The Journal of Applied Science **Vol. 20 No. 1: 28-38 [2021]**
วารสารวิทยาศาสตร์ประยุกต์ doi: 10.14416/j.appsci.2021.01.003
ISSN 1513-7805 Printed in Thailand ISSN 2586-9663 (Online)

Research Article

Extreme Value Modeling of Daily Maximum Temperature with the r-Largest Order Statistics

Piyapatr Busababodhin[1], Monchaya Chiangpradit[1], Nipada Papukdee[1], Jiraphon Ruechairam[1], Kettida Ruanthaisong[1] and Pannarat Guayjarernpanishk[2]*

[1]Department of Mathematics, Mahasarakham University, Mahasarakham, 44150, Thailand
[2]Faculty of Interdisciplinary Studies, Nong Khai Campus, Khon Kaen University, Nong Khai, 43000, Thailand
***E-mail**: Panngu@kku.ac.th
Received: 03/08/2020; Revised: 27/11/2020; Accepted: 07/12/2020

Abstract

This study aimed to model the daily maximum temperature in Mae Hong Son, Nan, and Uttaradit which are the first-three highest temperature provinces in the north of Thailand. These data sets were collected during 1984–2018 by the Thai Meteorological Department. Many researchers usually analyzed extreme value data by using the generalized extreme value distribution and focused on the first order of maximum value. These interested the researcher to create a model with the generalized extreme value distribution for the r-largest order statistics with different orders such as r = 2, 3, 4, …, n where n is the number of data and compare the model of the maximum value of any sequence with the deviance statistics. The researcher analyzed the data with the "EVA" package in R program. The results showed that Weibull distribution is the best model for the daily temperature data of Nan Meteorological and Nan Sorkorsor Meteorological station when r =4, Mae Hong Son Meteorological and Uttaradit Meteorological station when r = 3, and Mae Sariang Meteorological when r =2. Besides, Gumbel distribution is the best model for the daily maximum temperature data of Tha Wang Pha Meteorological station when r =4, and Thung Chang Meteorological station when r =2. Furthermore, based on the return levels of various return periods, the Mae Hong Son Meteorological station had the highest return level of maximum temperature for each return period. Therefore, to solve and prevent temperature problems, the Mae Hong Son Meteorological station should be the first considered

Keywords: generalized extreme value distribution, the r-largest order statistics, return level, deviance statistics

Introduction

Thailand is in a tropical zone with a general climate of torridity all year long. The average temperature throughout the year of Thailand is approximately 27-celsius degree. However, the temperature is different in each area and season. Areas in the mainland from the central region and upper eastern region up to the northern region have extremely different temperatures between in summer and in winter and between at day and at night. In summer, the highest temperature during the afternoon may reach 40-celsius degree or higher from March to May, especially April which is the month with the hottest weather in each year.

doi: 10.14416/j.appsci.2021.01.003

The climate of the northern region is tropical with seasonal rain which is influenced by monsoons from two directions including northeast monsoon during the middle of October to Mid-February or March in some years which brings coldness from China to Thailand to be winter with dry weather and southwest monsoon during Mid-May to the end of September which blows moisture from the equator in the rainy season. The average temperature in the northern region is 25-celsius degree, the highest temperature in April with approximately 42-celsius degree, and the lowest temperature in January with approximately 15-celsius degree. Chiang Rai has the lowest temperature. Uttaradit has the highest temperature. Meanwhile, the province with has the heaviest rain is Chiang Rai and Lampang has the lightest rain. Furthermore, Nan, Mae Hong Son, and Uttaradit were found to be the top three provinces with the highest temperature in Thailand (The Thai Meteorological Department, 2016)

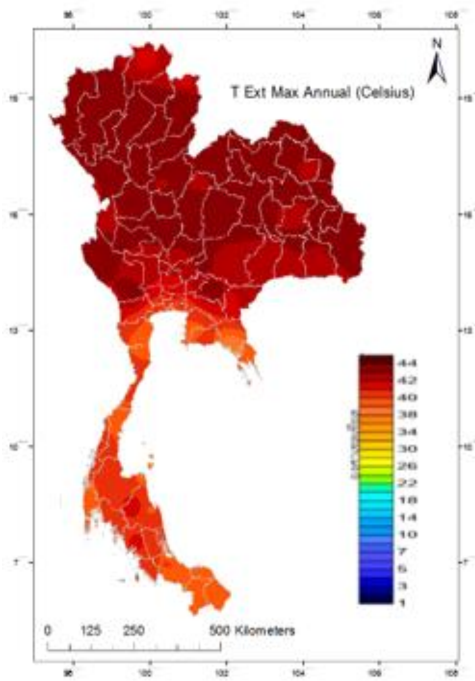

Figure 1 Areas with the highest temperature in Thailand
Source: Division of Meteorological Development (Hydrological Meteorological Division)

There were several research studies related to the application of extreme value theory with meteorological and hydrological data. Ying An, and M.D. Pandey (2007) studied the statistical estimation of the extreme wind speed using annually r-largest order statistics (r-LOS) extracted from the time series of wind data. The hourly wind speed data collected at 30 stations in Ontario, Canada, are analyzed in the paper. The result shows that the wind pressure data can be suitably model by the Gumbel distribution. Benjawan Charin (2014) studied a suitable model of the highest and the lowest temperature data of the middle northeastern region in Thailand by generalized extreme value distribution and generalized Pareto

 doi: 10.14416/j.appsci.2021.01.003

distribution. Additionally, she investigated the return level of the highest and the lowest temperature in 2, 5, 10, 25, and 50 years. Data of the highest and the lowest temperature during 1977-2013 from the Thai Meteorological Department in 6 stations including Khon Kaen Meteorological station, Tha Phra Agrometeorological station, Kosum Phisai Hydrological station, Kamalasai Hydrological station, Roi-Et Meteorological station, and Roi-Et Agrometeorological station were analyzed by R program. The results revealed that the models of the highest and the lowest temperature in every station were Gamma distribution. The return level of the highest temperature, Khon Kaen Meteorological station was the highest rate of return, and the return level of the lowest temperature, Roi-Et Agrometeorological station was the highest rate of return. Prapawan Senapeng and Piyapatr Busababodhin (2017) created appropriate models of the highest temperature data in the northeastern region of Thailand with generalized extreme value distribution (GEV) and generalized Pareto distribution (GPD) and investigated the return level of the highest temperature. They employed the data from 1985-2015 in 25 stations from the Thai Meteorological Department. For GEV, the findings demonstrated that Weibull distribution and Gumbel distribution were suitable distribution for the highest temperature in 15 stations and 10 stations, respectively. For GPD, the appropriate distributions were Gamma distribution in 24 stations and exponential distribution in 1 station. When considering the return level in 5, 10, 25, 50, and 100 years of temperature in the northeastern region of Thailand, the researchers found that the return rate of every station rose when years of return increased. Also, Nong Khai Weather station was the highest return rate. Thus, relevant organizations should prevent or solve the temperature problems at Nong Khai Weather station more than others.

According to the research study, there is no literature that examines extreme value modeling for daily maximum temperature by the r-largest order statistics. Therefore, we are interested to study extreme value modeling with the r-largest order statistics (GEV-r) applying with of daily maximum temperature of northeast Thailand. We apply data of Mae Hong Son, Nan, and Uttaradit which were the top three provinces with the highest temperature in the north of Thailand. We investigate the most suitable model with the deviance statistics. Moreover, the return levels of the highest temperature in the return period of 2, 5, 10, 20, and 50 years were demonstrated. These would be the highest benefit for dealing with temperature problems in advance.

Preliminaries

Generalized extreme value distribution (GEV)

Let $X_1, X_2, ..., X_n$ be a sequence of independent random variables having a common distribution function F and let $M_n = \max(X_1, X_2, ..., X_n)$. The theory of extreme type identifies for sequences of constant $a_n > 0$ and b_n as follow)Coles, 2001(

$$\Pr\left\{\left(\frac{M_n - b_n}{a_n}\right) \le x\right\} = F^n(a_n x + b_n) \to G(x)\ ;\ n \to \infty, \tag{1}$$

for a non-degenerate distribution function G, the G is a member of the GEV family

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \tag{2}$$

Vol. 20 No. 1: 28-38 [2021]
doi: 10.14416/j.appsci.2021.01.003

defined on $\left\{x : 1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0\right\}$, where $-\infty < \mu < \infty$ is location parameter, $\sigma > 0$ is scale parameter and $-\infty < \xi < \infty$ is shape parameter.

If $\xi \to 0$ then GEV is called the Gumbel distribution.
If $\xi > 0$ then GEV is called the Fréchet distribution.
If $\xi < 0$ then GEV is called the Weibull distribution.

The probability density function of GEV as shown in Figure 2.

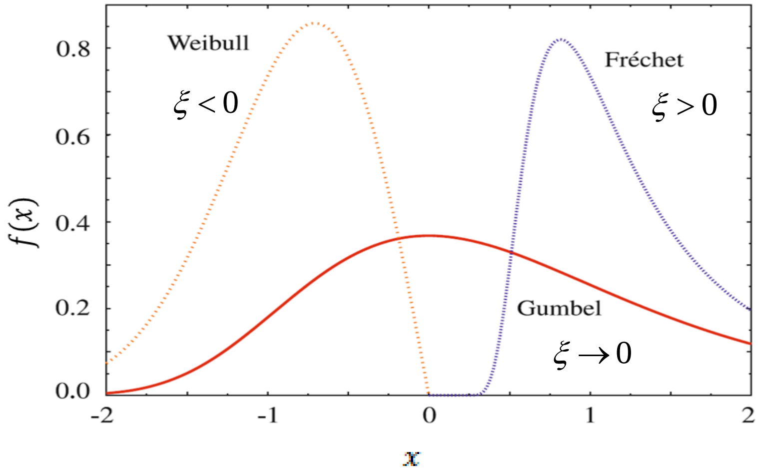

Figure 2 Probability density function of generalized extreme value distribution categorized by shape parameter

Return Level

We estimated situation rainfall from GEV with return level. The return level shows that event will appear at least one time in the next t years. Estimates of extreme quantiles of the GEV are then obtained by inverting (2):

$$x_p(G) = \begin{cases} \mu - \frac{\sigma}{\xi}\left\{1-\left[-\log(1-p)\right]^{-\xi}\right\}, & \xi \neq 0 \\ \mu - \sigma\log\left[-\log(1-p)\right], & \xi = 0 \end{cases} \tag{3}$$

where $G(x_p) = 1-p$, $0 < p < 1$. In common terminology, $x_p(G)$ is the return level associated with the return period $t = \frac{1}{p}$ (Gumbel, 1958; Coles, 2001).

Generalized extreme value distribution for r-largest order statistics (GEV-r)

Let $X_1, X_2, ..., X_n$ be a sequence of independent and identically distribution random variables, and $M_n^{(r)} = \text{r th largest of } \{X_1, X_2, ..., X_n\}$ is extreme order statistics when fixed r, as $n \to \infty$.

From (1) (Coles, 2001) the limiting distribution as $n \to \infty$ for fixed r of

 doi: 10.14416/j.appsci.2021.01.003

$$\tilde{M}_n^{(r)} = \left(\frac{M_n^{(1)} - b_n}{a_n}, \frac{M_n^{(2)} - b_n}{a_n}, ..., \frac{M_n^{(r)} - b_n}{a_n} \right).$$

Falls within the family having joint probability density function

$$f(x^{(1)},...,x^{(r)}) = \exp\left\{ -\left[1 + \xi\left(\frac{x^{(r)} - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \times \prod_{k=1}^{r} \frac{1}{\sigma}\left[1 + \xi\left(\frac{x^{(r)} - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}-1} \quad (4)$$

where $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \xi < \infty$, $x^{(r)} \leq x^{(r-1)} \leq ... \leq x^{(1)}$ and $x^{(k)} : 1 + \xi\left(\frac{x^{(k)} - \mu}{\sigma} \right) > 0$, for $k = 1, 2, ..., r$.

Maximum Likelihood Estimation (MLE)

Under the assumption that $X_1, X_2, ..., X_n$ are independent random having the GEV distribution, the log-likelihood for the GEV parameters $\left(l(\mu, \sigma, \xi) \right)$ are follows:

In case of $\xi \neq 0$

$$l(\mu, \sigma, \xi) = -m\log\sigma - \left(1 + \frac{1}{\xi} \right)\sum_{i=1}^{m}\log\left[1 + \xi\left(\frac{x_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^{m}\log\left[1 + \xi\left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}}, \quad (5)$$

provided that $1 + \xi\left(\frac{x_i - \mu}{\sigma} \right) > 0$, for $i = 1, 2, ..., m$.

In case of $\xi = 0$

$$l(\mu, \sigma) = -m\log\sigma - \sum_{i=1}^{m}\left(\frac{x_i - \mu}{\sigma} \right) - \sum_{i=1}^{m}\exp\left[-\left(\frac{x_i - \mu}{\sigma} \right) \right]. \quad (6)$$

Maximization of the pair of (5) and (6) with respect to the parameters (μ, σ, ξ) leads to the maximum likelihood estimate with respect to the entire GEV family $\left(\hat{\mu}, \hat{\sigma}, \hat{\xi} \right)$.

The MLE of μ, σ, ξ $\left(\hat{\mu}, \hat{\sigma}, \hat{\xi} \right)$ giving by $\frac{\partial l(\mu, \sigma, \xi)}{\partial \mu} = 0$, $\frac{\partial l(\mu, \sigma, \xi)}{\partial \sigma} = 0$ and $\frac{\partial l(\mu, \sigma, \xi)}{\partial \xi} = 0$, respectively (Coles, 2001).

Model Suitability Validation

We analyzed the suitable model with minimum deviance statistics and the deviance statistics (D) is the goodness-of-fit test of the model by determining that r_j is improved model and r_i is the basic model (Murendeni M. Nemukula & Caston Sigauke, 2018), the hypothesis testing and test statistics are set as follows

H_0: improved model is suitable

H_1: improved model is not suitable

$$D(i,\ j) = 2\ln\left(\frac{\lambda(\mathrm{r}_i)}{\lambda(\mathrm{r}_j)}\right)$$
$$= 2\left\{\ln\lambda(\mathrm{r}_i) - \ln\lambda(\mathrm{r}_j)\right\} \approx \chi_k^2 \qquad (7)$$

where $i, j = 1,2,...6$, $i \neq j$ and k is the difference of the number of parameters of both models. $\lambda(\mathrm{r}_i)$ and $\lambda(\mathrm{r}_j)$ are the highest probability function of r_i and r_j respectively. H_0 is rejected if $D(i,\ j) > C_\alpha$ (critical value) provided that C_α is critical value at position $(1-\alpha)$ of distribution of χ_k^2 (Coles, 2001; Smith, 1989; Soares & Scotto, 2004).

Analysis result

We analyzed data with R program by using the package " Extreme Value Analysis (EVA)" of Bader and Yan in 2015. The model analysis could be summarized as follows.

Initial maximum temperature analysis

We clean daily data temperature in 7 stations of Mae Hong Son, Nan, and Uttaradit. There is Mae Hong Son Meteorological station (300201), Mae Sariang Meteorological station (300202), Nan Meteorological station (331201), Nan Agrometeorological station (331301), Tha Wang Pha Army Headquarter Meteorological station (331401), Thung Chang Army Headquarter Meteorological station (331402) and Uttaradit Meteorological station (351201). Data were analyzed by generalized extreme value distribution for the r-largest order statistics.

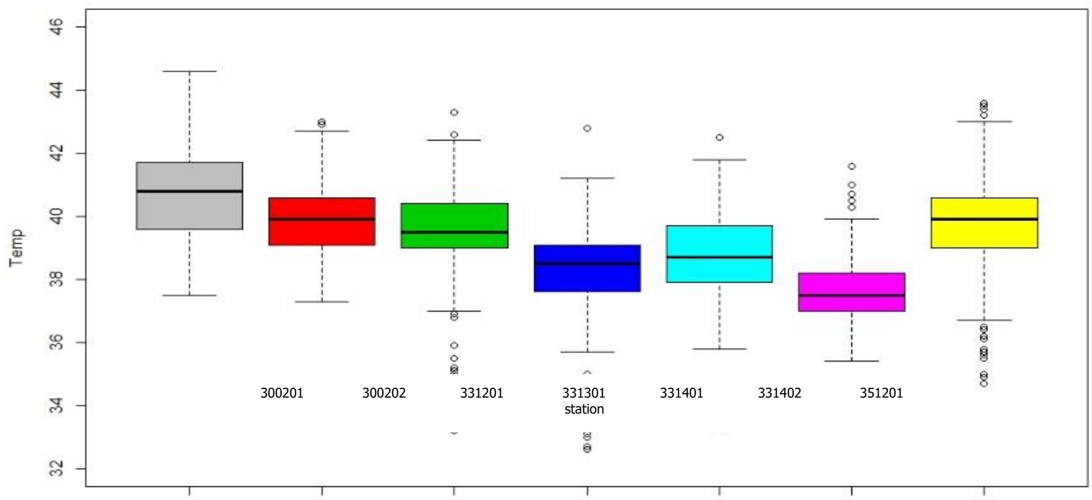

Figure 3 Boxplots of dialy maximum temperature

According to Figure 3, the daily maximum temperature of Mae Hong Son Meteorological station (300201) was the highest at 44.60-celsius degree, and Nan Agrometeorological station (331301) had the highest value at 37.50-celsius degree. We tested the appropriate model with deviance statistics, as shown in Table 1

doi: 10.14416/j.appsci.2021.01.003

Table 1 Deviance statistics test's results

Station	**Deviance statistics test**					**Suitable r- largest order statistics**
Mae Hong Son Meteorological (300201)	*D*(1,2)	*D*(2,3)	*D*(3,4)	*D*(3,5)	*D*(3,6)	3
	-33.65	-13.73	23.31	67.07	-7.03	
Mae Sariang Meteorological (300202)	*D*(1,2)	*D*(2,3)	*D*(2,4)	*D*(2,5)	*D*(2,6)	2
	-30.42	12.60	21.74	54.65	110.39	
Nan Meteorological (331201)	*D*(1,2)	*D*(2,3)	*D*(3,4)	*D*(4,5)	*D*(4,6)	4
	-41.49	-12.09	6.50	20.73	55.29	
Nan Agrometeorological (331301)	*D*(1,2)	*D*(2,3)	*D*(3,4)	*D*(4,5)	*D*(4,6)	4
	-45.52	-23.75	-0.16	24.56	60.87	
Tha Wang Pha Army Headquarter Meteorological (331401)	*D*(1,2)	*D*(2,3)	*D*(3,4)	*D*(4,5)	*D*(4,6)	4
	-53.45	-11.74	-11.31	28.75	57.74	
Thung Chang Army Headquarter Meteorological (331402)	*D*(1,2)	*D*(2,3)	*D*(3,4)	*D*(3,5)	*D*(3,6)	2
	-25.99	3.45	21.32	59.40	78.65	
Uttaradit Meteorological (351201)	*D*(1,2)	*D*(2,3)	*D*(3,4)	*D*(3,5)	*D*(3,6)	3
	-33.65	5.96	-0.31	28.62	71.88	

According to Table 1, for Mae Hong Son Meteorological station (300201), the deviance statistics test was performed to compare r_2 and r_1 e.g. $D(1, 2) = 2(-74.6214 - (-57.7967))$ =-33.6494. Since $D(1, 2)$, $D(2, 3)$, and $D(3, 6)$ had negative values, therefore they were not be considered because χ_1^2 can not be negative. Finally, considered $D(3, 4)$ and $D(3, 5)$ to select the best model with a significance level of 0.01 (χ_1^2 =6.64). In order to compare between r_4 and r_3, $D(3, 4) > 6.64$ implied that r_3 is a suitable model. The model of r_5 was compared with r_3, it was found that $D(3, 5) > 6.64$. Therefore, it reject the null hypothesis, so that r_3 was the suitable model. Thus, it could be concluded that r = 3 was the most suitable. Every station would consider the same way, and according to the deviance statistics test, the most suitable model for each station is shown in Table 2.

Best model of average daily temperature by using r-largest order statistic

The analysis could be used to summarize suitable model and distribution for annual maximum temperature categorized by stations as shown in Table 2

Vol. 20 No. 1: 28-38 [2021]
doi: 10.14416/j.appsci.2021.01.003

Table 2 Best model and suitable distribution of annual maximum temperature under GEV-r

Station	Best model	Parameter estimation			Suitable distribution	AIC
		$\hat{\mu}$ (se) (CI 95%)	$\hat{\sigma}$ (se) (CI 95%)	$\hat{\xi}$ (se) (CI 95%)		
Mae Hong Son Meteorological (300201)	r = 3	42.02 (0.16) (41.72, 42.33)	1.05 (0.07) (0.92, 1.18)	-0.35 (0.05) (-0.45, -0.24)	Weibull distribution	-156.9698
Mae Sariang Meteorological (300202)	r = 2	40.89 (0.151) (40.59, 41.18)	0.954 (0.08) (0.80, 1.10)	-0.34 (0.06) (-0.47, -0.22)	Weibull distribution	-125.8093
Nan Meteorological (331201)	r = 4	41.35 (0.11) (41.06, 41.63)	1.01 (0.05) (0.92, 1.10)	-0.50 (0.04) (-0.57, -0.42)	Weibull distribution	-169.5409
Nan Agrometeorological (331301)	r = 4	40.12 (0.16) (39.80, 40.45)	1.12 (0.06) (1.01, 1.23)	-0.39 (0.05) (-0.48, -0.30)	Weibull distribution	-173.5027
Tha Wang Pha Army Headquarter Meteorological (331401)	r = 4	40.50 (0.15) (40.20, 40.79)	1.06 (0.05) (0.96, 1.16)	-0.48 (0.04) (-0.56, 1.16)	Gumbel distribution	-189.0952
Thung Chang Army Headquarter Meteorological (331402)	r = 2	38.56 (0.20) (38.17, 38.94)	0.97 (0.11) (0.76, 1.19)	-0.18 (0.10) (-0.38, 0.01)	Gumbel distribution	-74.5829
Uttaradit Meteorological (351201)	r = 3	41.38 (0.15) (41.09, 41.67)	1.02 (0.05) (0.92, 1.12)	-0.39 (0.05) (-0.48, -0.30)	Weibull distribution	-157.7609

According to Table 2, we considered 95% confidence interval for ξ of Mae Sariang Meteorological station (300202) with order statistic r = 2, Mae Hong Son Meteorological station (300201) and Uttaradit Meteorological station (351201) with order statistic r = 3, Nan Meteorological station (331201) and Nan Agrometeorological station (331301) with order statistic r = 4. We found that Weibull distribution is a suitable model. Meanwhile, Tha Wang Pha Army Headquarter Meteorological station (331401) has order statistic r = 4, and Thung Chang Army Headquarter Meteorological station (331402) has order statistic r = 2 with Gumbel distribution.

Return level

Analysis of return level and effective return level in each return period was discovered. Annual maximum temperature data in 7 stations were suitable with parameters under a stable process. Therefore, the return level in each return period is shown in Figure 4

The Journal of Applied Science
วารสารวิทยาศาสตร์ประยุกต์

Vol. 20 No. 1: 28-38 [2021]
doi: 10.14416/j.appsci.2021.01.003

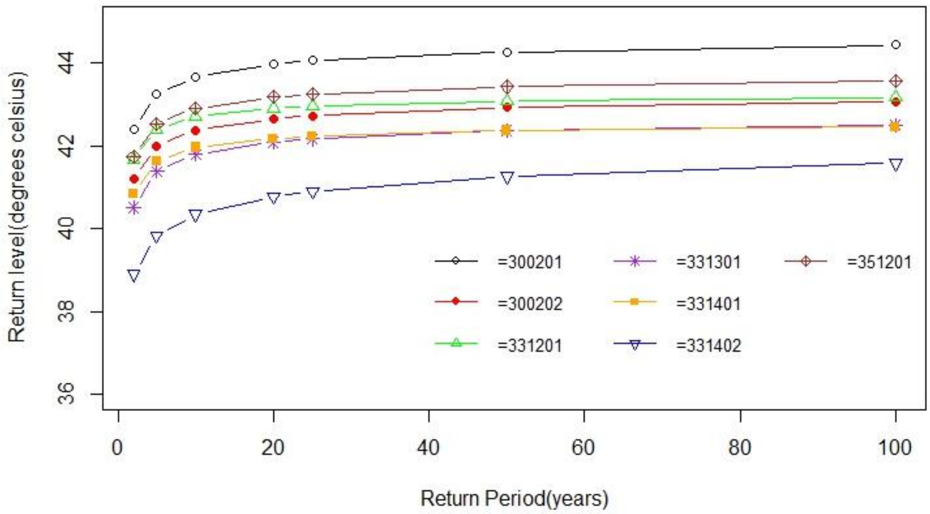

Figure 4 Return level in each return period by stations

According to Figure 4, the return level of annual maximum temperature at every station would increase when the return period rose. That is, the temperature in every area would be higher, especially Mae Hong Son Meteorological station (300201) followed by Uttaradit Meteorological station (351201), Nan Meteorological station (331201), Mae Sariang Meteorological station (300202), Nan Agrometeorological station (331301), Tha Wang Pha Army Headquarter Meteorological station (331401) and Thung Chang Army Headquarter Meteorological station (331402), respectively.

We applied the analysis result of the return period for 2, 10, 20, and 50 years in the geographic information system (GIS) program as shown in Figure 5 and Figure 6

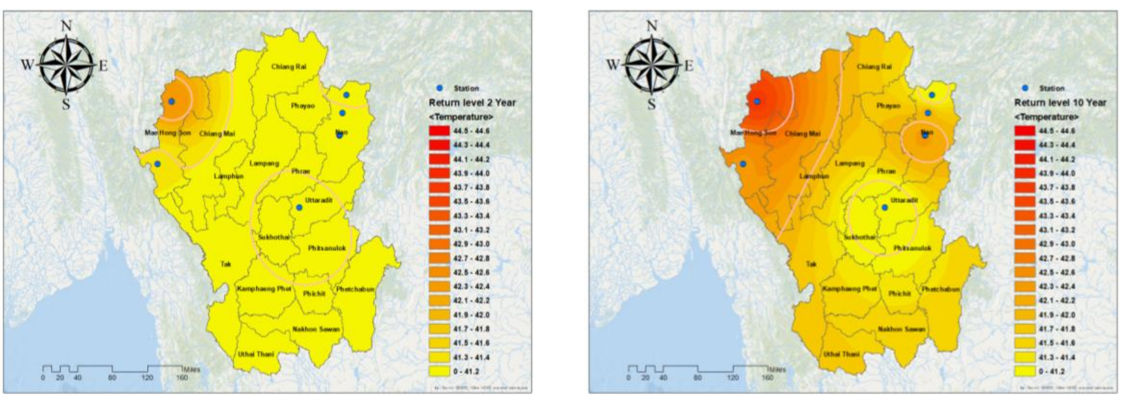

Figure 5 Return level of annual maximum temperature for return period at 2 and 10 years, respectively.

doi: 10.14416/j.appsci.2021.01.003

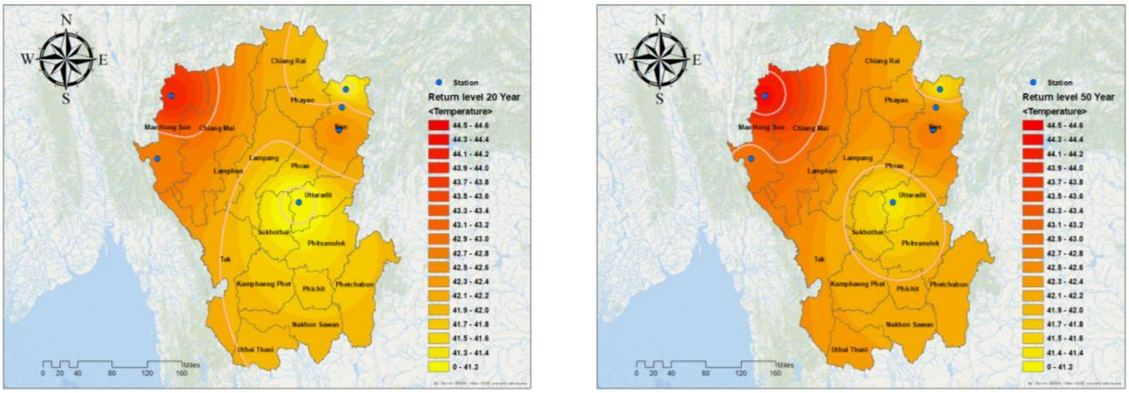

Figure 6 Return level of annual maximum temperature for return period at 20 and 50 years, respectively.

According to Figure 5 and Figure 6, return levels of 2, 10, 20, and 50 years of Mae Hong Son Thung Chang Army Headquarter Meteorological Station (300201) showed the highest temperature.

Conclusion and Discussion

The suitable model of daily maximum temperature by using r-largest order statistic with the theory of extreme value and finding return level of daily maximum temperature in annual with r-largest order statistic could be summarized and discussed as follows.

To consider the r-largest order statistic with extreme value theory, we found that r = 2 in Mae Sariang Meteorological station (300202) and Thung Chang Army Headquarter Meteorological station (331402), r = 3 in Mae Hong Son Thung Meteorological station (300201) and Uttaradit Meteorological station (351201) and r = 4 in Nan Meteorological station (331201), Nan Agrometeorological station (331301) and Tha Wang Pha Army Headquarter Meteorological station (331401). From the shape parameter of the suitable models, we could summarize that the distribution for daily maximum temperature in almost every station was Weibull distribution except Tung Chang Army Headquarter Meteorological station (331402) and Tha Wang Pha Army Headquarter Meteorological station (331401) that Gumbel distribution. It is corresponding with the study of Nemukula, M. M. and Sigauke, C. (2018) that the modeling of the highest average daily temperature with r-largest order statistics, based on data in South Africa, from 10 orders, was suitable when r $\leq$ 6.

Considering the return level in different return periods, we found that the return level increased when the return period rose in every station especially Mae Hong Son Meteorological Station (300201), which has the highest return level. Therefore, if relevant sectors desire to prevent or solve temperature problems, they should emphasize Mae Hong Son Meteorological station rather than others. Additionally, according to the report of the Thai Meteorological Department in February 2019, a number of provinces had increasing temperatures such as Lampang, Tak, Sukhothai, Nakhonsawan and Kanchanaburi. Hence, further study should include these areas.

Acknowledgement

The authors would like to express sincere gratitude to the Thai Meteorological Department for the kind support of daily maximum temperature data and the Faculty of Science, Mahasarakham University, for providing a venue and financial support in this research.

References

Bader, B., & Yan, J. (2016). eva: Extreme value analysis with goodness-of-fit testing. *R package version 0.2*, *4*.

Benjawan, C. (2014). *Extreme Value Model of Temperature in Mid-Northeastern Region of Thailand* [Unpublished master's thesis]. Khon Kaen University. (in Thai)

Coles, S., Bawa, J., Trenner, L., & Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*. London: Springer.

Enrique, C., Ali, S. Hadi, N., Balakrishnan, J., & Sarabia, M. (2004). *Extreme Value and Related Models with Applications in Engineering and Science*. The United States of America: Wiley.

Gumbel, E. J. (1958). *Statistics of extremes*. New York: Columbia University Press.

Nemukula, M. M., & Sigauke, C. (2018). Modeling average maximum daily temperature using r largest order statistics: An application to South African data. *Jàmbá: Journal of Disaster Risk Studies*, *10*(1), 1-11.

Senapeng, P., & Busababodhin, P. (2017). Modeling of maximum temperature in Northeast Thailand. *Burapha Science Journal*, 92-107. (in Thai)

Thai Meteorological Department. (2018). *Highest Temperature*. https://www.tmd.go.th/programs/uploads/tempstat/max_stat_latest.pdf

An, Y., & Pandey, M. D. (2007). The r largest order statistics model for extreme wind speed estimation. *Journal of wind engineering and industrial aerodynamics*, *95*(3), 165-182.