where the colour
 296 denotes the value of averaged annual rainfall or the GEV parameter μ , and the size of the circles
 297 denotes the value of standard deviation of the annual rainfall or the GEV parameter σ .

298 6 Conclusions

299 This paper presents a study on the spatial variation of extreme rainfall using two-century
 300 long datasets covering Great Britain and Australia. The annual maximum daily rainfall (AMDR)
 301 series extracted from regions of interest (ROI, 11,011 in total) with various spatial properties
 302 (location, size and shape), are individually fitted with GEV distributions whose parameters are
 303 then analysed over the space. Four generalised linear models (GLMs) are developed to quantify
 304 these variations by involving the effect from the geographical location, area size and shapes.
 305 From the results discussed previously, the following conclusions can be drawn:

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

324 These findings offer a new quantitative insight in understanding the spatial variation of
325 large-scale climatology of rainfall. Not only are they supported and consistent with many
326 previous studies on rainfall distributions, the quantification of the extreme rainfall and its spatial
327 dependencies are of great practical value in engineering design, e.g. designed rainfall/floods for
328 constructions. The methods employed by this study are specifically designed for large grid-
329 based datasets, and thus can be readily applied to climate projections for evaluating the spatial
330 heterogeneity of climate change impact, such as flooding and droughts. It should be noted that
331 the quality of the underlying datasets, which have undergone a series of quality control measures,
332 may still bring in large amount of uncertainties and should be addressed in further work. In
333 addition, impact of the density of the underlying data observations, i.e., rain-gauges, and its
334 variation over long term also need to be further studied.

335 **Acknowledgments**

336 The authors would like to thank the Centre of Hydrology and Ecology (CEH) and The Bureau of
337 Meteorology, Australia for providing the datasets. The open-source toolbox of spatial random
338 sampling for grid-based data analysis (SRS-GDA toolbox, doi: 10.5281/zenodo.4044626)
339 developed by the authors used in this study. This research is supported by the Chinese
340 Scholarship Council and the College of Engineering, Swansea University, UK via their PhD
341 scholarships offered to the co-author Han Wang and the Royal Academy of Engineering UK-
342 China Urban Flooding Programme Grant (REF: UUFRIPI\10021), which are gratefully
343 acknowledged.

344 **References**

- 345 Banwell, N., Rutherford, S., Mackey, B., Street, R., & Chu, C. (2018). Commonalities between
346 disaster and climate change risks for health: A theoretical framework. *International*
347 *journal of environmental research and public health*, 15(3), 538.

- 348 Buishand, T. A., De Haan, L., & Zhou, C. (2008). On spatial extremes: with application to a
349 rainfall problem. *The Annals of Applied Statistics*, 2(2), 624-642.
- 350 Buytaert, W., Celleri, R., Willems, P., De Bievre, B., & Wyseure, G. (2006). Spatial and
351 temporal rainfall variability in mountainous areas: A case study from the south
352 Ecuadorian Andes. *Journal of hydrology*, 329(3-4), 413-421.
- 353 Chandler, R. E., & Wheeler, H. S. (2002). Analysis of rainfall variability using generalized linear
354 models: a case study from the west of Ireland. *Water Resources Research*, 38(10), 10-11-
355 10-11.
- 356 Coe, R., & Stern, R. D. (1982). Fitting models to daily rainfall data. *Journal of Applied*
357 *Meteorology*, 21(7), 1024-1031.
- 358 CSIRO, & Australian Bureau of Meteorology (2018). Previous State of the Climate reports.
359 Retrieved from [https://www.csiro.au/en/Research/OandA/Areas/Assessing-our-](https://www.csiro.au/en/Research/OandA/Areas/Assessing-our-climate/State-of-the-Climate-2018/Previous-State-of-the-Climate-reports)
360 [climate/State-of-the-Climate-2018/Previous-State-of-the-Climate-reports](https://www.csiro.au/en/Research/OandA/Areas/Assessing-our-climate/State-of-the-Climate-2018/Previous-State-of-the-Climate-reports).
- 361 Feng, S., Nadarajah, S., & Hu, Q. (2007). Modeling annual extreme precipitation in China using
362 the generalized extreme value distribution. *Journal of the Meteorological Society of*
363 *Japan. Ser. II*, 85(5), 599-613.
- 364 Hosking, J. R. M., Wallis, J. R., & Wood, E. F. (1985). Estimation of the generalized extreme-
365 value distribution by the method of probability-weighted moments. *Technometrics*, 27(3),
366 251-261.
- 367 James, G. M. (2002). Generalized linear models with functional predictors. *Journal of the Royal*
368 *Statistical Society: Series B (Statistical Methodology)*, 64(3), 411-432.
- 369 Jones, D. A., Wang, W., & Fawcett, R. (2009). High-quality spatial climate data-sets for
370 Australia. *Australian Meteorological and Oceanographic Journal*, 58(4), 233.
- 371 Jung, Y., Shin, J. Y., Ahn, H., & Heo, J. H. (2017). The spatial and temporal structure of extreme
372 rainfall trends in South Korea. *Water*, 9(10), 809.
- 373 Kendon, E. J., Fosser, G., Murphy, J., Chan, S., Clark, R., Harris, G., Lock, A., Lowe, J., Martin,
374 G., & Pirret, J. (2019). UKCP Convection-permitting model projections: Science report.
- 375 Kendon, M., MacCarthy, M., & Jevrejeva, S. (2015). State of the UK Climate 2014: Met Office.
- 376 Kendon, M., McCarthy, M., Jevrejeva, S., Matthews, A., & Legg, T. (2018). State of the UK
377 climate 2017. *International Journal of Climatology*, 38, 1-35.
- 378 Kendon, M., McCarthy, M., Jevrejeva, S., Matthews, A., & Legg, T. (2019). State of the UK
379 climate 2018. *International Journal of Climatology*, 39, 1-55.
- 380 Li, Z., Li, Z., Zhao, W., & Wang, Y. (2015). Probability modeling of precipitation extremes over
381 two river basins in northwest of China. *Advances in Meteorology*, 2015.
- 382 Lowe, J. A., Bernie, D., Bett, P., Bricheno, L., Brown, S., Calvert, D., Clark, R., Eagle, K.,
383 Edwards, & T., Fosser, G. (2018). UKCP18 science overview report. Met Office Hadley
384 Centre: Exeter, UK.
- 385 Martins, E. S., & Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-
386 value quantile estimators for hydrologic data. *Water Resources Research*, 36(3), 737-744.

- 387 McCullagh, P. (1989). *Generalized linear models*: Routledge.
- 388 McCullagh, P. (2018). *Generalized linear models*: Routledge.
- 389 Millán, M. M., Estrela, M. J., & Miró, J. (2005). Rainfall components: variability and spatial
390 distribution in a Mediterranean Area (Valencia Region). *Journal of Climate*, *18*(14),
391 2682-2705.
- 392 Overeem, A., Buishand, T. A., Holleman, I., & Uijlenhoet, R. (2010). Extreme value modeling
393 of areal rainfall from weather radar. *Water Resources Research*, *46*(9).
- 394 Pedersen, L., Jensen, N. E., Christensen, L. E., & Madsen, H. (2010). Quantification of the
395 spatial variability of rainfall based on a dense network of rain gauges. *Atmospheric*
396 *research*, *95*(4), 441-454.
- 397 Peleg, N., Marra, F., Fatichi, S., Paschalis, A., Molnar, P., & Burlando, P. (2018). Spatial
398 variability of extreme rainfall at radar subpixel scale. *Journal of hydrology*, *556*, 922-933.
- 399 Pierce, D. A., & Schafer, D. W. (1986). Residuals in generalized linear models. *Journal of the*
400 *American Statistical Association*, *81*(396), 977-986.
- 401 Prein, A. F., Rasmussen, R. M., Ikeda, K., Liu, C., Clark, M. P., & Holland, G. J. (2017). The
402 future intensification of hourly precipitation extremes. *Nature Climate Change*, *7*(1), 48-
403 52.
- 404 Rashid, M., Beecham, S., Chowdhury, R., Piantadosi, J., Anderssen, R. S., & Boland, J. (2013).
405 Simulation of extreme rainfall from CMIP5 in the Onkaparinga catchment using a
406 generalized linear model.
- 407 Rogger, M., Agnoletti, M., Alaoui, A., Bathurst, J. C., Bodner, G., Borga, M., Chaplot, V.,
408 Gallart, F., Glatzel, G., & Hall, J. (2017). Land use change impacts on floods at the
409 catchment scale: Challenges and opportunities for future research. *Water Resources*
410 *Research*, *53*(7), 5209-5219.
- 411 Schaefer, M. G. (1990). Regional analyses of precipitation annual maxima in Washington State.
412 *Water Resources Research*, *26*(1), 119-131.
- 413 Segond, M. L., Onof, C., & Wheeler, H. S. (2006). Spatial-temporal disaggregation of daily
414 rainfall from a generalized linear model. *Journal of hydrology*, *331*(3-4), 674-689.
- 415 Stern, R. D., & Coe, R. (1984). A model fitting analysis of daily rainfall data. *Journal of the*
416 *Royal Statistical Society: Series A (General)*, *147*(1), 1-18.
- 417 Tanguy, M., Dixon, H., Prosdocimi, I., Morris, D. G., & Keller, V. D. J. (2016). Gridded
418 estimates of daily and monthly areal rainfall for the United Kingdom (1890-2015)[CEH-
419 GEAR]. NERC Environmental Information Data Centre.
- 420 Villarini, G., Smith, J. A., Baeck, M. L., Sturdevant-Rees, P., & Krajewski, W. F. (2010). Radar
421 analyses of extreme rainfall and flooding in urban drainage basins. *Journal of hydrology*,
422 *381*(3-4), 266-286.
- 423 Wang, H., & Xuan, Y. (2020). SRS-GDA: A spatial random sampling toolbox for grid-based
424 hydro-climatic data analysis in environmental change studies. *Environmental Modelling*
425 *& Software*, *124*, 104598.

- 426 Wang, P. (1987). Residual plots for detecting nonlinearity in generalized linear models.
 427 *Technometrics*, 29(4), 435-438.
- 428 Westra, S., Alexander, L. V., & Zwiers, F. W. (2013). Global increasing trends in annual
 429 maximum daily precipitation. *Journal of Climate*, 26(11), 3904-3918.
- 430 Westra, S., Fowler, H. J., Evans, J. P., Alexander, L. V., Berg, P., Johnson, F., Kendon, E. J.,
 431 Lenderink, G., & Roberts, N. M. (2014). Future changes to the intensity and frequency of
 432 short-duration extreme rainfall. *Reviews of Geophysics*, 52(3), 522-555.
- 433 Yan, Z., Bate, S., Chandler, R. E., Isham, V., & Wheater, H. (2002). An analysis of daily
 434 maximum wind speed in northwestern Europe using generalized linear models. *Journal*
 435 *of Climate*, 15(15), 2073-2088.
- 436 Yang, C., Chandler, R. E., Isham, V. S., & Wheater, H. S. (2005). Spatial-temporal rainfall
 437 simulation using generalized linear models. *Water Resources Research*, 41(11).
- 438 Yoon, P., Kim, T. W., & Yoo, C. (2013). Rainfall frequency analysis using a mixed GEV
 439 distribution: a case study for annual maximum rainfalls in South Korea. *Stochastic*
 440 *Environmental Research and Risk Assessment*, 27(5), 1143-1153.
- 441 Zheng, Y., Xue, M., Li, B., Chen, J., & Tao, Z. (2016). Spatial characteristics of extreme rainfall
 442 over China with hourly through 24-hour accumulation periods based on national-level
 443 hourly rain gauge data. *Advances in Atmospheric Sciences*, 33(11), 1218-1232.