Maximum daily rainfall in South Korea

Saralees Nadarajah[1] and Dongseok Choi[2]

[1]*School of Mathematics, University of Manchester, Manchester M60 1QD, UK.*
e-mail: saralees.nadarajah@manchester.ac.uk
[2]*Department of Public Health and Preventive Medicine, Oregon Health & Science University,*
Portland, Oregon 97201-3098, USA.

Annual maxima of daily rainfall for the years 1961–2001 are modeled for five locations in South Korea (chosen to give a good geographical representation of the country). The generalized extreme value distribution is fitted to data from each location to describe the extremes of rainfall and to predict its future behavior. We find evidence to suggest that the Gumbel distribution provides the most reasonable model for four of the five locations considered. We explore the possibility of trends in the data but find no evidence suggesting trends. We derive estimates of 10, 50, 100, 1000, 5000, 10,000, 50,000 and 100,000 year return levels for daily rainfall and describe how they vary with the locations. This paper provides the first application of extreme value distributions to rainfall data from South Korea.

1. Introduction

Extreme value theory deals with the stochastic behavior of the extreme values in a process. For a single process, the behavior of the maxima can be described by the three extreme value distributions – Gumbel, Fréchet and negative Weibull – as suggested by Fisher and Tippett (1928). Kotz and Nadarajah (2000) indicated that the extreme value distributions could be traced back to the work done by Bernoulli in 1709. The first application of extreme value distributions was probably made by Fuller in 1914. Thereafter, several researchers have provided useful applications of extreme value distributions to rainfall data from different regions of the world: see Nguyen *et al* (1998, 2002) for applications in Canada; Koutsoyiannis and Baloutsos (2000) for applications in Greece; Ferro (1993) and Parida (1999) for applications in India; Cannarozzo *et al* (1995), Aronica *et al* (2002) and Crisci *et al* (2002) for applications in Italy; Elnaqa and Abuzeid (1993) for applications in Jordan; Zalina *et al* (2002) for applications in Malaysia; Withers and Nadarajah (2000)

for applications in New Zealand; and, Miroslava (1991, 1992) for applications in Yugoslavia. For a review of applications of extreme value distributions to climate data see Farago and Katz (1990).

The only work known to us on extremes of Korean rainfall are that of Park *et al* (2001) and Park and Jung (2002). Park *et al* (2001) modeled the summer extreme rainfall data (time series of annual maximum of daily and 2-day precipitation) at 61 gauging stations over South Korea by using the Wakeby distribution with the method of *L*-moments estimates. Park and Jung (2002) modeled the same data using the Kappa distribution with the maximum likelihood estimates. The main drawback with these approaches is the use of distributions which are not extreme value distributions. Theoretically, there is no justification to model annual maximum daily rainfall by either the Wakeby or the Kappa distribution (although, in practice, they may provide a reasonable fit). In this paper, we provide the first application of extreme value distributions to model rainfall data from South Korea.

Traditionally, the three extreme value distributions are applied to annual maximum daily rainfall. In this paper we use the generalized extreme value (GEV) distribution, which has all the flexibility of the three extreme value distributions – Gumbel, Fréchet and negative Weibull. The GEV distribution was developed by Jenkinson (1955) – see also Hosking *et al* (1985) and Galambos (1987). The cumulative distribution function (cdf) of the GEV distribution is:

$$F(x) = \exp\left\{ - \left(1 + \xi \frac{x - \mu}{\sigma} \right)^{-1/\xi} \right\}, \quad (1)$$

for $1 + \xi(x-\mu)/\sigma > 0$, where μ, σ and ξ are referred to as the location, scale and shape parameters, respectively. The particular case of (1) for $\xi = 0$,

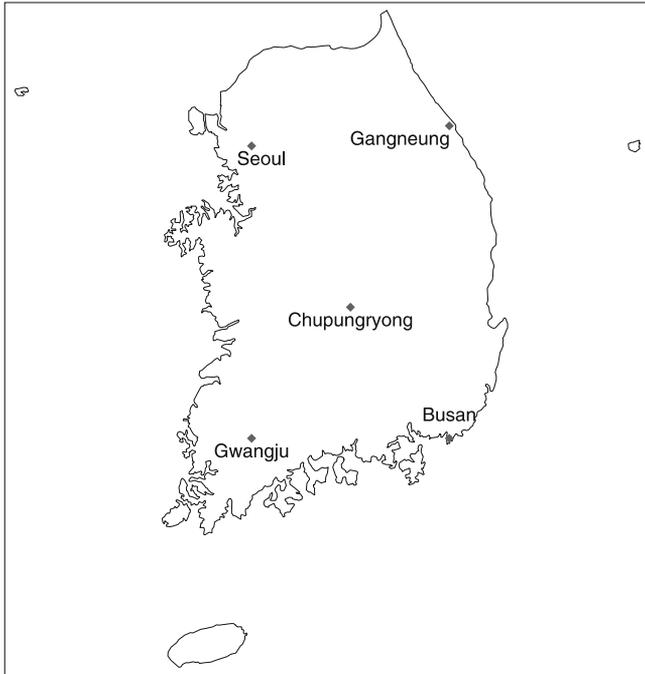$$F(x) = \exp\left\{ - \exp\left(-\frac{x - \mu}{\sigma} \right) \right\}, \quad (2)$$

(for $-\infty < x < \infty$) is the Gumbel distribution, while the cases $\xi > 0$ and $\xi < 0$ are known as the Fréchet and the negative Weibull distributions, respectively. For rainfall data, usually $\xi > 0$, although sometimes the Gumbel distribution is adequate.

The paper is organized as follows. Five carefully chosen locations in South Korea and the corresponding datasets are described in section 2. The data used are the annual maximum daily rainfall from 1961 to 2001. The models and the fitting procedures are described in section 3. We use the method of maximum likelihood for estimation and the profile likelihood method for the corresponding confidence intervals. Finally, the results of the fitted models and their implications are discussed in section 4. In particular, estimates are given for the 10, 50, 100, 1000, 5000, 10000, 50000 and 100,000 year return levels.

2. Data

The data consist of annual maximum daily rainfall for the years from 1961 to 2001 for the five locations pointed out in figure 1. The data were extracted from the website http://www.kma.go.kr/ of Korea Meteorological Organization, which lists the daily and monthly rainfall figures for 77 locations in South Korea. But there are only 15 sites that have data going back to 1961 (the earliest year for which data are available). Our analyses are limited to 5 of these sites. These five locations – Seoul in the west, Gangneung in the east, Busan in the south-east, Gwangju in the south-west and Chupungryong in the middle – were chosen carefully to give a good geographical representation of the country. Table 1 gives the latitude and the longitude of the five locations and some summary statistics of the corresponding data sets.

Figures 2 to 4 show how the annual maximum daily rainfall has varied from 1961 to 2001 for the five locations. For the sake of comparison we have used the same scale. It is evident that Chupungryong has the least annual maximum daily rainfall.

Figure 1.   The five locations.

Table 1. *Locations and some summary statistics.*

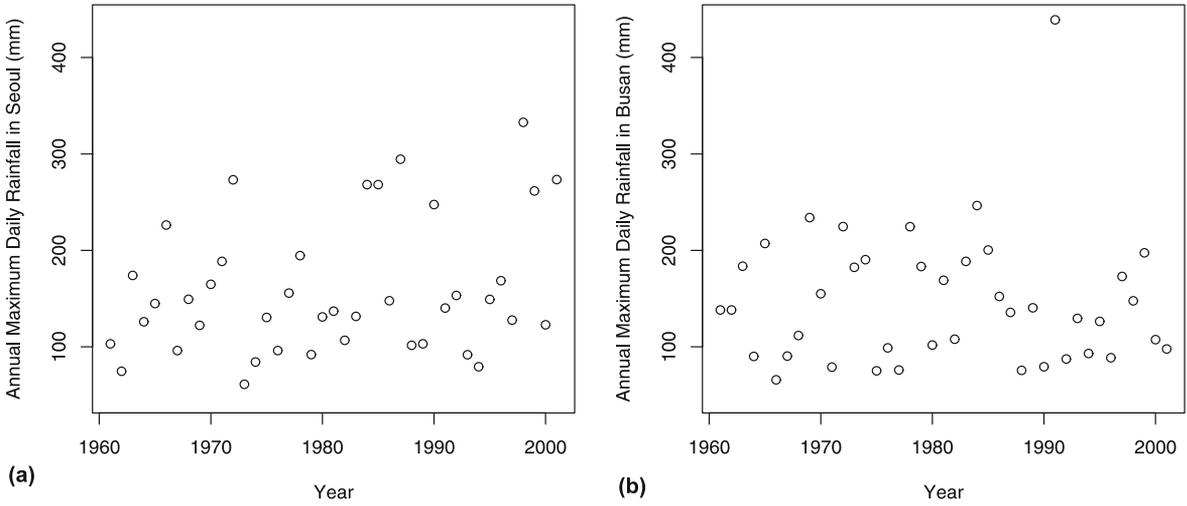| Location	Latitude	Longitude	Mean	Standard deviation	Minimum	Maximum
Seoul	127°00′	37°30′	158.4	68.9	61.2	332.8
Busan	129°00′	35°05′	147.2	69.3	65.8	439.0
Gwangju	127°00′	35°05′	125.1	52.9	47.1	335.6
Gangneung	129°00′	37°40′	142.0	63.7	66.8	297.5
Chupungryong	128°00′	36°10′	104.1	39.1	50.4	215.4

Figure 2. Annual maxima daily rainfall recorded in **(a)** Seoul and **(b)** Busan for 1961 to 2001.

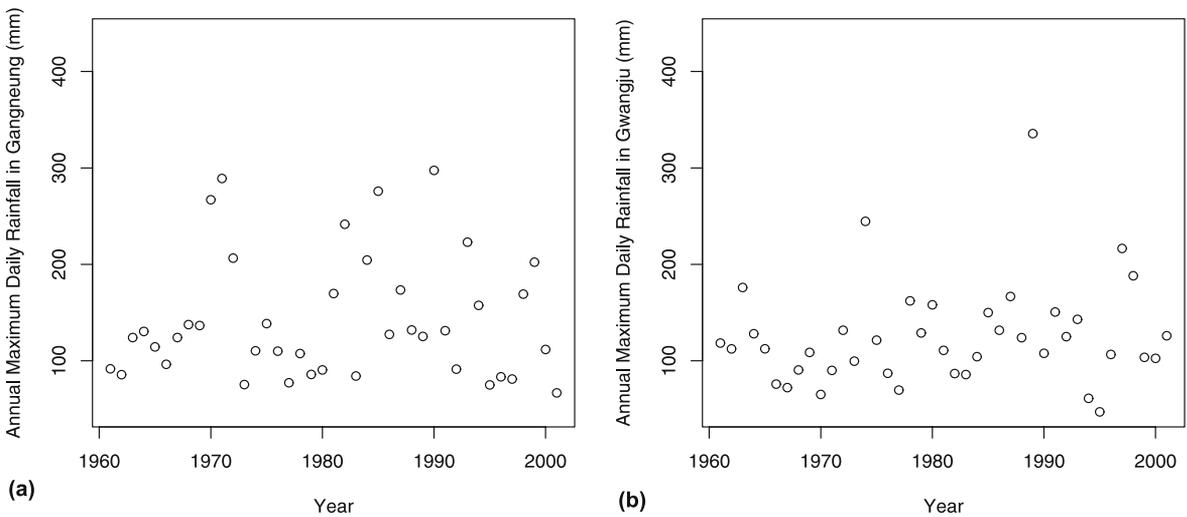Figure 3. Annual maxima daily rainfall recorded in **(a)** Gangneung and **(b)** Gwangju for 1961 to 2001.

Figure 4. Annual maxima daily rainfall recorded in Chupungryong for 1961 to 2001.

3. Methodology

Suppose $x_1, \ldots, x_n$ denote annual maximum daily rainfall for n years from a given location. The method of maximum likelihood was used to fit (1) to these data. Assuming independence of the data, the likelihood is the product of the densities of (1) for the observations $x_1, \ldots, x_n$. Mathematically,

$$L(\mu, \sigma, \xi) = \frac{1}{\sigma^n} \prod_{i=1}^{n} \left(1 + \xi \frac{x_i - \mu}{\sigma} \right)^{-(1/\xi+1)}$$

$$\times \exp \left\{ -\sum_{i=1}^{n} \left(1 + \xi \frac{x_i - \mu}{\sigma} \right)^{-1/\xi} \right\}. \quad (3)$$
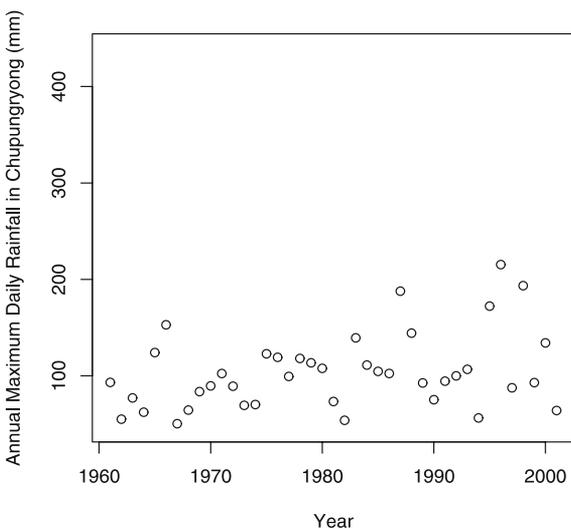
The estimates of μ, σ and ξ – say $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$ – are taken to be those values which maximize the

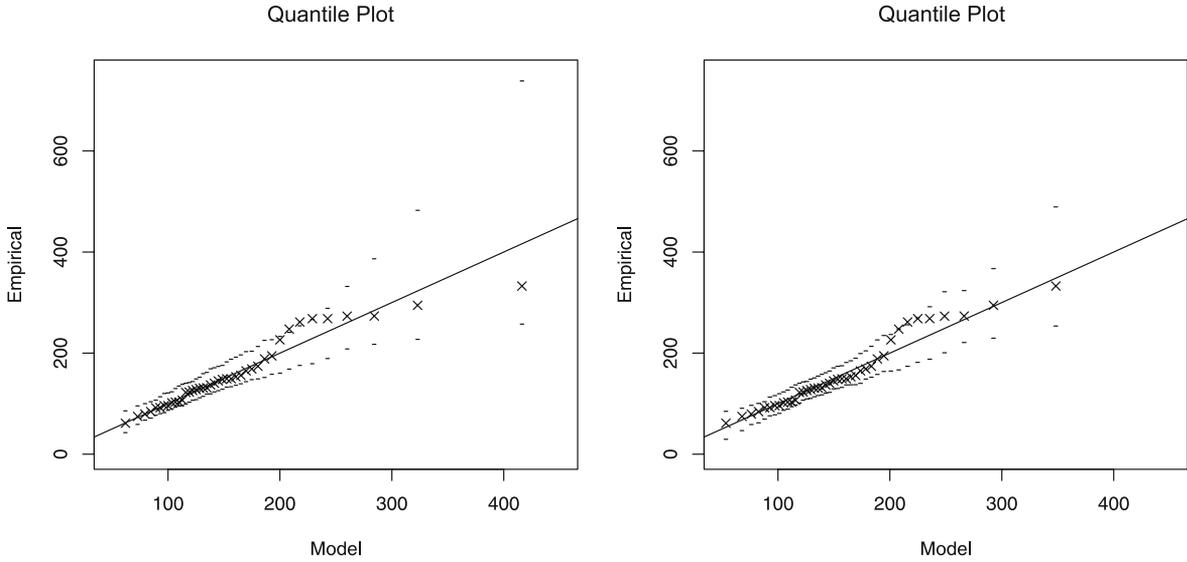Quantile Plot   Quantile Plot

Figure 5.   QQ plots with simulated 95% confidence intervals for model 1 (left) and model 2 (right) for Seoul.

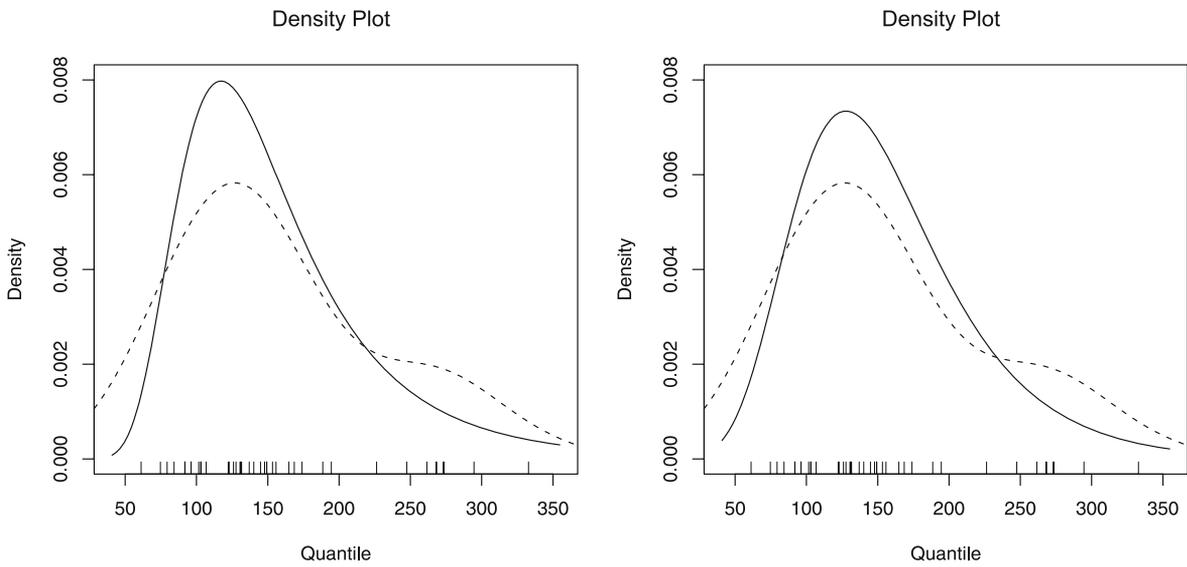Density Plot   Density Plot

Figure 6.   Fitted (solid) as well as the non-parametric (broken) densities for model 1 (left) and model 2 (right) for Seoul.

likelihood L. This maximization was performed using a quasi-Newton iterative algorithm. The standard errors of the estimates were computed by inverting the Fisher information matrix (Prescott and Walden 1980).

The basic model fitted was (1) with μ, σ, and ξ constant (to be referred to as model 1). As mentioned in section 1, sometimes the Gumbel distribution gives as good a fit as (1) for rainfall data, so we also fitted (2) with μ and σ constant (to be referred to as model 2).

Model 2 is a submodel of model 1, so a standard way of determining the best fit model is the likelihood ratio test (Wald 1943). If L_1 is the maximum likelihood for the three-parameter model 1 and L_2 is the maximum likelihood for the two-parameter model 2, then under the simpler model the test statistic $\lambda = -2\log(L_2/L_1)$ would be assumed to be distributed as a chi-square variable with 1 degree of freedom (since the numbers of parameters differ by 1). In hypothesis testing problems this would be asymptotically true as the number of data tends to infinity. Thus, at the 5% significance level, the simpler two-parameter model would be preferred if $-2\log(L_2/L_1) < \chi^2_{1,0.95} = 3.841$. In practice, because of the lack of complete independence of the annual maxima, this would probably have to be interpreted conservatively.

Figures 2 to 4 suggest that extreme rainfall could possibly exhibit some trends with respect to time. To investigate this, we tried the following variations of models 1 and 2:

Model 3: $\mu = a + b \times \text{Year}$, $\sigma = \text{constant}$,

$\xi = \text{constant}$,

a four-parameter model with μ allowed to vary linearly with respect to time, and

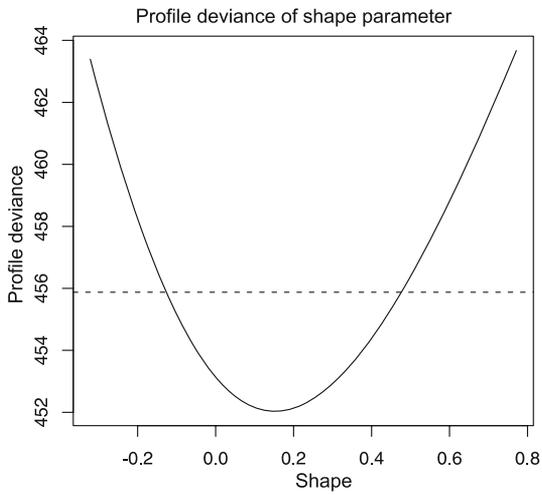Model 4: $\quad \mu = c + d \times$ Year, $\quad \sigma =$ constant, $\xi = 0,$

Figure 7. Profile deviance for the ξ parameter in model 1 for Seoul.

a three-parameter model with μ allowed to vary linearly with respect to time. The standard likelihood ratio test was used to determine whether the trends described by the models 3 and 4 are significant or not.

The goodness of fit of these models was examined by QQ and density plots. A QQ plot is where the observed quantile is plotted against the quantile predicted by the fitted model. For example, to check the goodness of fit of model 1, we would plot the sorted values (in the ascending order) of the observed annual maximum daily rainfall versus expected quantiles y_i determined by $F(y_i) = (i - 0.375)/(n + 0.25)$, where F is given by (1). Similarly to check the goodness of fit of model 2, we would plot the sorted values of the observed annual maximum daily rainfall versus expected quantiles y_i determined by $F(y_i) = (i - 0.375)/(n + 0.25)$, where F is given by (2). A density plot is one that compares the fitted density of the model with a non-parametric version computed directly from the annual maximum daily data. The fitted densities for

Table 2. *Best fitted models and parameter estimates.*

| Location | Best model | Parameter estimates | | | K–S test |
		$\hat{\mu}$ (s.e.)	$\hat{\sigma}$ (s.e.)	$\hat{\xi}$ (s.e.)	p value
Seoul	Model 2	127.8 (8.2)	50.1 (6.4)		0.041
Busan	Model 2	118.4 (7.7)	46.8 (6.0)		0.048
Gwangju	Model 2	103.3 (6.0)	36.4 (4.5)		0.042
Gangneung	Model 1	107.0 (6.9)	36.1 (6.0)	0.345 (0.188)	0.042
Chupungryong	Model 2	86.8 (4.8)	29.1 (3.6)		0.038

Table 3. *Return level estimates for $T = 10, 50, 100$.*

| Location | Return level x_T (95% confidence interval) | | |
	$T = 10$	$T = 50$	$T = 100$
Seoul	240.9 (208.8, 283.5)	323.8 (275.1, 389.9)	359.5 (302.9, 435.2)
Busan	224.0 (194.0, 263.6)	301.5 (255.8, 362.9)	331.8 (281.9, 405.1)
Gwangju	185.5 (162.7, 215.6)	245.9 (211.5, 292.3)	275.1 (232.0, 325.2)
Gangneung	227.7 (185.2, 351.7)	317.9 (−)	374.5 (291.1, 982.5)
Chupungryong	152.3 (134.0, 176.8)	200.2 (172.8, 238.2)	220.7 (189.0, 264.3)

Table 4. *Return level estimates for $T = 500, 1000, 5000$.*

| Location | Return level x_T (95% confidence interval) | | |
	$T = 500$	$T = 1000$	$T = 5000$
Seoul	451.6 (366.6, 541.3)	489.8 (393.8, 587.0)	554.2 (458.4, 689.9)
Busan	434.2 (339.1, 508.1)	473.6 (363.8, 552.6)	560.5 (421.5, 654.2)
Gwangju	340.9 (278.6, 402.0)	366.1 (298.9, 434.4)	421.0 (346.1, 509.2)
Gangneung	456.2 (346.0, 1320.1)	469.8 (−)	541.2 (395.6, 1569.1)
Chupungryong	268.0 (226.5, 324.8)	288.6 (242.5, 350.9)	334.3 (279.8, 411.4)

Table 5. *Return level estimates for $T = 10000, 50000, 100000$.*

	Return level x_T (95% confidence interval)		
Location	$T = 10,000$	$T = 50,000$	$T = 100,000$
Seoul	615.3 (484.0, 739.0)	689.9 (548.9, 842.2)	722.4 (576.3, 886.8)
Busan	599.3 (445.9, 698.6)	636.4 (510.7, 784.9)	704.9 (532.3, 835.7)
Gwangju	452.7 (365.9, 542.9)	533.3 (409.6, 625.2)	564.4 (428.8, 650.0)
Gangneung	590.7 (414.3, 487.8)	650.5 (–)	685.1 (–)
Chupungryong	357.1 (295.7, 437.6)	409.3 (332.8, 498.6)	427.2 (348.8, 524.4)

models 1 and 2 would be given by

$$f(x) = \frac{1}{\hat{\sigma}} \left(1 + \hat{\xi} \frac{x - \hat{\mu}}{\hat{\sigma}} \right)^{-\left(1/\hat{\xi} + 1\right)}$$

$$\times \exp \left\{ - \left(1 + \hat{\xi} \frac{x - \hat{\mu}}{\hat{\sigma}} \right)^{-1/\hat{\xi}} \right\},$$

and

$$f(x) = \frac{1}{\hat{\sigma}} \exp \left(-\frac{x - \hat{\mu}}{\hat{\sigma}} \right)$$

$$\times \exp \left\{ - \exp \left(-\frac{x - \hat{\mu}}{\hat{\sigma}} \right) \right\},$$

respectively. The non-parametric estimate was computed by using the kernel method (Silverman 1986).

Once the best model for the data has been determined, the interest is in deriving the return levels of rainfall. The T year return level, say x_T, is the level exceeded on average only once in T years. For example, the 2-year return level is the median of the distribution of the annual maximum daily rainfall. If model 1 is assumed then on inverting $F(x_T) = 1 - 1/T$ we get the expression

$$x_T = \mu - \frac{\sigma}{\xi} \left\{ 1 - \left[-\log \left(1 - \frac{1}{T} \right) \right]^{-\xi} \right\}. \quad (4)$$

If on the other hand, model 2 is assumed then the corresponding expression is:

$$x_T = \mu - \sigma \log \left\{ -\log \left(1 - \frac{1}{T} \right) \right\}. \quad (5)$$

On substituting $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$ into (4)–(5), we have the maximum likelihood estimates of the return levels.

Confidence intervals for return level estimates are usually based on the delta method (Rao 1973) or resampling techniques. Here we employ the profile likelihood method, which is usually regarded superior to any of the existing methods. The following is a brief description of the method: reparameterize the likelihood (3) in terms of (x_T, σ, ξ) and calculate the profile likelihood $L_p(x_T)$ given by

$$L_p(x_T) = \max_{\sigma, \xi} L(x_T, \sigma, \xi).$$

Then the $100(1 - \alpha)\%$ confidence interval for x_T is given by the set

$$\left\{ \theta : 2 \log \left(\frac{L(\hat{\mu}, \hat{\sigma}, \hat{\xi})}{L_p(\theta)} \right) < \chi^2_{1,1-\alpha} \right\}, \quad (6)$$

where $\chi^2_{1,1-\alpha}$ denotes the $100(1 - \alpha)\%$ quantile of the chi-squared distribution with one degree of freedom. We would refer to the term on the left of the inequality in (6) as the profile deviance for x_T.

4. Results and discussion

Models 1 to 4 were fitted to annual maxima daily rainfall from each of the five locations. The method of maximum likelihood was used throughout. We shall let L_i denote the maximized likelihood for model i for $i = 1, 2, 3, 4$.

For Seoul data, model 1 gave the estimates $\hat{\mu} = 123.8$, $\hat{\sigma} = 46.6$ and $\hat{\xi} = 0.153$ with $-2 \log L_1 = 452.0$. On the other hand, model 2 gave the estimates $\hat{\mu} = 127.8$ and $\hat{\sigma} = 50.1$ with $-2 \log L_2 = 453.1$. Since $-2 \log(L_2/L_1) = 453.1 - 452.0 < 3.841 = \chi^2_{1,0.95}$ it follows by the standard likelihood ratio test that model 2 should be preferred to model 1 (in other words, the Gumbel distribution provides as good a fit as the more general GEV distribution). This finding is supported by figures 5 and 6, where we have plotted the QQ and density plots for models 1 and 2. There is little to choose between the two QQ plots and the two density plots. If anything, model 2 appears to give a better fit in the

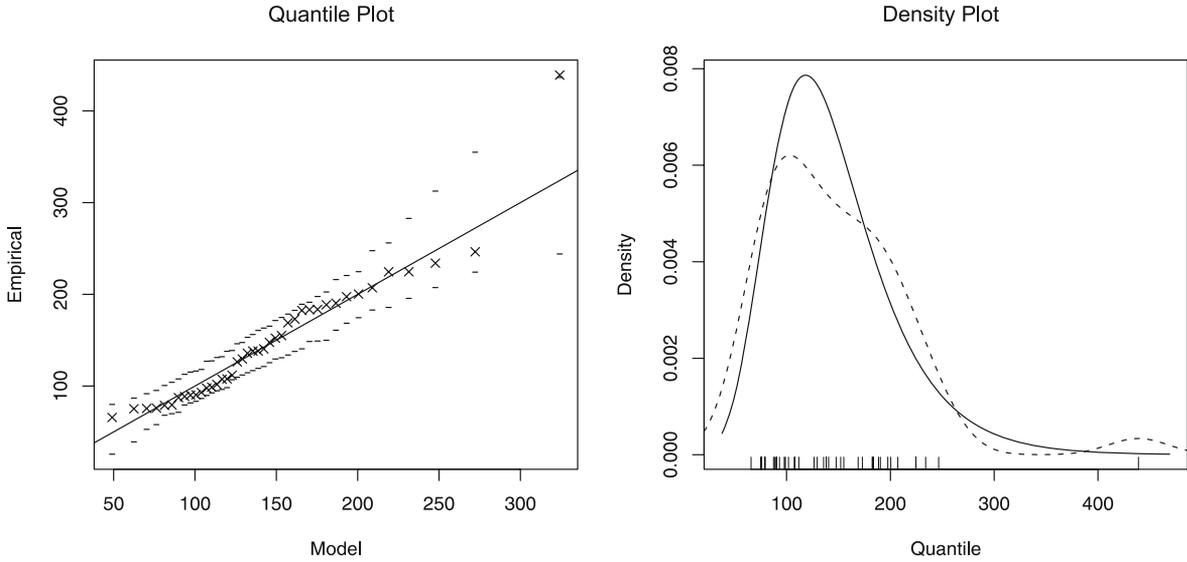Quantile Plot

Density Plot

Figure 8. Diagnostics for model 2 for Busan: QQ plot with simulated 95% confidence intervals (left); fitted (solid) as well as the non-parametric (broken) densities (right).
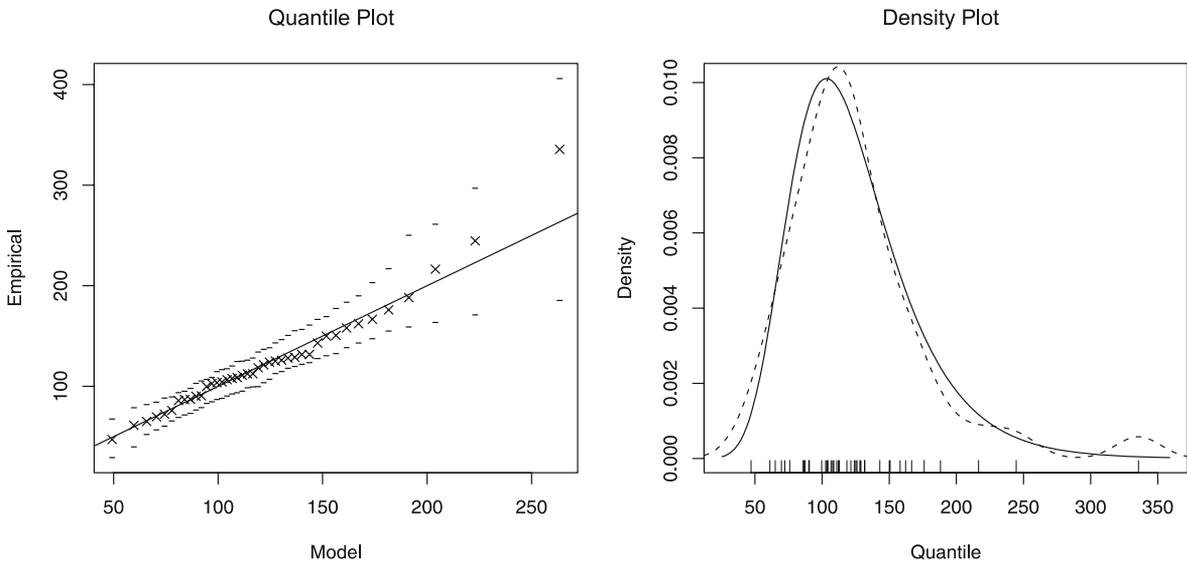
Quantile Plot

Density Plot

Figure 9. Diagnostics for model 2 for Gwangju: QQ plot with simulated 95% confidence intervals (left); fitted (solid) as well as the non-parametric (broken) densities (right).

upper tail area (which is the area of most interest). A further support is provided by the profile confidence interval for ξ displayed in figure 7 – the value $\xi = 0$ is well contained within the confidence interval.

When models 3 and 4 were fitted to Seoul data, we obtained $-\log L_3 = 450.7$ and $-\log L_4 = 451.6$. Since both $-2\log(L_1/L_3) = 452.0 - 450.7 < 3.841$ and $-2\log(L_2/L_4) = 453.1 - 451.6 < 3.841$ we see no evidence to suggest significant trends with respect to time. We also fitted the following variations of models 1 and 2:

Model 5: $\mu = \text{constant}$,

$\sigma = \exp(a + b \times \text{Year})$, $\xi = \text{constant}$,

Model 6: $\mu = \text{constant}$,

$\sigma = \exp(a + b \times \text{Year})$, $\xi = 0$,

Model 7: $\mu = a + b \times \text{Year}$,

$\sigma = \exp(c + d \times \text{Year})$, $\xi = \text{constant}$,

and

Model 8: $\mu = a + b \times \text{Year}$,

$\sigma = \exp(c + d \times \text{Year})$, $\xi = 0$.

None of these provided a significant fit. The non-parametric densities in figure 6 appear to show a small second peak. To check the significance of
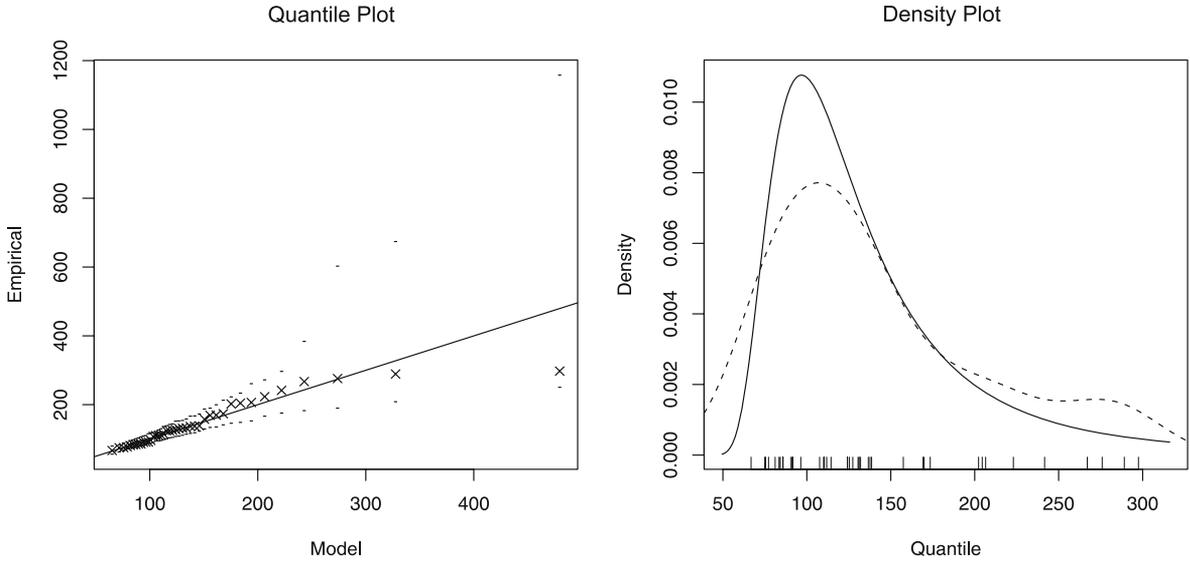
Quantile Plot

Density Plot

Figure 10.   Diagnostics for model 1 for Gangneung: QQ plot with simulated 95% confidence intervals (left); fitted (solid) as well as the non-parametric (broken) densities (right).
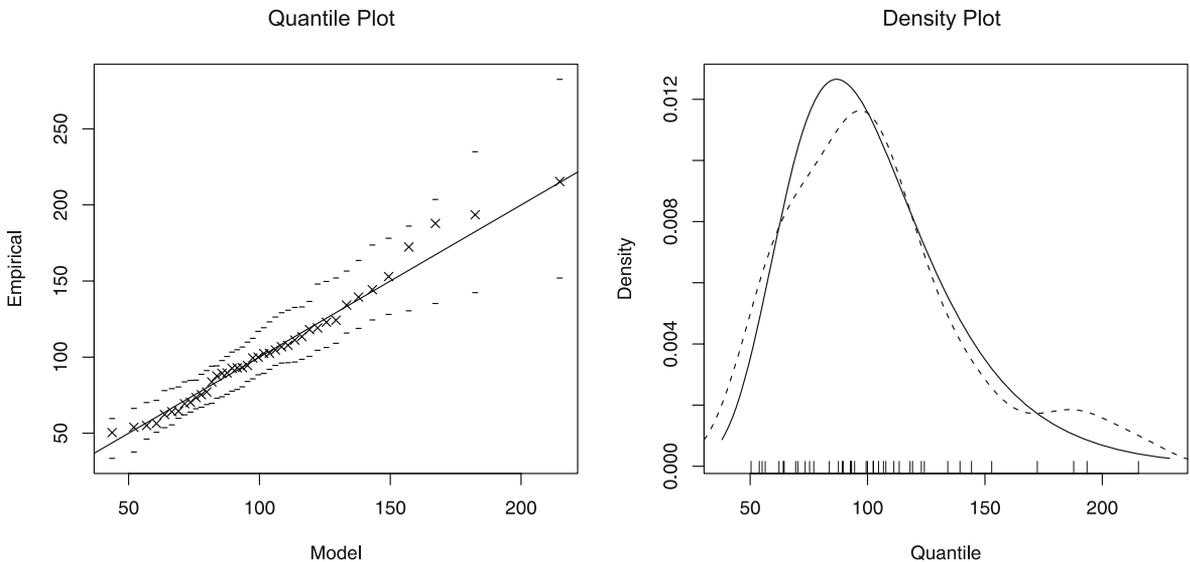
Quantile Plot

Density Plot

Figure 11.   Diagnostics for model 2 for Chupungryong: QQ plot with simulated 95% confidence intervals (left); fitted (solid) as well as the non-parametric (broken) densities (right).

this, we fitted the following two mixture variation of model 2:

$$F(x) = \alpha \exp \left\{ - \exp \left(- \frac{x - \mu_1}{\sigma_1} \right) \right\}$$

$$+ (1 - \alpha) \exp \left\{ - \exp \left(- \frac{x - \mu_2}{\sigma_2} \right) \right\}$$

for $0 < \alpha < 1$. Again, the fit was not significant. Hence, we conclude that the most reasonable model for extreme rainfall in Seoul is model 2 (the Gumbel distribution).

The above analyses were repeated for the other four locations. Table 2 summarizes the findings. It gives the best model chosen for each location, its parameter estimates and standard errors. We note that there is no significant evidence of trend for any of the five locations considered. Model 2 (Gumbel distribution) appears to give the best description for four of the five locations. The QQ and density plots for the best fitting models are displayed in figures 5, 6, 8, 9, 10 and 11. The horizontal bars on the QQ plots represent simulated 95% confidence intervals. Evidently the chosen models provide good fits. The last column of table 2 provides the p values of the Kolmogorov–Smirnov test of the best fitting

models. These p values are a final confirmation of the goodness of fit of the chosen models.

Tables 3–5 give estimates of the return level x_T for the best fitting models for $T = 10, 50, 100, 500, 1000, 5000, 10,000, 50,000, 100,000$ years. The estimates and the associated 95% confidence intervals are computed using (4)–(5) and for the profile likelihood method, see (6). For a few cases (annotated by (–)) confidence intervals are not given because of regularity problems to do with the inversion of the likelihood.

The estimates for 10 and 100-year return levels above are very consistent with those given by Park and Jung (2002, page 62). Among the locations considered, the capital Seoul in the west appears to be associated with the highest return levels. This finding is again consistent with Park and Jung (2002), in which the highest 100-year return levels were found in the mid-western part of the Korean peninsula. The second highest return levels appear to be shared by Gangneung and Busan (no significant difference between these two locations). As noted in section 2, the inland rural area of Chupungryong has the lowest return levels.

5. Conclusions

We have conducted a statistical analysis of extreme daily rainfall in South Korea using extreme value distributions. We believe that this is the first such study undertaken. We have established that the Gumbel distribution can be used as a reasonable model for extreme daily rainfall. We have provided estimates of the return level of daily rainfall and the corresponding 95% confidence intervals for five locations in South Korea, which were chosen to give a good geographical representation of the country. These estimates could be used as measures of flood protection.

This paper only provides an initial study of extreme daily rainfall in South Korea. This study can be extended in many ways. One avenue is to use distributions that are more flexible than the GEV, such as the four-parameter stable distribution. This distribution is specified by its characteristic function and suffers from the fact that it does not have closed form expressions for its distributional properties, even for its probability density function. Furthermore, the estimation procedures associated with this distribution are highly nonlinear. Whereas, most of the tables and the figures in this paper make use of closed form properties of the GEV distribution. Thus, if one has to use the four-parameter stable distribution then computations will have to use multiple numerical integration. This might lead to numerical difficulties.

However, there are other distributions that are at least as flexible as the GEV and allow closed form distributional properties. These include: the generalized Planck distribution, exponentiated Weibull distribution, exponentiated Gumbel distribution, Burr type distributions, generalized gamma distribution, generalized F distribution, generalized Pareto distribution and Markov chain based models. One could fit these distributions to the Korean rainfall data and see whether improvements can be achieved. We leave this exercise to a future study.

Acknowledgements

References

Aronica G, Cannarozzo M and Noto L 2002 Investigating the changes in extreme rainfall series recorded in an urbanised area; *Water Science and Technology* **45** 49–54.

Cannarozzo M, Dasaro F and Ferro V 1995 Regional rainfall and flood frequency-analysis for Sicily using the 2-component extreme-value distribution; *Hydrol. Sci. J.* **40** 19–42.

Crisci A, Gozzini B, Meneguzzo F, Pagliara S and Maracchi G 2002 Extreme rainfall in a changing climate: regional analysis and hydrological implications in Tuscany; *Hydrol. Processes* **16** 1261–1274.

Elnaqa A and Abuzeid N 1993 A program of frequency-analysis using Gumbel's method; *Ground Water* **31** 1021–1024.

Farago T and Katz R 1990 Extremes and design values in climatology; *World Meteorological Organization* WCAP-14, WMO/TD-No. 386.

Ferro V 1993 Rainfall intensity–duration–frequency formula for India – discussion; *Journal of Hydraulic Engineering–ASCE* **119** 960–962.

Fisher R A and Tippett L H C 1928 Limiting forms of the frequency distribution of the largest or smallest member of a sample; *Proceedings of the Cambridge Philosophical Society* **24** 180–290.

Galambos J 1987 *The Asymptotic Theory of Extreme Order Statistics*, 2nd edn. (Melbourne, Florida: Krieger).

Hosking J R M, Wallis J R and Wood E F 1985 Estimation of the generalized extreme value distribution by the method of probability weighted moments; *Technometrics* **27** 251–261.

Jenkinson A F 1955 The frequency distribution of the annual maxima (or minima) values of meteorological elements; *Quarterly Journal of the Royal Meteorological Society* **81** 158–171.

Kotz S and Nadarajah S 2000 *Extreme Value Distributions: Theory and Applications.* (London: Imperial College Press).

Koutsoyiannis D and Baloutsos G 2000 Analysis of a long record of annual maximum rainfall in Athens, Greece, and design rainfall inferences; *Natural Hazards* **22** 29–48.

McLean D M 1985 Deccan Traps mantle degassing in the terminal Cretaceous marine extinctions; *Cret. Res.* **6** 235–259.

Miroslava U 1991 The extreme value distribution of 5-min rainfall data Belgrade; *Theoretical and Applied Climatology* **44** 223–228.

Miroslava U 1992 The extreme value distribution of rainfall data at Belgrade, Yugoslavia; *Atmosfera* **5** 47–56.

Nguyen V T V, Nguyen T D and Wang H 1998 Regional estimation of short duration rainfall extremes; *Water Science and Technology* **37** 15–19.

Nguyen V T V, Nguyen T D and Ashkar F 2002 Regional frequency analysis of extreme rainfalls; *Water Science and Technology* **45** 75–81.

Parida B P 1999 Modelling of Indian summer monsoon rainfall using a four-parameter Kappa distribution; *Int. J. Climatol.* **19** 1389–1398.

Park J S and Jung H S 2002 Modelling Korean extreme rainfall using a Kappa distribution and maximum likelihood estimate; *Theoretical and Applied Climatology* **72** 55–64.

Park J S, Jung H S, Kim R S and Oh J H 2001 Modelling summer extreme rainfall over the Korean peninsula using Wakeby distribution; *Int. J. Climatol.* **21** 1371–1384.

Prescott P and Walden A T 1980 Maximum likelihood estimation of the parameters of the generalized extreme-value distribution; *Biometrika* **67** 723–724.

Rao C R 1973 *Linear Statistical Inference and Its Applications*, 2nd edn. (New York: John Wiley and Sons).

Silverman B W 1986 *Density Estimation for Statistics and Data Analysis.* (London: Chapman & Hall).

Wald A 1943 Tests of statistical hypotheses concerning several parameters when the number of observations is large; *Transactions of the American Mathematical Society* **54** 426–483.

Withers C S and Nadarajah S 2000 Evidence of trend in return levels for daily rainfall in New Zealand; *J. Hydrol. (NZ)* **39** 155–166.

Zalina M D, Desa M N M, Nguyen V T V and Kassim A H M 2002 Selecting a probability distribution for extreme rainfall series in Malaysia; *Water Science and Technology* **45** 63–68.