Statistical Modelling of Extreme Rainfall in Taiwan

Michael McAleer[1,2,3,4]
[1]Econometric Institute, Erasmus, School of Economics, Erasmus University Rotterdam;
[2]Tinbergen Institute, The Netherlands;
[3]Institute of Economic Research, Kyoto University, Japan;
[4]Department of Quantitative Economics, Complutense University of Madrid, Spain;
michael.mcaleer@gmail.com;

Lan-Fen Chu
National Science and Technology Center for Disaster Reduction (NCDR), Taiwan;
lanfen.chu@gmail.com,
Tel: 886-2-8195-8689;
Corresponding Author.

Ching-Chung Chang
Institute of Economics, Academia Sinica, Taiwan.
emily@econ.sinica.edu.tw

Abstract—In this study, the annual maximum daily rainfall from 1911 to 2010 are modeled for 6 stations in Taipei, Taichung, Tainan, Hengchun, Hualien, and Taitung. We fit the rainfall data both with stationary and non-stationary generalized extreme value (GEV) distribution, and further estimated future patterns based on the best fitted model. The non-stationary model used the location and scale parameter in the GEV distribution as a function of time to detect the temporal trend of the maximum rainfall. The results suggest that non-stationary GEV and Gumbel distribution provide best fit for Haulien and Taitung, respectively. The return levels of the stationary models for 4 stations and their 95% confidence intervals (CI) are provided. In addition, the return period of Typhoon Morakot is also calculated based on stationary models in this study.

Keywords-Extreme Value Theory; Extreme Rainfall; Return Level; Typhoon Morakot

I. INTRODUCTION

All extreme rainfall events are a primary cause of flooding hazards worldwide, and thus, considerable attention has been paid to the modeling of extreme rainfall to help prevent flooding hazards, and for water-related structures, agricultural, monitoring climate changes, and other issues. Taiwan is a small island located off the east coast of the Euro-Asian continent. It is often struck by typhoons in summer and fall that cause significant damage to Taiwan, especially in highly concentrated population and property areas. An increasing trend in extreme rainfall has been recently observed following typhoons, which implies that Taiwan has a higher probability of suffering damage from this type of natural disaster. Thus, understanding these extreme patterns and their future behavior is increasingly important in Taiwan.

Several researchers have studied useful applications of generalized extreme theory (GEV) on the rainfall data from different parts of the world: for example [6], [7], [8], [9], [21], [10], [11], [12], [13], and [2]. However, no research has examined extreme rainfall via generalized extreme theory in Taiwan, or used 100-year time-series data to establish the pattern of extreme rainfall, and to estimate future patterns for a certain time span.

Rainfall data are difficult to analyze [21]. In Taiwan, certain researchers have analyzed extreme climate by using general circulation model (GCM) scenarios provided in The Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Downscaling both dynamic and statistical efforts provides information on a finer scale relative to the data produced by GCMs, such as GCM project climate parameters at a resolution of 250 km^2; downscaled models provide projections at 50 km^2. However, this method is usually costly and requires considerable computer resources; it is difficult for a researcher with a financial deficit to simulate climate variation. Although downscaling can show relationships between small- and large-scale variables to overcome the drawbacks of the GCM method, it still cannot reflect the diverse topography in Taiwan, which has a full range of climate zones. Thus, using statistical method to analyze the distribution of extreme rainfall can enrich the related researches on climate change in Taiwan, this is also the distribution of this study.

The remainder of this paper is organized as follows: Section II presents the data used in this study; Section III introduces the methodology; Section IV shows the results; Section V provides the return period of Typhoon Morakot; and lastly, Section VI offers a conclusion.

II. DATA

The original data consisted of daily rainfall records from 1911 to 2010, which were provided by the Central Weather Bureau (CWB) of Taiwan. Table I shows the summary statistics of rainfall data of six stations, including station number, latitude, and the longitude for each station. In this study, extreme rainfall is defined as the maximum daily rainfall for each year using the block maxima method for defining extreme values [1].

TABLE I. DESCRIPTIVE STATISTICS OF RAINFALL DATA (UNIT: MM)

Lo-cation	Station No.	Longi-tude	Lati-tude	Min	Max	Med	Mean
Tai-pei	46692	121°51'	25°04'	21	425	146	164
Taic-hung	46749	120°68'	24°15'	26	660	166	187
Tai-nan	46741	120°18'	22°20'	13	523	196	206
Heng-chun	46759	120°74'	22°01'	17	484	215	232
Hua-lien	46699	121°61'	23°98'	13	465	210	217
Tai-dung	46766	121°15'	22°76'	8	484	201	245

Note: "Min", "Max", "Med" and "Mean" mean the minimum, maximum median and the expected value of the extreme rainfall, respectively.

III. METHODOLOGY

Generalized extreme value (GEV) is based on Gumbel, Fréchet and the Weibull distribution. It developed by Jenkinson [3] who combined the above three distribution (see [4], [5]). The cumulative distribution function (cdf) of the GEV distribution is:

$$\text{GEV}_0: F(x) = \exp\left\{-(1+\xi(x-\mu)/\sigma)^{-1/\xi}\right\},$$

$$1+\xi(x-\mu)/\sigma > 0 \qquad (1)$$

where μ, $\sigma > 0$ and ξ are the location parameter, scale and shape parameter, respectively. The case of (1) for $\xi = 0$

$$\text{Gum}_0: F(x) = \exp\left\{-\exp(-(x-\mu)/\sigma)\right\},$$

$$-\infty < x < \infty \qquad (2)$$

is a Gumbel distribution, whereas the case $\xi > 0$ and $\xi < 0$ of (1) are known as the Fréchet family and the negative Weibull distributions, respectively. The method of maximum likelihood was used to fit the related models throughout the paper, and the maximization was performed using a quasi-Newton iterative algorithm. The basic model fitted was GEV (to be referred to as GEV_0; (1)) with constant μ, σ and ξ. Reference [19] indicated that sometimes 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 Gum_0). We follow the same statistical methods as [15], [16], and [17]. Consequently, some descriptions throughout this paper is the same as theirs.

Assuming the independence of the data, the likelihood function is the product of the assumed densities for the observations $x_1, x_2, ..., n$. For GEV_0 model, we have

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

$$\times \exp\left\{-\sum_{i=1}^{n}\left(1+\xi\frac{x_i-\mu}{\sigma}\right)^{-1/\xi}\right\} \qquad (3)$$

provided that

$$1+\xi\frac{x_i-\mu}{\sigma} > 0 \text{ for } i=1,...,n \qquad (4)$$

The estimates of μ, σ and ξ, denoted as $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$, are taken to be those values that maximize the likelihood function L. A standard way of determining the best fit model is the likelihood ration test. If L_i is the maximum likelihood for model i, and L_j is the maximum likelihood for model j, then under the simpler model, the test statistic λ =-2 log(L_j / L_i) is assumed to be distributed as a chi-square variable with v degrees for freedom, where v is the difference in the number of parameters between the models i and j. In hypothesis testing problems, this would be asymptotically true as the amount of data tends to be infinite. Thus, at the 5% significance level, model j is preferred if λ =-2 log(L_j / L_i)> $x^2_{v,0.95}$. In practice, because annual maxima lack complete independence, this is likely to be conservatively interpreted [15].

To investigate the existence of a trend in extreme rainfall with respect to time, we applied the following variations of Model GEV_0 and Gum_0:

$$\text{GEV}_1: \mu = a + b \text{ (Year-}t_0+1),$$

$$\sigma =\text{constant}, \quad \xi =\text{constant} \qquad (5)$$

a four-parameter model with μ allowed to vary linearly with respect to time, and "constant" means that the parameter is not time dependent and is subject to the estimation;

$$\text{Gum}_1: \mu = a + b \text{ (Year-}t_0+1),$$

$$\sigma =\text{constant}, \quad \xi =0 \qquad (6)$$

a three-parameter model, where μ vary with respect to time;

$$\text{GEV}_2: \mu =\text{constant}, \xi =\text{constant}$$

$$\sigma = \exp(a + b \times (\text{Year-}t_0\text{-1})), \qquad (7)$$

a four-parameter model with σ vary exponentially with time, and

Gum$_2$: μ =constant, ξ =0

$$\sigma = \exp(a + b \times (\text{Year-t}_0\text{-1})) , \qquad (8)$$

a three-parameter model with σ vary exponentially with time. In all models, t_0 denotes the year the record started. None of the GEV$_2$ and Gum$_2$ Model provided a significant fit. A similar technique has been used by many published researchers, including [20], [19], [16] and [13].

Once the best models for the data are determined, the next step is to derive the return levels for rainfall. The *T*-year return level, x_T , is the level exceeded on average only once in every *T* years [18]. If Model GEV$_0$ is assumed then on inverting $F(x_T) = 1 - 1/T$, we obtain the expression:

$$x_T = \mu - \frac{\sigma}{\xi}\left\{1 - \left[-\log(1-1/T)\right]^{-\xi}\right\} \qquad (9)$$

If Model Gum$_0$ is assumed then the corresponding expression is:

$$x_T = \mu - \sigma \log\left\{-\log(1-1/T)\right\} \qquad (10)$$

On substituting $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$ into (9)-(10), we have the maximum likelihood estimates of the return levels.

IV. RESULT

All estimated results, including the estimates and standard errors (SE), are shown in Table II. Significant evidence of a non-stationary increasing trend was observed for the Hualien and Taitung stations. Hualien station was suitable for the generalized extreme-value (GEV) distribution, and Taitung station was suitable for the Gumbel distribution. The estimator of $\hat{b}$ is the change in extreme rainfall from one year to the next. The stationary Gumbel distribution showed a good fit for Taipei, Taichung, and Tainan, and the stationary GEV distribution provided a good fit for Hengchun.

TABLE II.   BEST FITTED MODELS AND PARAMETER ESTIMATES

Location	Model	$\hat{a}$ (SE)	$\hat{b}$ (SE)	$\hat{\sigma}$ (SE)	$\hat{\xi}$ (SE)
Taipei	Gum$_0$	131.41 (5.55)		52.84 (4.12)	
Taichung	Gum$_0$	147.26 (7.09)		67.56 (5.32)	
Tainan	Gum$_0$	165.40 (7.77)		73.60 (5.48)	
Hengchun	GEV$_0$	196.33 (8.78)		79.06 (6.14)	-0.14 (0.06)
Hualien	GEV$_1$	157.29 (15.71)	52.21 (26.99)	76.16 (5.77)	-0.15 (0.05)

Taitung	Gum$_1$	144.61 (14.66)	67.01 (25.40)	69.15 (5.07)	

Table III provides the estimates of the return level based on the best fitted and the staionary models for T=10, 20, 50 and 100 years. The estimates and the associated 95% confidence intervals are also provided. Among the four considered locations, the Tainan and Hengchun in the southern of Taiwan to be estimated with the highest return levels.

TABLE III.   RETURN LEVEL ESTIMATES FOR 10, 20, 50, AND 100 YEARS

Location	10-year	20-year	50-year	100-year
Taipei	250 (189,310)	288 (228,349)	338 (277,397)	374 (314,345)
Taichung	299 (201,397)	347 (249,447)	410 (312,509)	458 (359,556)
Tainan	331 (213,449)	384 (265,502)	452 (334,571)	503 (385, 622)
Hengchun	348 (323,381)	387 (357, 434)	431 (394, 503)	461 (416, 555)

Note: Confidence intervals for return level estimates are based on delta method [22].

V. THE RETURN PERIOD OF TYPHOON MORAKOT

Our estimated results from the return levels can help explain the rarity of the extreme rainfall from a typhoon. Taking Typhoon Morakot as an example, it struck Taiwan from August 6 to August 10, 2009 and brought the heaviest rainfall that was ever recorded in the historic data. A record-high 1,162 mm of accumulated rainfall on August 7 was measured in Alishan. However, the rainfall of the first day of the typhoon at the same station was 420 mm, which is still significantly high for Taiwan. Therefore, we used the accumulated rainfall of 420 and 1162 mm from Typhoon Morakot as an extreme event to estimate the return period for Taichung, Tainan, and Hengchun. In addition, in this case, the maximum accumulated rainfall for Taichung, Tainan, and Hengchun peaked at 347, 523, and 406 mm, respectively; therefore, these data were used as the level of accumulation to analyze return periods for Taichung, Tainan, and Hengchun stations, to predict the future behavior of extreme rainfall.

The estimated results show that, with an accumulated rainfall of 420 mm a day, the return period for Taichung, Tainan, and Hengchun is approximately 37, 59, and 32 years, respectively; under an accumulated rainfall of 406 mm a day, the return period for Taichung, Tainan, and Hengchun is approximately 28, 48, and 27 years, respectively. If this tpye of typhoon hits Taiwan in the future, with an accumulated rainfall of up to 347 mm a day, Taichung and Hengchun will experience extreme rainfall with a return period of 10 years and 12 year, respectively.

TABLE IV.   THE RETURN PERIOD OF TYPHOON MORAKOT

Alishan	Taichung	Tainan[*]	Hengchun[*]

	420 mm	1,162 mm	347 mm	523mm	406 mm
Taichung	37	--	10	500	28
Tainan	59	--	20	250	48
Hengchun	32	--	12	125	27

Note: "*" means the probability is closed to 0, and the return period is close to infinite.

VI. CONCLUSION

We analyzed extreme rainfall in Taiwan based on a statistical analysis of extreme daily rainfall, and we calculated the return level from the best fitted models and provided the corresponding 95% CI for these chosen models. Significant evidence was observed of a non-stationary increasing trend for the Hualien and Taitung stations. Hualian station is suitable for GEV application, and Taitung station is suitable for Gumbel distribution application. The stationary Gumbel distribution profiled a good fit for Taipei, Taichung, and Tainan, and the stationary GEV distribution provided a good fit for Hengchun.

In this study, we also used Typhoon Morakot as an extreme rainfall event to estimate the return period of this type of event in the future. The maximum 420 and 1,162 mm accumulated rainfall per day that Typhoon Morakot generated for Alishan station were used as the baseline for estimating the return periods for Taichung, Tainan, and Hengchun. In addition, the maximum accumulated rainfall that Typhoon Morakot generated in Taichung (347 mm), Tainan (523 mm), and Hengchun (406 mm) were also used to calculate the return periods for the three stations. The calculated results show that, with the maximum accumulation at Alishan station, the return periods for Taichung, Tainan, and Hengchun are 37, 59, and 32 years, respectively. With the maximum accumulated rainfall that Typhoon Morakot generated for Taichung station, the return periods for Taichung and Hengchun are 10 and 12 years, respectively, and the return period for Tainan is 20 years. For the maximum accumulated rainfall that Typhoon Morakot generated for Hegnchun, the return periods for Taichung, Tainan, and Hengchun are 28 years, 48 years, and 27 years, respectively. The results show an important implication: rainfall at 347 and 406 mm can be regarded as a threshold for providing a warning system, especially in Taichung and Hengchun. The caculated return period from the case of Typhoon Morakot can be considered as short and medium-term schedule for flooding planning from extreme rainfall in the future in Taiwan.

ACKNOWLEDGMENT

REFERENCES

[1] E. J. Gumbel, Statistics of Extremes. Columbia University Press, New York, 1985.

[2] K. Adamowski and J. Bougadis, "Detection of trends in annual extreme rainfall," Hydrol. Process, vol. 17, pp. 3547-3560.

[3] A. F. Jenkinson, "The frequency distribution of the annual maximum (or minimum) values of meteorological elements," Quart. J. Roy. Meteor. Soc., vol. 81, 1955, pp. 158−171.

[4] J. R. M. Hosking and J. R. Wallis, "Parameter and quantile estimation for the generalized Pareto distribution," Technometrics, vol. 29(3), 1987, pp. 339−349.

[5] J. Galambos, The Asymptotic Theory of Eextreme Order Statistics, 2nd edn. (Melbourne, Florida: Krieger), 1987.

[6] V. T. V. Nguyen, T. D. Nguyen, and H. Wang, "Regional estimation of short duration rainfall extremes," Wat. Sci. Technol., vol. 37, 1998, pp. 15-19.

[7] V. T. V. Nguyen, T. D. Nguyen, and F. Ashkar, "Regional frequency analysis of extreme rainfalls," Wat. Sci. Technol., vol. 45, 2002, pp. 75-81.

[8] D. Koutsoyiannis and G. Baloutsos, "Analysis of a long record of annual maximum rainfall in Athens, Greece, and design rainfall inferences," Nat. Hazards, vol. 22, 2002, pp. 29–48.

[9] M. Cannarozzo, F. Dasaro, and V. Ferro, "Regional rainfall and flood frequency-analysis for sicily using the 2-component extreme-value distribution," Hydrol. Sci. J., vol. 40, 1995, pp. 19–42.

[10] J. S. Park, H. S. Jung, R. S. Kim, and J. H. Oh, "Modelling summer extreme rainfall over the Korean peninsula using Wakeby distribution," Int. J. Climatol., vol. 21, 2001, pp. 1371-1384.

[11] G. Aronica, M. Cannarozzo, and L. Noto, "Investigating the changes in extreme rainfall series recorded in an urbanised area," Wat. Sci. Technol., vol. 45, 2002, pp. 49–54.

[12] M. D. Zalina, M. N. M. Desa, V. T. V. Nguyen, and A. H. M. Kassim, "Selecting a probability distribution for extreme rainfall series in Malaysia," Wat. Sci. Technol., vol. 45, 2002, pp. 63–68.

[13] J. S. Park and H. S. Jung, 2002, "Modelling Korean extreme rainfall using a Kappa distribution and maximum likelihood estimate," Theor. Appl. Climatol., vol. 72, 2002, pp. 55–64.

[14] C. S. Whithers and S. Nadarajah, "Evidence of trend in return levels for daily rainfall in New Zealand," J. Hydrol., vol. 39(2), 2002, pp. 155-166.

[15] S. Nadarajah, "Extremes of daily rainfall in West Central Florida," Clim. Change, vol. 69, 2005, pp. 325-342.

[16] S. Feng, S. Nadarajah, and Q. Hu, "Modeling annual extreme precipitation in China using the generalized extreme value distribution," J. Meteorol. Soc. Jpn., vrol. 85, 2007, pp. 599–613.

[17] J-S Park, H-S Kang, Y-S, Lee, and M-K Kim, "Changes in extreme daily rainfall in South Korea," Int. J. Climatol., 2010.

[18] S. Coles, An Introduction to Statistical Modeling of Extreme Values, Springer Series in Statistics, Springer Verlag London, 2001.

[19] S. Nadarajah and D. Choi, 2007, "Maximum daily rainfall in South Korea," J. Earth Syst. Sci., vol. 116, 2007, pp. 311–320.

[20] R. W. Katz, M. B. Parlang, P. Naveau, "Statistics of extremes in hydrology," Adv. Wat. Resour., vol. 25, 2002, pp. 1287–1304.

[21] C. S. Withers and S. Nadarajah, "Evidence of trend in return levels for daily rainfall in New Zealand," J. Hydrol. (NZ), vol. 39, 2000, pp. 155-166.

[22] C. R. Rao, Linear Statistical Inference and Its Applications, 2nd eds. New York: John Wilsey and Sons, 1973.