

Article

Adaptive Parameter Estimation of the Generalized Extreme Value Distribution Using Artificial Neural Network Approach

Tossapol Phoophiwfa ¹, Teerawong Laosuwan ², Andrei Volodin ³, Nipada Papukdee ⁴,
Sujitta Suraphee ¹ and Piyapatr Busababodhin ^{1,*}

- ¹ Digital Innovation Research Cluster for Integrated Disaster Management in the Watershed, Mahasarakham University, Kantarawichai, Maha Sarakham 44150, Thailand; 62010253001@msu.ac.th (T.P.); sujitta.s@msu.ac.th (S.S.)
- ² Department of Physics, Faculty of Science, Mahasarakham University, Maha Sarakham 44150, Thailand; teerawong@msu.ac.th
- ³ Department of Mathematics and Statistics, University of Regina, Regina, SK S4S 0A2, Canada; andrei.volodin@uregina.ca
- ⁴ Department of Applied Statistics, Rajamangala University of Technology Isan Khon Kaen Campus, Khon Kaen 40000, Thailand; nipada.pa@rmu.ac.th
- * Correspondence: piyapatr.b@msu.ac.th; Tel.: +66-92-542-6396

Abstract: Parameter estimation strategies have long been a focal point in research due to their significant implications for understanding data behavior, including the dynamics of big data. This study offers an advancement in these strategies by proposing an adaptive parameter estimation approach for the Generalized Extreme Value distribution (GEVD) using an artificial neural network (ANN). Through the proposed adaptive parameter estimation approach, based on ANNs, this study addresses the parameter estimation challenges associated with the GEVD. By harnessing the power of ANNs, the proposed methodology provides an innovative and effective solution for estimating the parameters of the GEVD, enhancing our understanding of extreme value analysis. To predict the flood risk areas in the Chi river watershed in Thailand, we first determine the variables that are significant in estimation of the three GEVD parameters μ , σ , and ξ by considering the respective correlation coefficient and then estimating these parameters. The data were compiled from satellite and meteorological data in the Chi watershed gathered from the Meteorological Department and 92 meteorological stations from 2010 to 2021, and consist of such variables as the Normalized Difference Vegetation Index (NDVI), climate, rainfall, runoff, and so on. The parameter estimation focuses on the GEVD. Taking into consideration that the processes could be stationary (parameters are constant over time, S) or non-stationary (parameters change over time, NS), maximum likelihood estimation and ANN approaches are applied, respectively. Both cases are modeled with the GEVD for the monthly maximum rainfall. The Nash-Sutcliffe coefficient (NSE), is used to compare the performance and accuracy of the models. The results illustrate that the non-stationary model was suitable for 82 stations, while the stationary model was suitable for only 10 stations. The NSE values in each model range from 0.6 to 0.9. This indicated that all 92 models were highly accurate. Furthermore, it is found that meteorological variables, geographical coordinates, and NDVI, that are correlated with the shape parameter in the ANN model, are more significant than others. Finally, two-dimensional maps of the return levels in the 2, 5, 10, 20, 50, and 100-year return periods are presented for further application. Overall, this study contributes to the advancement of parameter estimation strategies in the context of extreme value analysis and offers practical implications for water resource management and flood risk mitigation.

Keywords: generalized extreme value; artificial neural network; rainfall; runoff; normalized difference vegetation index



Citation: Phoophiwfa, T.; Laosuwan, T.; Volodin, A.; Papukdee, N.; Suraphee, S.; Busababodhin, P. Adaptive Parameter Estimation of the Generalized Extreme Value Distribution Using Artificial Neural Network Approach. *Atmosphere* **2023**, *14*, 1197. <https://doi.org/10.3390/atmos14081197>

Academic Editors: Iqbal Hossain and Abdullah Gokhan Yilmaz

Received: 19 June 2023
Revised: 19 July 2023
Accepted: 19 July 2023
Published: 25 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Climate changes have a direct impact on rainfall distribution and atmospheric fluctuations, affecting water resource management and hydrology. The northeastern region of Thailand frequently experiences flooding that causes significant damage. Over 75% of the country and about 15% of the total population suffer from water shortages every year, with an estimated annual damage value of around 19 million USA dollars. Given the recurring problem, there is an urgent need for efficient prediction and management of these extreme weather events. There were seven recent major floods, in the years 1983, 1995, 1996, 2002, 2006, 2011, and 2019, that caused direct damage to many lives and properties that cannot be assessed [1,2].

In 2011, Gale and Saunders [3] presented the cause of the 2011 major floods in Thailand and future flood forecasts, which showed that more flooding could occur within the next two to three decades unless flood defenses and flood management practices are improved. The study discovered that such locations were frequently damaged by flooding and the Chi River Basin area had a flood every year, according to the flood situation report [4–6]. In addition, the Chi Watershed has been experiencing flooding in many forms, including flooding in Roi Et, Kalasin, and Khon Kaen provinces, water overflowing the bank and wild water flows and mudslide in Kalasin and Chaiyaphum. The main causes of flooding problems were as follows: (1) long wet days and heavy rainfall occurring in the watershed area due to the southwest monsoon, northwest monsoon, and depression from the South China Sea; (2) the upstream area has a mountainous terrain, which is the origin of the Chi River, and many main streams with high slopes and increasing amount of forest destruction, reducing water retention and promoting fast run-off following rain; (3) the lower parts of the Chi watershed, especially in Roi Et province and Ubon Ratchathani province, are a plain. Many rivers flow together and it is also the point where the Chi River meets the Mun River before it flows into the Mekong River, causing problems in drainage from the watershed area; and (4) water management, the water in large reservoirs during the rainy season of some years must be drained in large quantities because the upstream area has a large amount of annual rainfall combined with the amount of water discharged from nearby reservoirs [7,8].

In the field of hydrology, the Generalized Extreme Value Distribution (GEVD) is commonly used to analyze extreme rainfall data [9]. Thus, parameter estimation methods for GEVD, and identifying significant explanatory variables for its parameters, are essential for modeling such extremes. Previous research has attempted parameter estimation using linear models and random forest models, focusing largely on climatic indices. Others have used the Index of Connectivity and Normalized Difference Vegetation Index (NDVI) for modeling rainfall-runoff and vegetation density, respectively. As [10,11] estimated and predicted the change in the parameters based on data from river basins in the United States and developed a model suitable for each area. The straight line and random forest models were compared with the simple Naive method of Jin et al. [12]. The parameters were found to mostly depend on climatic indices, while the selected prediction model was a linear one. Note that the methodology discussed in our article, can be applied to other statistical distributions related to the extreme value analysis. On this regard, we mention two articles [13,14]. In [13], a useful software for regional and at-site statistical analyses of annual maxima time series is proposed. The software showed good performance with the annual maxima series for the Italian rain-gauge network by implementing Gumbel, GEV and Two Component Extreme Value distributions. In [14], the results about the derivation of the exact distribution of maximum annual daily precipitation are summarized and discussed, with special attention on compound/super statistical distributions. The results have a 20% higher accuracy in terms of the Root Mean Square Error (RMSE) value, compared to the Naive approach [11]. Haniyeh et al. [15] created a rainfall-runoff model with the Index of Connectivity, which is considered as a hydro-geomorphic tool to investigate the flow connection within catchments. The NDVI is the index that indicates the density and abundance of vegetation [16]. The artificial neural network (ANN) model by [17,18] had

an efficiency of 97% when applied to the Haughton River in Australia. Rotjanakusol and Laosuwan [19] studied the relationship between the NDVI with the amount of rainfall obtained from the measurement station of the Thai Meteorological Department. The NDVI had a value between 80–97 in the lower northeast region of Thailand.

This study aims to contribute to the ongoing efforts of modeling extreme rainfall events by proposing an adaptive parameter estimation method using an ANN approach. We focus on the monthly maximum rainfall data for the Chi river watershed in Thailand, employing two models: a stationary model that uses maximum likelihood estimation, and a non-stationary model using ANN. This study presents the application of ANN to estimate time-changing GEVD parameters, a novel approach in this context. We also estimate the 2, 5, 10, 20, 50, and 100-year return periods for the Chi watershed with the R program and illustrate this with spatial maps using the Quantum Geographic Information System program. Supplementary Material includes technical specifics, tables, and figures.

2. Study Area

The data used in this study were provided by the Meteorological and Royal Irrigation Departments of Thailand and Terra-MODIS MOD13Q1 package satellite image data for the years 2010–2021 [20]. Figure 1 shows the locations of all 92 meteorological stations along the Chi watershed for the 12 provinces in the northeastern region of Thailand. Table 1 shows the attribute types, notation, and predictor variables for our models.

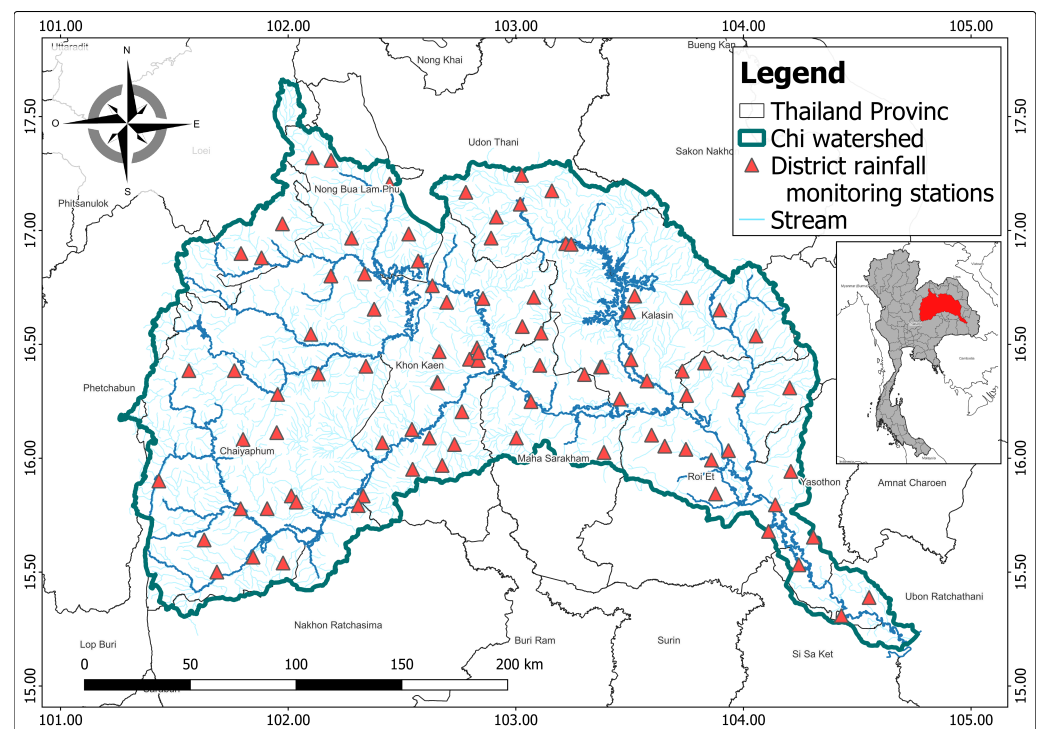


Figure 1. Location of all 92 stations along Chi Watershed in northeastern region of Thailand.

Table 1. Attribute types, notation, and predictor variables.

Attribute Type	Attribute	Notation
GEVD attributes	Location parameter	μ or μ (x_1)
	Scale parameter	σ or σ (x_2)
	Shape parameter	ξ or ξ (x_3)
Geographical coordinates	Latitude	LAT (x_4)
	Longitude	LON (x_5)
Meteorological variables	Maximum rainfall	max_rain (x_6)
	Average rainfall	average_rain (x_7)
	Cumulative rainfall	sum_rain (x_8)
	Average minimum rainfall	min_average_rain (x_9)
	Maximum wind speed	max_wind (x_{10})
	Average wind speed	average_wind (x_{11})
	Maximum temperature	max_temp (x_{12})
	Minimum temperature	min_temp (x_{13})
	Average temperature	average_temp (x_{14})
	Average relative humidity	average_RH (x_{15})
Satellite images	NDVI	NDVI (x_{17})
Hydrological variables	Maximum runoff	max_runoff (x_{18})
	Average runoff	average_runoff (x_{19})

Note: The parameters μ , σ and ξ in this table are deterministic and represent the location, scale, and shape parameters of the GEVD, respectively.

Figure 2 shows the histogram distribution of the data for the 19 predictor variables. Each of the 92 meteorological stations are specified by their geographical coordinates of latitude (x_4) and longitude (x_5). For each station, we observed the monthly meteorological, satellite image, and hydrological attribute type, which are the fourteen variables ($x_6 - x_{19}$) for the 12 years of 2010–2021. Thus, for each of the 92 stations we have $12 \times 12 = 144$ observations for each of the fourteen variables. Variables $x_4 - x_{19}$ are considered to be *input* variables for our model. The *output* variables are the GEVD parameters (ξ , μ , and σ) for each of the 92 stations, which are x_1 , x_2 , and x_3 in the notation of Table 1.

**Figure 2.** Histogram showing the distribution of data for all 19 variables.

3. Methodology

3.1. The GEVD

The GEVD was initially developed in [21], where three extreme distribution functions can be written: Gumbel distribution, Frechet distribution, and Weibull distribution [22,23]. The block maxima method was used to select the maximum observation (or minimum observation) $M_n = \max(x_1, x_2, \dots, x_n)$ for analysis determined by the process without changing the parameters. For the estimating GEVD parameters by MLE, the probability distribution function was structured as shown in Equation (1):

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

defined on the set $1 + \xi(x - \mu)/\sigma > 0$, where the parameters satisfy $-\infty < \mu < \infty, \sigma > 0$ and $-\infty < \xi < \infty$. This is the generalized extreme value (GEV) family of distributions. The model has three parameters: location parameter μ ; scale parameter σ , and shape parameter ξ . The type II and type III classes of extreme value distributions correspond to the cases $\xi > 0$ and $\xi < 0$, respectively, in this parametrization. The subset of the GEV family with $\xi = 0$ is interpreted as $\xi \rightarrow 0$, leading to the Gumbel family [22].

3.2. The Maximum Likelihood Estimation

The parameters of the GEVD can be estimated using the MLE approach. Although the MLE approach leads to underestimation for small sample sizes, it has a small variance for large sample sizes. The MLE is often considered better than L-moment estimation for large sample sizes [24]. For $\xi \neq 0$, the log-likelihood function can be obtained directly from Equation (1) [22]:

$$\begin{aligned} \ell(\mu, \sigma, \xi) = & -n \log \sigma - \left(1 + \frac{1}{\xi} \right) \sum_{i=1}^n \log \left(1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right) \\ & - \sum_{i=1}^n \left(1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right)^{-1/\xi}, \end{aligned}$$

provided that $1 + \xi \left(\frac{x - \mu}{\sigma} \right) > 0$.

3.3. The Return Level

The return level or quantiles is used to interpret extreme values. The return level for the return period T for the GEVD is defined as shown in Equation (2)

$$\hat{Z}_T = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} \left(1 - \left(-\log \left(1 - \frac{1}{T} \right) \right)^{-\hat{\xi}} \right). \quad (2)$$

The return level associated with the return period $T = \frac{1}{p}$, where p is the probability of the year in which $\hat{Z}_T$ exceeds the annual maximum, since, to a reasonable degree of accuracy, the level $\hat{Z}_T$ is expected to be exceeded on average once every T years. More precisely, $\hat{Z}_T$ is exceeded by the annual maximum in any particular year with probability p [22].

3.4. The ANN Approach

The ANNs are a Machine Learning algorithms based on some non-linear functions for the weighted sums of input variables. The ANN procedure can be visualized (Figure 3) as a graph with nodes (neurons) and connections between them that are called edges. Neurons and edges are typically assigned with weights w_{ij} that adjust the learning procedure [25]. The weight increases or decreases according to the strength of the signal at the connection. Neurons may have a threshold such that a signal is only sent if the aggregate signal crosses that threshold. Typically, neurons are aggregated into layers. Different layers may perform different transformations on their inputs. Signals travel from the first layer x_i (the input

layer) to the last layer (the output layer), possibly after traversing the layers multiple times. The weights w_{ij} , bias b_j , the weight summing junction v_j , and the activation function φ provide the output [26] as shown in Equation (3):

$$\text{Output} = \left(\sum_{j=1}^m \left(\sum_{i=1}^n (w_{ij}x_i) + b_j \right) v_j \right) \varphi. \quad (3)$$

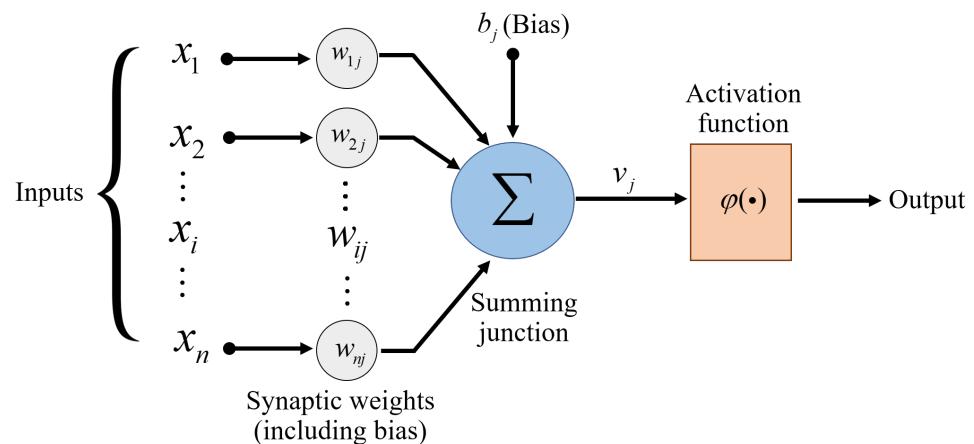


Figure 3. Structure of ANNs.

The structure of the ANN model is presented in the form of ANN($a - b - c$), where a, b and c are the number of input variables, number of hidden layers, and number of output variables, respectively.

ANN modeling is a process used to build models that can learn and predict data efficiently. It has the following steps.

Step 1. Validate and prepare the data with the independent variables being GEVD parameters and four groups of dependent variables: Geographical coordinates, Meteorological variables, Satellite images, and Hydrological variables (altogether 19 variables).

Step 2. Dataset division: Divide the dataset into training set and test data set (test set) for training and evaluating model performance using a ratio of 70% to 30%.

Step 3. Define model structure and parameters: Select and define the appropriate ANN structure for the desired task, such as the number of layers (layers) and the number of nodes (nodes) in each layer. In our study, we start with 1–20 in each layer and after find the suitable number of nodes, package “neuralnet” in R program [27]. We got 740 models per parameter.

Step 4. Build and prepare ANN model using R program.

Step 5. Evaluation and improvement: The performance of the ANN model is evaluated using the test data set. After, we improve the structure or parameters of the model to achieve the best results with NSE and RMSE.

3.5. Pearson Product-Moment Correlation Coefficient

The Pearson correlation coefficient is used to measure the relationship between two paired variables $(x_i, y_i), i = \{1, 2, \dots, n\}$. The variables should be on an interval or ratio scale. The values of the correlation coefficient lie between -1 to 1 and can be calculated as shown in Equation (4):

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (4)$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ and $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ [28].

3.6. Model Accuracy Verification

3.6.1. The RMSE

The RMSE is used to compare the estimated and the real values of three parameters in the GEVD. It is defined as show in Equation (5):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\theta_{ik} - \hat{\theta}_{ik})^2}, \quad k = \{1, 2, 3\}, \quad (5)$$

where θ_{ik} are the real values of parameters μ, σ, ξ , $\hat{\theta}_{ik}$ are the estimated values $\hat{\mu}, \hat{\sigma}, \hat{\xi}$, and n is the number of data.

3.6.2. Nash–Sutcliffe Model Efficiency Coefficient (NSE)

In 1970 [29], Nash and Sutcliffe assessed the suitability of models with the NSE coefficient, which is a measure of model accuracy, or model performance. The NSE values range from $-\infty$ to a maximum of 1, and compares the data value with the forecast value using the same quantile position, as show in Equation (6):

$$NSE = 1 - \frac{\sum_{i=1}^n (Q_i - Q_i^{pred})^2}{\sum_{i=1}^n (Q_i - \bar{Q}_i)^2}, \quad (6)$$

where Q_i are the quantile function of observed monthly maximum rainfall, Q_i^{pred} is the quantile function of the modeled monthly maximum rainfall, $\bar{Q}_i$ is the mean of the quantile function of observed monthly maximum rainfall, and n is the number of data. The summary of the operational procedures and analytical methods employed in this study for ANN modeling to estimate the parameters of the GEVD is illustrated in Figure 4. It shows the data collection phase, four types of data were gathered, including satellite images, meteorological data, hydrological data, and geographical coordinates. The collected dataset was then subjected to a data cleansing process. Following this, the parameters of the GEVD were estimated using both stationary and non-stationary approaches. The NSE was utilized to select appropriate models, and flood risk areas were predicted through the estimation of return levels.

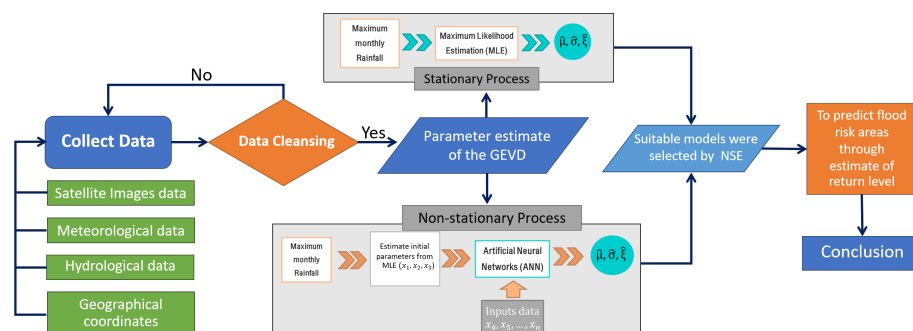


Figure 4. Map of the research methodology.

4. Results and Discussion

4.1. Parameter Estimation of the GEVD

The results of the parameter estimation of the GEVD parameters μ, σ , and ξ from the monthly maximum rainfall data from 92 meteorological stations using the MLE approach are shown in Figures 5–7:

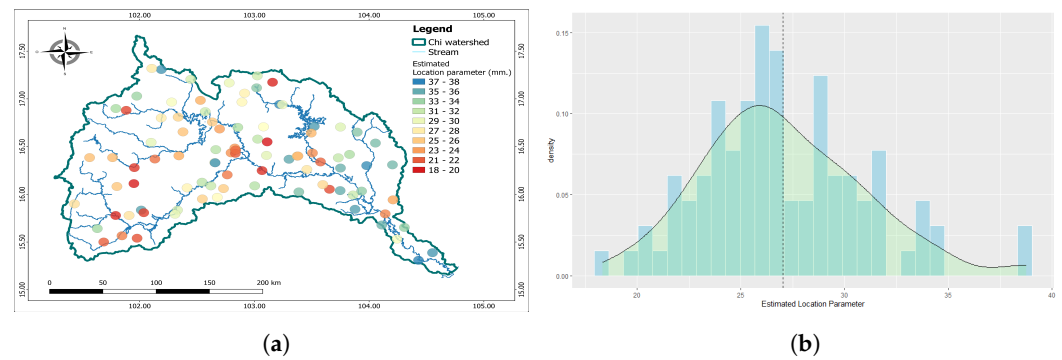


Figure 5. Estimates of location parameter ($\hat{\mu}$) for 92 stations based on monthly maximum rainfall. (a) Map of estimated location parameter ($\hat{\mu}$). (b) Histogram of estimated location parameter ($\hat{\mu}$).

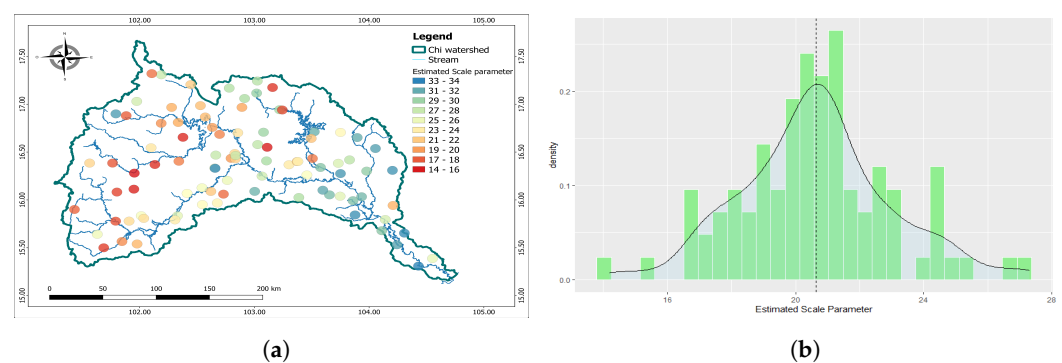


Figure 6. Estimates of scale parameter ($\hat{\sigma}$) for 92 stations based on monthly maximum rainfall. (a) Map of estimated scale parameter ($\hat{\sigma}$). (b) Histogram of estimated scale parameter ($\hat{\sigma}$).

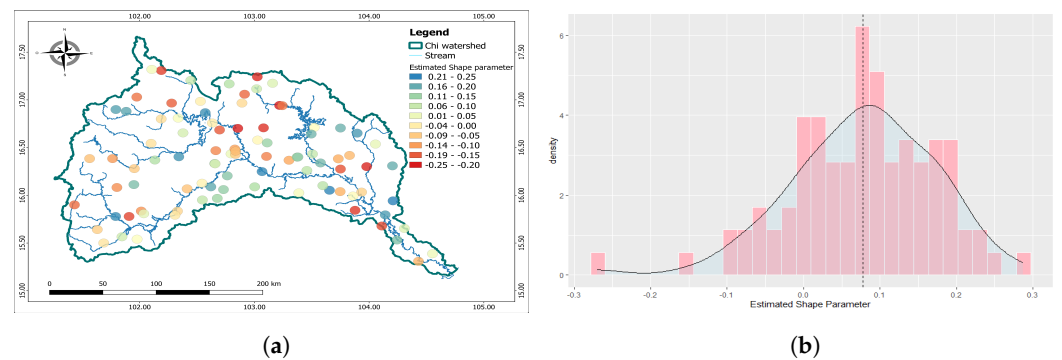


Figure 7. Estimates of shape parameter ($\hat{\xi}$) for 92 stations based on monthly maximum rainfall. (a) Map of estimated shape parameter ($\hat{\xi}$). (b) Histogram of estimated shape parameter ($\hat{\xi}$).

For the estimation results for the location parameter μ of the GEVD, the values ranged from 20 to 38 with an average of 27 (Figure 5a), and the distribution of these values did not conform to a specific parametric distribution, showing skewness to the right (Figure 5b). These values were then used as the initial parameter value for the estimation procedure for the other variables and applied to the ANN models as well.

With respect to the estimation results for the scale parameter σ of the GEVD, the values ranged from 14 to 33 with an average of 21 (Figure 6a), and the data were symmetrically distributed (Figure 6b). The initial parameter values for our models, including the ANN model, were chosen based on the distribution of the location parameter, μ , of the GEVD. These values provided a sensible starting point for the optimization procedures in our models, as they reflect the empirical distribution of the data. By using these as initial estimates, we could enhance the efficiency and stability of the parameter estimation process. In the revised manuscript, we have included a more detailed explanation of this rationale for our choice of initial parameter values.

Finally, for the estimation results for the shape parameter ξ of the GEVD, the values ranged from -0.25 to 0.25 with an average of 0.09 (Figure 7a), but the data distribution was non-parametric being skewed to the left (Figure 7b). These values were used as the initial parameter value for the estimation procedure of the other variables and applied to the ANN models as well.

4.2. Relationship between Variables

The Pearson correlation coefficient is applied to analyze any correlation between the variables; that is, the relationship between the parameters μ , σ and ξ and the 16 other variables. The results of the relationship analysis between the variables are shown in Figure 8.

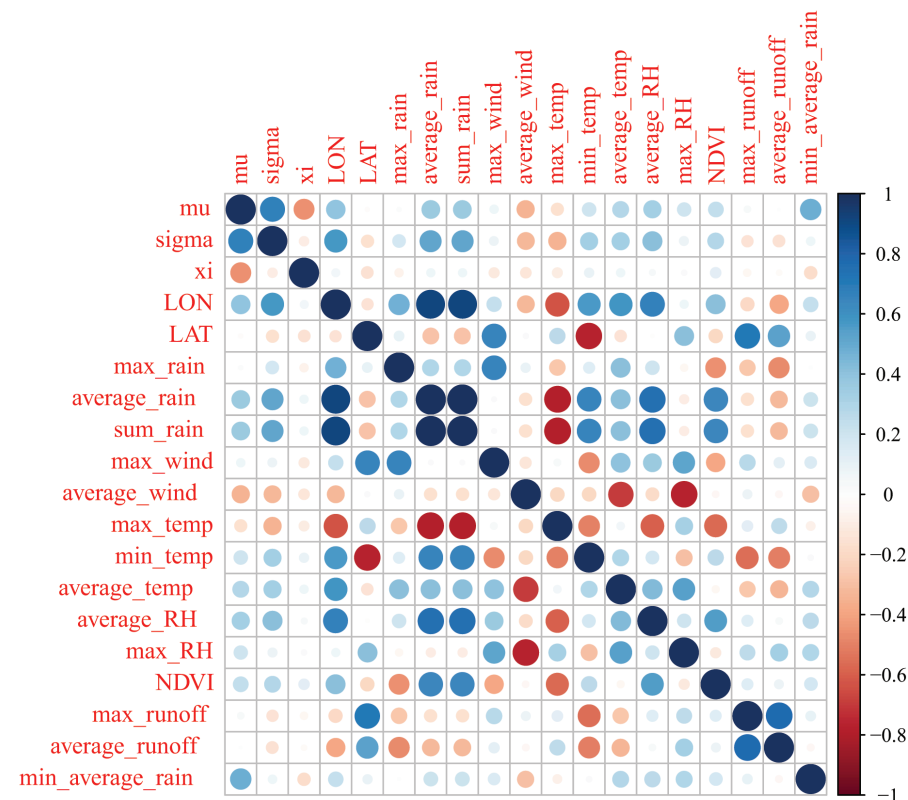


Figure 8. Correlations between the predictor variables.

Figure 8 shows that the variables σ , average minimum rainfall, ξ , NDVI, longitude, average rainfall, and cumulative rainfall are all significantly correlated with the location parameter μ . The variables μ , longitude, average rainfall, cumulative rainfall, average relative humidity, and NDVI are all significantly correlated with the scale parameter σ . The variables μ , maximum Rainfall, NDVI, and maximum temperature are significantly correlated with the shape parameter ξ .

4.3. The ANN Model

In this section, we develop an ANN for estimating the three GEVD parameters. To validate the derived model, we use the cross-section method with using 70% of the data for training the ANN model and 30% for testing it. Tables 2–4 show the best ANN model in bold font for estimation of each GEVD parameter μ , σ and ξ , while Figures 9–11 show the structure of the best ANN model for each parameter.

Table 2. Structure of the ANN models for estimating the GEVD location parameter μ with structure modeling and evaluation of the model performance by RMSE and NSE.

Model	Input Variable	Structure	RMSE	NSE
ANN01	$\sigma, \xi, \text{LAT, LON, max_rain, average_rain, sum_rain, min_average_rain, max_wind, average_wind, max_temp, min_temp, average_temp, average_RH, max_RH, NDVI, runoff_max, average_runoff}$	18-1-1	2.1848	0.7541
ANN02	$\sigma, \xi, \text{LAT, LON, max_rain, average_rain, sum_rain, min_average_rain, max_wind, average_wind, max_temp, min_temp, average_temp, average_RH, max_RH, NDVI, max_runoff, Average_Runoff}$	18-20-1	2.4424	0.6927
ANN03	$\sigma, \xi, \text{LAT, LON, max_rain, average_rain, sum_rain, min_average_rain, max_wind, average_wind, max_temp}$	11-4-1	2.3731	0.6837
ANN04	$\sigma, \text{LON, average_rain}$	3-14-1	2.8328	0.6047
ANN05	$\sigma, \text{LON, average_rain}$	3-16-1	2.8452	0.6012
ANN06	$\sigma, \text{LON, average_rain}$	3-11-1	2.8453	0.6012
ANN07	$\text{max_rain, average_rain, sum_rain}$	3-4-1	3.0436	0.2307
ANN08	$\text{max_rain, average_rain, sum_rain}$	3-16-1	3.0573	0.2237
ANN09	$\text{max_rain, average_rain, sum_rain}$	3-5-1	3.0665	0.2190

Note: Structure of ANN or ANN($a - b - c$), where a, b , and c are the number of input variables, number of hidden layers and number of output variables, respectively.

As it is shown in Table 2, the best model for estimating the GEVD location parameter μ is ANN01 based on it having the smallest RMSE (2.1848) and largest NSE (0.7541) values. The model can be seen more clearly in Figure 9, where we present the structure of ANN01 model.

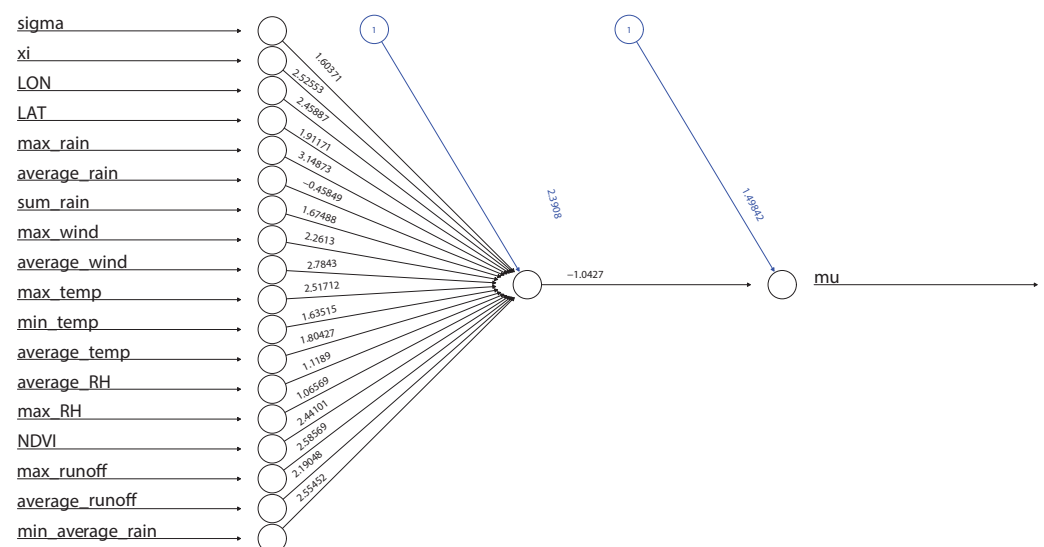
**Figure 9.** Structure of the ANN model ANN01 (18-1-1).

Table 3. Structure of the ANN models for estimating the GEVD scale parameter σ with structure modeling and evaluation of the model performance by RMSE and NSE.

Model	Input Variable	Structure	RMSE	NSE
ANN10	μ , ζ , LAT, LON, max_rain, average_rain, sum_rain, min_average_rain, max_wind	9-1-1	1.6799	0.5998
ANN11	μ , ζ , LAT, LON, max_rain, average_rain, sum_rain, min_average_rain, max_wind, average_wind, max_temp, min_temp, average_temp	13-14-1	1.7302	0.5920
ANN12	μ , ζ , LAT	3-10-1	1.7400	0.5330
ANN13	μ	1-1-1	1.9115	0.4450
ANN14	μ	1-2-1	1.9151	0.4429
ANN15	μ	1-3-1	1.9146	0.4432
ANN16	max_rain, average_rain, sum_rain, max_wind, average_wind, max_temp, min_temp, average_temp, average_RH, max_RH, NDVI, max_runoff	12-9-1	1.6006	0.5323
ANN17	ζ , average_rain, sum_rain, max_wind, average_wind, max_temp, min_temp, average_temp, average_RH, max_RH, NDVI, max_runoff	12-1-1	1.6455	0.5057
ANN18	max_rain, average_rain, sum_rain, max_wind, average_wind, max_temp	6-20-1	2.1025	0.4847

Note: Structure of ANN or ANN($a - b - c$), where a , b , and c are the number of input variables, number of hidden layers and number of output variables, respectively.

As shown in Table 3, the best model for estimating the GEVD scale parameter σ is ANN10, based on it having the smallest RMSE (1.6799) and largest NSE (0.5998) values. The structure of the ANN10 model is summarized more clearly in Figure 10.

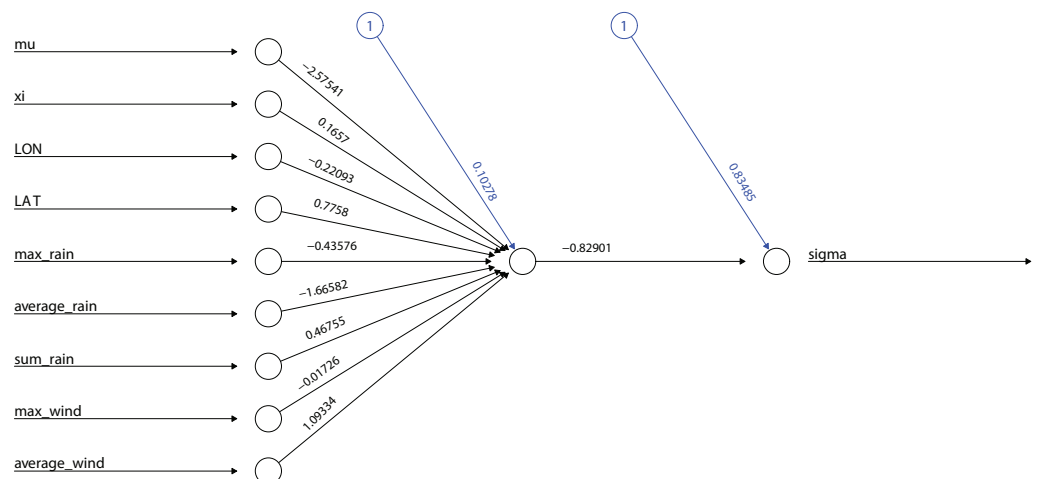
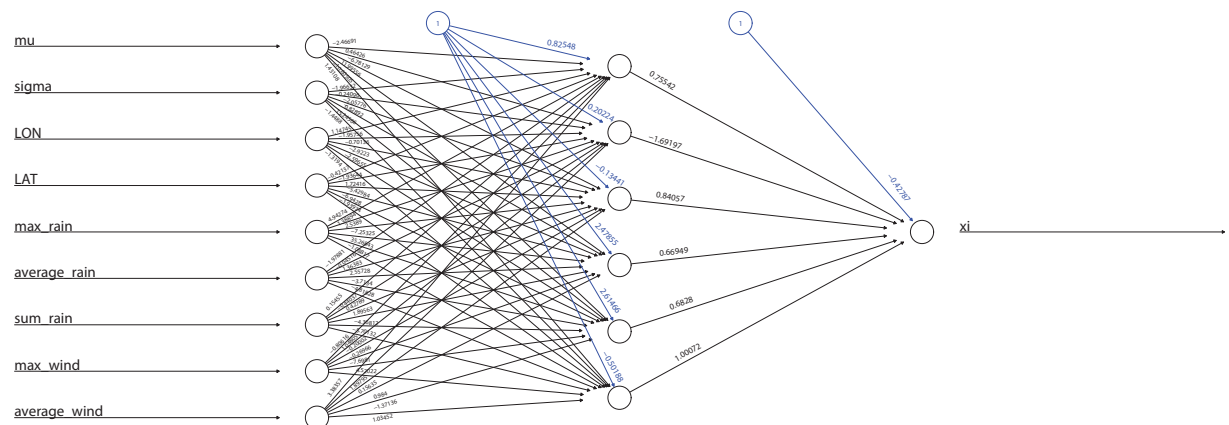
**Figure 10.** Structure of the ANN model ANN10 (9-1-1).

Table 4. Structure of the ANN models for estimating the GEVD shape parameter ξ with structure modeling and evaluation of the model performance by RMSE and NSE

Models	Input Variable	Structure	RMSE	NSE
ANN19	$\mu, \sigma, \text{LAT, LON, max_rain, average_rain, sum_rain, min_average_rain, max_wind, average_wind}$	9-6-1	0.0740	0.4866
ANN20	$\mu, \sigma, \text{LAT, LON, max_rain, average_rain}$	6-2-1	0.0807	0.4669
ANN21	$\mu, \sigma, \text{LAT, LON, max_rain, average_rain}$	6-16-1	0.0833	0.4319
ANN22	$\mu, \text{LAT, LON, average_rain, sum_rain}$	5-17-1	0.0651	0.4734
ANN23	$\mu, \text{LAT, LON, average_rain, sum_rain}$	5-2-1	0.0673	0.4374
ANN24	$\mu, \text{LAT, LON, average_rain, sum_rain, average_wind, max_temp, min_temp, average_temp, average_RH, NDVI, min_average_rain}$	12-4-1	0.0843	0.4250
ANN25	max_rain, average_rain, sum_rain, max_wind	4-19-1	0.0803	0.0931
ANN26	max_rain, average_rain, sum_rain, max_wind	4-16-1	0.0804	0.0903
ANN27	max_rain, average_rain, sum_rain, max_wind	4-10-1	0.0807	0.0826

Note: Structure of ANN or ANN($a - b - c$), where a, b , and c are the number of input variables, number of hidden layers and number of output variables, respectively.

**Figure 11.** Structure of the ANN model ANN19 (9-6-1).

As shown in Table 4, the best model for estimating the GEVD shape parameter ξ is ANN19, based on having the smallest RMSE (0.0740) and largest NSE (0.4866) values. For clarity, the structure of the model ANN19 is shown in Figure 11.

In the next step, the three selected best ANN models (ANN01(18-1-1), ANN10 (9-1-1), and ANN19 (9-6-1)) are applied to the parameter estimation of the GEVD. Each of three GEVD parameters can be estimated by the (S) or (NS) procedures. Hence, there are $2^3 = 8$ possible estimation procedures (Table 5).

For this study, we deployed our adaptive method for the data from the 92 meteorological stations, and the eight possible GEVD estimation procedures are applied to each station with the results shown in Figure 12. The top three best procedures are found to be GEVD08(NS-S-NS), GEVD07(NS-S-S), and GEVD02(S-NS-S).

There are 10 stations for which the S procedure model was optimal for all three parameters, namely, for GEVD01(S-S-S), while there were seven stations that are best fitted by the NS procedure for all three parameters, namely, for GEVD05(NS-NS-NS) (Table 5). From these results, and those in Figure 12, we can conclude that mixed (S and NS) procedures give a better performance than those with a single (S or NS) procedure model for all three GEVD parameters. In particular, there are 16 stations along the Chi watershed that are suitable for the GEVD08(NS-S-NS) procedure, followed by GEVD07(NS-S-S) and GEVD02(S-NS-S) for 15 stations.

Table 5. Parameter estimation procedures with the number of suitable stations.

Model	Parameters Estimations Process			Number of Stations (Percentage)
	μ (mu)	σ (sigma)	ξ (xi)	
GEVD01	S	S	S	10 (10.87%)
GEVD02	S	NS	S	15 (16.30%)
GEVD03	S	NS	NS	11 (11.96%)
GEVD04	S	S	NS	7 (7.61%)
GEVD05	NS	NS	NS	7 (7.61%)
GEVD06	NS	NS	S	11 (11.96%)
GEVD07	NS	S	S	15 (16.30%)
GEVD08	NS	S	NS	16 (17.39%)

Note: S is stationary process and NS is a non-stationary process from the ANN model.

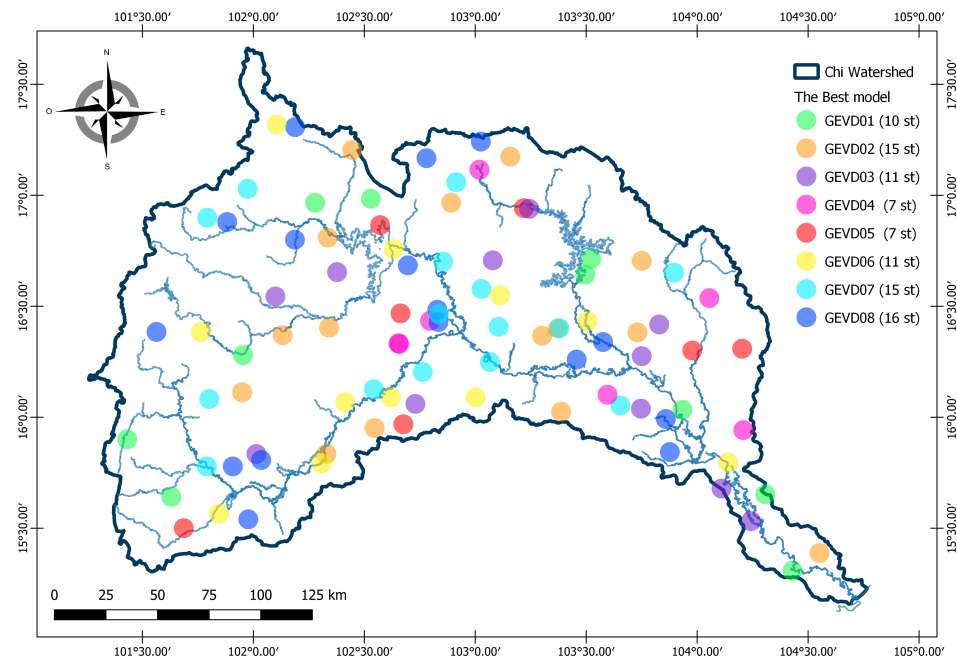


Figure 12. Map showing the suitable estimation procedure for each meteorological station.

4.4. Return Level Estimation

Figure 13 shows return level estimation for 2, 5, 10, 20, 50 and 100-year return period for the annual maximum monthly rainfall data in the Chi watershed area.

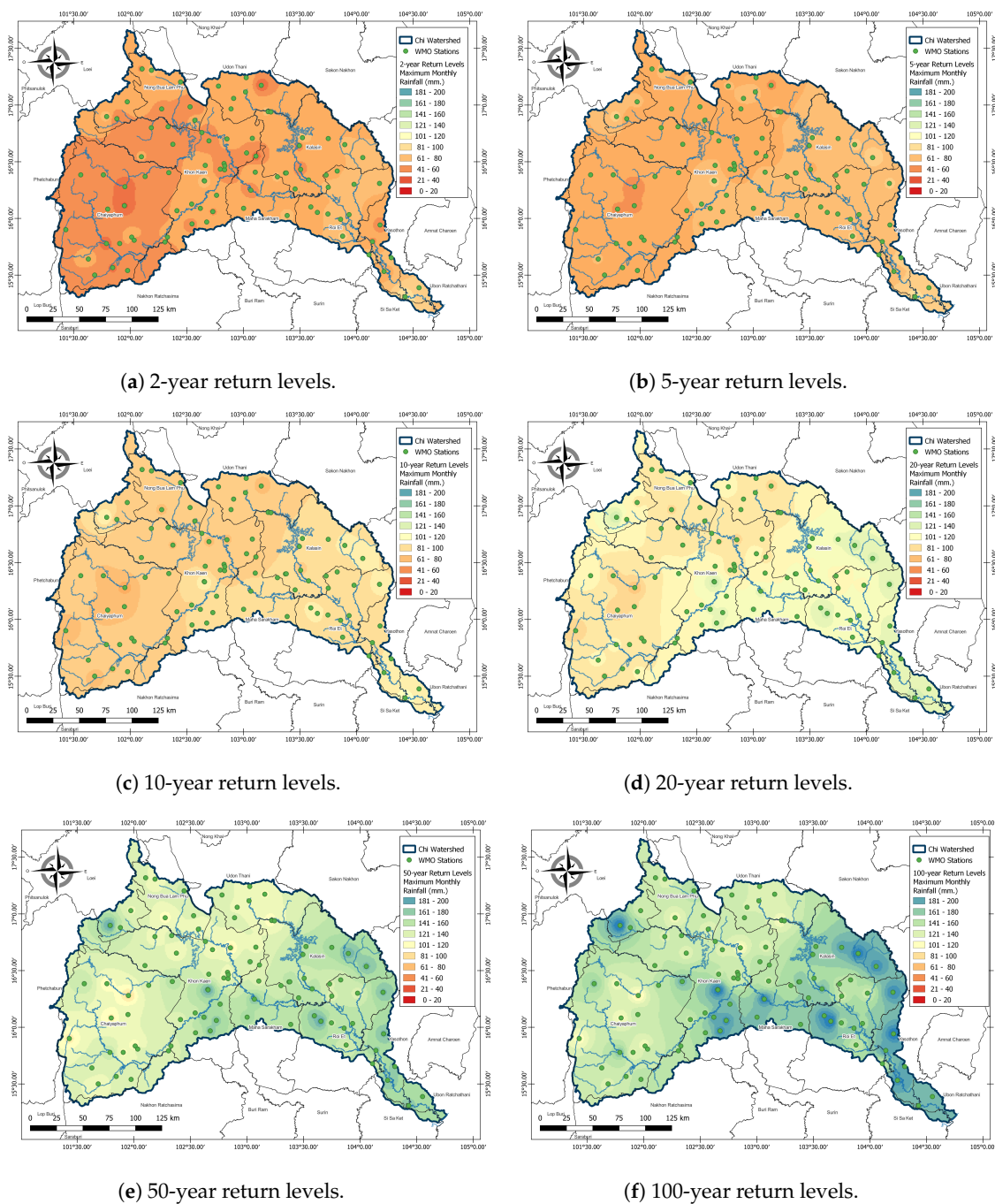


Figure 13. Maps of the estimated 2, 5, 10, 20, 50 and 100-year return period for the annual maximum monthly rainfall data in the Chi watershed area.

The return level maps for a 2, 5, 10, 20, 50 and 100-year return period of the monthly maximum rainfall data show the areas at risk of flooding within that time period. The flood risk areas are found to be in the provinces located on the lower left edge of the Chi watershed: Yasothorn, Ubon Ratchathani, Roi Et, Kalasin, and Maha Sarakham. These maps were created using the Q-GIS program, and show that the return level increases every year for all stations. This clearly means that planning for future rainfall management is essential.

5. Discussion

This study models the estimation procedure of the location, scale, and shape parameters of the GEVD with usage of the maximum rainfall and satellite image data. The main analytical approach is to use an ANN, with the main purpose being to determine the variables affecting the change in the three GEVD parameters μ , σ , and ξ when the monthly rainfall data are considered. The correlation between the variables is investigated using the Pearson correlation coefficient and then the parameters of the GEVD are modeled using the ground and satellite based data. The ANN models are then used to predict flood prone areas by producing two-dimensional maps of the return level within specified period of a 2 to 100-year period. The conclusion of the analysis can be summarized into 3 parts.

(1) The analysis of all 19 variables (Table 1) and the Pearson correlation coefficient (Figure 8) revealed that (a) there are seven variables σ , average minimum rainfall, ξ , NDVI, longitude, average rainfall, and cumulative rainfall that are significantly correlated with the estimates of GEVD location parameter μ with Pearson's correlation coefficients of 0.68, 0.48, 0.45, -0.46 , 0.39, 0.36, 0.36, and 0.34, respectively (b) there are six variables μ , longitude, average rainfall, cumulative rainfall, average relative humidity, and NDVI that are significantly correlated with the estimates of the GEVD scale parameter σ with Pearson's correlation coefficients of 0.68, 0.57, 0.51, 0.51, 0.42, and 0.36, respectively (c) there are four variables μ , maximum rainfall, NDVI, and maximum temperature that are significantly correlated with the estimates of the GEVD shape parameter ξ with Pearson's correlation coefficients -0.46 , -0.29 , 0.23 and -0.21 , respectively.

(2) The results from the ANN model for estimating the three GEVD parameters μ , σ , and ξ are as follows. The cross-validation method for the ANN model used 70% of the data for training and 30% for testing. The model structure consists of an input layer, a hidden layer, and an output layer. The best models for μ , σ , and ξ are ANN01, ANN10 and ANN19, respectively, based on having the smallest RMSE and highest NSE values. The suitable models can be written as:

$$\begin{aligned}\hat{\mu} &= \left(\sum_{j=1}^1 \left(\sum_{i=2}^{19} (w_{ij}y_i) + b_j \right) v_j \right) \varphi, \\ \hat{\sigma} &= \left(\left((w_{1,1}x_1 + w_{3,1}x_3 + w_{4,1}x_4 + w_{5,1}x_5 \right. \right. \\ &\quad \left. \left. + w_{6,1}x_6 + w_{7,1}x_7 + w_{8,1}x_8 + w_{9,1}x_9 + w_{10,1}x_{10}) + b_1 \right) v_1 \right) \varphi, \\ \hat{\xi} &= \left(\sum_{j=1}^6 \left((w_{1,j}x_1 + w_{2,j}x_2 + w_{4,j}x_4 + w_{5,j}x_5 + w_{6,j}x_6 \right. \right. \\ &\quad \left. \left. + w_{7,j}x_7 + w_{8,j}x_8 + w_{9,j}x_9 + w_{10,j}x_{10} + w_{11,j}x_{11}) \right) + b_j \right) v_j \varphi.\end{aligned}$$

(3) Comparing the model performance between the S and NS procedures using NSE, is concluded that the mixed (S and NS) procedures gave a better performance than the single (S or NS) procedure model for estimating all three GEVD parameters. In particular, there are 16 stations along the Chi watershed that are suitable for the GEVD08 (NS-S-NS) procedure, followed by GEVD07 (NS-S-S) and GEVD02 (S-NS-S) for 15 stations.

The Modelling of the three GEVD parameters from the monthly maximum rainfall using MLE and ANN is consistent with Hristos et al. [11], who found that the given parameters were related to the weather conditions and terrain characteristics. They also found that the NDVI affects the parameters. In accord Haniyeh Asadi et al. [15] developed an ANN model for rainfall-runoff forecasts using the NDVI and Hydrological Connectivity Index, two climatic and hydrological factors.

Considering both S and NS procedures to model the GEVD, the NS process is caused by a number of factors, with the main contributing factors being climate change and topography, which results in the lower efficiency and accuracy of the S process modeling. On the other hand, the NS process was suitable for the global environment that was constantly changing.

6. Conclusions

The conclusion of this study encompasses several important findings that contribute to the understanding and modeling of extreme rainfall events using the GEVD.

Firstly, the analysis revealed significant correlations between certain variables and the estimates of the GEVD parameters (μ, σ, ξ). The variables that exhibited strong correlations include average minimum rainfall, NDVI, longitude, average rainfall, cumulative rainfall, and average relative humidity. These correlations provide valuable insights into the factors influencing extreme rainfall behavior, such as local climate conditions and geographical characteristics. Understanding these relationships can enhance our ability to predict and manage flood risks. Secondly, the adaptive parameter estimation approach using the ANN demonstrated its effectiveness in estimating the GEVD parameters. The ANN models, specifically ANN01, ANN10, and ANN19, showed superior performance in terms of the RMSE and NSE values. This highlights the capability of ANN in capturing the complex relationships between input variables and GEVD parameters, enabling more accurate parameter estimation and improved flood risk assessments. Furthermore, the comparison between S and NS procedures revealed the importance of considering both approaches in GEVD modeling. The mixed S-NS procedures outperformed individual S or NS procedure models in estimating the GEVD parameters. This indicates the influence of climate change and topography on extreme rainfall patterns, emphasizing the need to account for temporal variability and non-stationarity in modeling extreme events. Neglecting these factors can lead to biased estimations and inadequate risk assessments.

The findings of this study have practical implications for flood risk management and water resource planning. By understanding the key variables and their correlations with GEVD parameters, decision-makers can better identify areas prone to extreme rainfall events and allocate resources for effective mitigation strategies. The ANN-based approach offers a valuable tool for parameter estimation and prediction of flood-prone areas, enabling timely and accurate decision-making.

In conclusion, this study provides valuable insights into the estimation of GEVD parameters for extreme rainfall modeling. The significant correlations observed between variables and GEVD parameters, along with the successful implementation of the ANN approach, enhance our understanding of extreme rainfall behavior. The consideration of non-stationarity further improves the modeling accuracy. These findings contribute to better flood risk management and water resource planning in the face of changing climate conditions.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/atmos14081197/s1>.

Author Contributions: Conceptualization, P.B. and T.P.; methodology, P.B.; software, T.P.; validation, T.P., A.V. and S.S.; formal analysis, T.L.; investigation, N.P.; resources, P.B.; data curation, T.P.; writing—original draft preparation, P.B.; writing—review and editing, P.B., S.S. and N.P.; visualization, T.P.; supervision, A.V.; project administration, P.B. and S.S.; funding acquisition, P.B. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported under that framework of international cooperation program managed by the Mahasarakham University, Thailand. Piyapatr's work was supported by Mahasarakham University (No.6517004/2565). And Piyapatr's work also was funded by the Agricultural Research Development Agency (Public Organization) of Thailand, (ARDA).

Institutional Review Board Statement: Not applicable: This study does not involve human participants or animals. Therefore, ethical review and approval were not required.

Informed Consent Statement: Not applicable: This study does not involve human participants. Therefore, informed consent was not required.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors are grateful to the reviewers for their valuable and constructive comments. The last listed author is thankful to Mahasarakham University for giving an opportunity to work on this project in December 2022. Observational data in the Thailand were provided by Thai Meteorological Department (TMD accessed on 10 January 2023) at <https://www.tmd.go.th/>.

Conflicts of Interest: The authors declare no potential conflict of interest.

References

- Promchote, P.; Wang, S.Y.S.; Johnson, P.G. The 2011 Great Flood in Thailand: Climate Diagnostics and Implications from Climate Change. *J. Clim.* **2016**, *29*, 367–379. [\[CrossRef\]](#)
- Singkran, N. Flood risk management in Thailand: Shifting from a passive to a progressive paradigm. *Int. J. Dis. Risk Red.* **2017**, *25*, 92–100. [\[CrossRef\]](#)
- Gale, E.L.; Saunders, M.A. The 2011 Thailand flood: Climate causes and return periods. *Weather* **2013**, *68*, 233–237. [\[CrossRef\]](#)
- Kunitiyawichai, K.; Schultz, B.; Uhlenbrook, S.; Suryadi, F.; Corzo, G. Comprehensive flood mitigation and management in the Chi River Basin, Thailand. *Lowland Technol. Int.* **2011**, *13*, 10–18.
- Arunyanart, N. Limsiri, C.; Uchaipichat, A. Flood hazards in the Chi River Basin, Thailand: Impact management of climate change. *Appl. Ecol. Environ. Res.* **2017**, *15*, 841–861. [\[CrossRef\]](#)
- Prahadchai, T.; Shin, Y.; Busababodhin, P.; Park, J.S. Analysis of maximum precipitation in Thailand using non-stationary extreme value models. *Atmos. Sci. Lett.* **2022**, *24*, e1145. [\[CrossRef\]](#)
- Department, R.I. Summarizes the Situation of the Chi-Mun Watershed Flooding in 2019. 2019. Available online: <http://water.rid.go.th> (accessed on 15 January 2023)
- Meteorological, T. Climatic Informations. 2021. Available online: <https://www.tmd.go.th> (accessed on 10 January 2023)
- Shin, Y.; Shin, Y.; Hong, J.; Kim, M.K.; Byun, Y.H.; Boo, K.O.; Chung, I.U.; Park, D.S.R.; Park, J.S. Future Projections and Uncertainty Assessment of Precipitation Extremes in the Korean Peninsula from the CMIP6 Ensemble with a Statistical Framework. *Atmosphere* **2021**, *12*, 97. [\[CrossRef\]](#)
- Lima, C.H.; Lall, U.; Troy, T.; Devineni, N. A hierarchical Bayesian GEV model for improving local and regional flood quantile estimates. *J. Hydrol.* **2016**, *541*, 816–823. [\[CrossRef\]](#)
- Tyralis, H.; Papacharalampous, G.; Tantane, S. How to explain and predict the shape parameter of the generalized extreme value distribution of streamflow extremes using a big dataset. *J. Hydrol.* **2019**, *574*, 628–645. [\[CrossRef\]](#)
- Jin, X.; Zhou, W.; Bie, R. Multinomial event naive Bayesian modeling for SAGE data classification. *Comput. Stat.* **2007**, *22*, 133–143. [\[CrossRef\]](#)
- De Luca, D.L.; Napolitano, F. A user-friendly software for modelling extreme values: EXTRASTAR (EXTRemes Abacus for STATistical Regionalization). *Environ. Model. Softw.* **2023**, *161*, 105622. [\[CrossRef\]](#)
- De Michele, C. Advances in deriving the exact distribution of maximum annual daily precipitation. *Water* **2019**, *11*, 2322. [\[CrossRef\]](#)
- Asadi, H.; Shahedi, K.; Jarihani, B.; Sidle, R.C. Rainfall-Runoff Modelling Using Hydrological Connectivity Index and Artificial Neural Network Approach. *Water* **2019**, *11*, 212. [\[CrossRef\]](#)
- Lillesand, T.; Kiefer, R.; Chipman, J. *Remote Sensing and Image Interpretation*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
- Guarnieri, R.A.; Pereira, E.; Chou, S. Solar radiation forecast using artificial neural networks in South Brazil. In Proceedings of the 8th International Conference on Southern Hemisphere Meteorology and Oceanography—8 ICSHMO, Foz do Iguaçu, Brazil, 24–28 April 2006; pp. 1777–1785.
- Ahmed, S. Performance of derivative free search ANN training algorithm with time series and classification problems. *Comp. Stat.* **2013**, *28*, 1881–1914. [\[CrossRef\]](#)
- Rotjanakusol, T.; Laosuwan, T. Drought Evaluation with NDVI-Based Standardized Vegetation Index in Lower Northeastern Region of Thailand. *Geogr. Tech.* **2019**, *14*, 118–130. [\[CrossRef\]](#)
- Land Processes Distributed Active Archive Center (LP DAAC). MOD13Q1 Version 6 Vegetation Indices. Available online: <https://lpdaac.usgs.gov/products/mod13q1v006/> (accessed on 10 March 2023).
- Jenkinson, A.F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Q. J. R. Meteorol. Soc.* **1955**, *81*, 158–171. [\[CrossRef\]](#)
- Coles, S. *An Introduction to Statistical Modeling of Extreme Values*; Springer: New York, NY, USA, 2001.
- Busababodhin, P. *Extreme Value Analysis*; Rajabhat MahaSarakham University: Maha Sarakham, Thailand, 2018.
- Papukdee, N.; Park, J.S.; Busababodhin, P. Penalized likelihood approach for the four-parameter kappa distribution. *J. Appl. Stat.* **2022**, *49*, 1559–1573. [\[CrossRef\]](#)
- Moré, J.J. The Levenberg-Marquardt algorithm: Implementation and theory. In *Numerical Analysis*; Watson, G.A., Ed.; Springer: Berlin/Heidelberg, Germany, 1978; pp. 105–116.
- Kim, T.W.; Valdés, J.B. Nonlinear Model for Drought Forecasting Based on a Conjunction of Wavelet Transforms and Neural Networks. *J. Hydrol. Eng.* **2003**, *8*, 319–328. [\[CrossRef\]](#)
- Fritsch, S.; Guenther, F.; Guenther, M.F. Package ‘neuralnet’. *Train. Neural Netw.* **2019**, *2*, 30.

28. Pearson, K. Notes on the History of Correlation. *Biometrika* **1920**, *13*, 25–45. [[CrossRef](#)]
29. Nash, J.; Sutcliffe, J. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.