



Modeling of Two-Meter Extreme Air Temperature in Iraq Using Extreme-Value Copulas

Manaf Hazim Ahmed  and Zaid Al-Khaledi 

Department of Statistics and Informatics, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq

Article information

Article history:

Received July 2, 2023

Accepted August 17, 2023

Available online December 1, 2023

Keywords:

Spatial Extremes

Extreme Value Copulas

2 Meter Air temperature

Pickands dependence function

Correspondence :

Manaf Hazim Ahmed

manaf.ahmed@uomosul.edu.iq

Abstract

The UN's International Organization for Migration (IOM) reported that Iraq is the fifth-most affected country by soaring temperatures. This requires assessing the risks associated by accurately understanding the behaviour of these extreme events. A deep investigation of the behaviour of 2m air temperature has been done, by spatially modeling this event using extreme value copula with Pickands dependence. The investigation of the 2m extreme air temperature in Iraq concluded that the symmetric extreme-value copula models are suitable to consider in the modeling. Nine extreme value copula models were constructed from one parameter family copulas (Hüsler-Reiss, Gumbel, and Galambos), and two parameters model t-EV, after adopting the spatial context. Fifty locations were randomly sampled from 1517 locations divided into two parts, 40 for modeling, and 10 for validation. The Composite Maximum Pseudo-Likelihood estimation method has been used in the modeling. According to the AIC information criterion, we selected 4 models as candidates (Hüsler-Reiss A and B; Gumbel, and Galambos). Due to the slight difference in the criterion values among the four candidate models, the Kullback-Leibler (KL) divergence method between the non-parametric and parametric pairwise extreme-value copulas has been evaluated by the validation dataset, to choose the best-fitted model. The Hüsler-Reiss A was the best-fitted model, due to the high KL density around zero of all the pairwise in the validation dataset.

DOI: <https://doi.org/10.33899/ijqjoss.2023.0181192> , ©Authors, 2023, College of Computer Science and Mathematics, University of Mosul Iraq.
This is an open access article under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

Extreme temperature events, such as heat waves, have significant impacts on various domains, such as agriculture, energy, and public health. Modeling these extreme events accurately is crucial for better understanding their behaviour and for effective planning and mitigation of the risks associated with them. Extreme value models have shown promising performance in modeling such types of data. In a spatial context, a max-stable spatial process is considered, and the models of this process will be in multivariate case [1]. This situation itself poses a challenge because these events follow multivariate extreme value (MGEV) distributions, and no existing models can capture the dependence structure of these events. Additionally, ignoring the dependencies among the locations and treating them as independent locations using the Generalized Extreme Value (GEV) distribution for each location will provide an unreal representation of the events. Despite the models of the max-stable in the bivariate case existing, the limited number of these models, and the relatively high number of parameters in most of the models also state as restrictions in the modeling.

For what is mentioned above, there is a need for a statistical tool that can combine the multivariate extreme-value theory with models more simple than classical ones, (Smith, Brown-Resnick, and Schlather), so that can be considered appropriate models for the dependence structure among the locations of the extreme event. Extreme value copula has gained a lot of attention in recent years for Modeling the dependence structure between extreme random variables. It is based on the extreme value theory, so these extreme copulas provide a functional link between multivariate distribution functions and their univariate margins [2, 3]. In spatial extremes, extreme value copulas play a crucial role. They enable the characterization of the dependence structure of the extreme event occurring at different locations. By considering the tail behaviour of these events, extreme value copulas can accurately capture the underlying dependence patterns. This, in turn, leads to improved modeling and analysis of spatial extremes. For examples, see [4], [5], [6], [7], and [8].

The report of the International Organization for Migration (IOM) in the UN concerning climate change published on 11 August 2022 puts Iraq as the fifth-most vulnerable country to climate breakdown, affected by soaring temperatures, and this requires preparing for assessing the risks associated with this climate change. Choosing a 2m air temperature to investigate its behaviour in this study was motivated by the outputs of this report. To address this breakdown, one should first understand the behaviour of the extreme 2m air temperature. This will be done by Modeling this event via the extreme-value copula. The 2m air temperature was collected from the fifth generation of the European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric, land and oceanic climate global dataset ERA5 [9]. This study is devoted to investigating the behaviour of the two-meter air temperature in Iraq through in-depth analysis by extreme-value copula with Pickands dependence functions. The modeling of this event has been done by following the statistical inference on extreme-value copulas introduced in [10] and adaptation of extreme-value copula to spatial context by considering the parameters are functions of distance among the locations. The Composite Maximum Pseudo-Likelihood estimation method introduced in [11] has been used in the modeling.

The paper is organized as follows: the theoretical concepts of extreme-value copula models, and corresponding Pickands dependence functions. Furthermore, adapting these concepts to the spatial context has been presented first. Then, the Composite pseudo-likelihood method is used in the parameters estimation of the copula models presented in Extreme-Value copula section. Preparing the 2m air temperature in Iraq dataset by pre-processing it (examining the stationary, isotropy, tail dependencies, and symmetry properties), modeling, and choosing the best-fitted model have been done in modeling the 2m air temperature in Iraq dataset. Finally, the discussions and conclusions of the main results obtained were presented.

Extreme-Value Copula

In this section, extreme-value copula models and their extension to spatial context used in modeling the 2m air temperature have been presented. The extreme-value copula will be defined via Pickands dependence function. Pickands function is a major and important key in extreme-value copula, so choosing different functions of Pickands leads to different copula models. Since the dataset that will be discussed in the pre-processing section is symmetric, a symmetric Pickands function, i.e., symmetric extreme-value, will be the focus of this section. The main fundamental concepts concerning Copula can be found in [12].

Let $X = (X_1, \dots, X_d)^T$ be a random vector with multivariate probability distribution function $H(x_1, \dots, x_d)$, and marginals $G_j, j = 1, \dots, d$. A function $C: [0,1]^d \rightarrow [0,1]$ is said to be a multivariate copula $C(u_1, \dots, u_d) = \mathbb{P}r(U_1 \leq u_1, \dots, U_d \leq u_d)$, with dimension d if and only if

$$C(u_1, \dots, u_d) = H(G_1^{-1}(u_1), \dots, G_d^{-1}(u_d)), \quad (1)$$

where $(u_1, \dots, u_d) \in [0,1]$. The copula C is unique if H is continuous [13]. In extreme value context, let $X^k = (X_1^k, \dots, X_d^k), k = 1, 2, \dots$ be a random vector with i.i.d replications and multivariate copula C_X , and let $B_n := (B_1, \dots, B_d)$ non-overlapping block maxima, such that $B_n = \max_{k=1, \dots, n} \{X^k\}$, and has max-stable copula

$$C_{B_n}(u_1, \dots, u_d) = C_X(u_1^{1/n}, \dots, u_d^{1/n})^n. \quad (2)$$

The extreme-value copula C exists if and only if

$$C_{B_n}(u_1, \dots, u_d) \rightarrow C(u_1, \dots, u_d), \quad \text{as } n \rightarrow \infty, \quad (3)$$

such that

$$C(u_1, \dots, u_d) = \exp\{\ell(-G_1^{-1}(u_1), \dots, -G_d^{-1}(u_d))\}, \quad (4)$$

where ℓ is a max-stable tail dependence measure [10, 14].

Bivariate extreme-value copula via Pickands dependence function

Since the multivariate extreme events have tail dependence, a Pickands dependence function $A: [0,1]^d \rightarrow [1/d, 1]$ is the most reasonable measure able to quantify the dependence strength among the variables [15]. It has the capability on analyzing rare events, such as extreme weather events. More specifically, Pickands dependence function is considered an essential tool in bivariate extreme-value copula, because it can reduce the mapping to one dimension, and Copula is fully characterized in this dimension [13]. Without loss of the generality, we will define $A(\cdot)$ when $d = 2$. In extreme context, and under max-stability tail dependence assumption, we can define Pickands dependence function, so that for all $u_j = G_j(X_j) \in [0,1], j = 1,2$

$$A(u_1, u_2) = \frac{\ell(-G_1^{-1}(u_1), -G_2^{-1}(u_2))}{G_1^{-1}(u_1) + G_2^{-1}(u_2)}. \quad (5)$$

Therefore, in the bivariate case, the extreme-value copula in Equation (4) can be expressed by Pickands dependence function as follows

$$C(u_1, u_2) = \exp\left(\left(G_1^{-1}(u_1) + G_2^{-1}(u_2)\right)A(u_1, u_2)\right) \quad (6)$$

By Theorem 2.22 in [10] formula of C in (6) will be

$$C(u_1, u_2) = \exp\left(\left(G_1^{-1}(u_1) + G_2^{-1}(u_2)\right)A\left[\frac{G_2^{-1}(u_2)}{G_1^{-1}(u_1) + G_2^{-1}(u_2)}\right]\right), \quad (7)$$

or equivalently

$$C(u_1, u_2) = \exp\left(\log(u_1 u_2)A\left[\frac{\log(u_2)}{\log(u_1 u_2)}\right]\right). \quad (8)$$

Referring that $G_j^{-1}(u_j)$ is the inverse of marginals corresponding to the $X_j, j = 1,2$. Since this study concerns to block maxima case, then G will follow GEV distribution with location μ , scale σ , and shape ξ parameters. $A(w)$ is a convex function with inequality $\max\{w, 1 - w\} \leq A(w) \leq 1$, respectively correspond to complete dependence and independence. Different models of C are determined by models of A . We will present the most common parametric and symmetric copulas, that will be used in the modeling in this paper.

1. **Hüsler-Reiss copula:** the spirit of this model is the standard Gaussian model with correlation function ρ , which can introduce via Pickands dependence function, so that

$$A(w) = (1 - w)\Phi(k(1 - w)) + w\Phi(k(w)), \quad (9)$$

where

$$k(w) = \frac{1}{\lambda} + \frac{\lambda}{2} \log \frac{w}{1-w}, \quad w = \frac{u_1}{u_1 + u_2}, \quad (10)$$

and Φ is the univariate standard Gaussian distribution. Let $\rho_n \rightarrow 1$ as $n \rightarrow \infty$, then $\lambda = \sqrt{(1 - \rho_n) \log(n)} \in [0, \infty)$ as $n \rightarrow \infty$. In this model, λ will be the parameter of such model, under the assumption that the correlation between the pairwise (U_1, U_2) increase as the size of sample increase also [16].

2. **Gumbel copula:** this model is one of the Archimedean copulas able to evaluate the dependence straight in asymptotic limits of maxima (upper tail) of the pairwise (U_1, U_2) . The Pickands dependence function is

$$A(w) = [w^\theta + (1 - w)^\theta]^{1/\theta}, \quad \theta \in [1, \infty). \quad (11)$$

The attraction of this model is the domain of the Logistic distribution function [17].

3. **Galambos copula:** this model is the negative of Gumbel copula (derived from negative logistic distribution), e.g., if C_G is Gumbel model, and \bar{C}_G is the corresponding survival copula, then we can consider $\bar{C}_G(u_1, u_2)$ is a distribution function of the pairwise $(1 - U_1, 1 - U_2)$, so that the corresponding Pickands dependence function is

$$A(w) = 1 - [w^{-\delta} + (1 - w)^{-\delta}]^{-1/\delta}, \quad \delta \in [0, \infty) \quad (12)$$

4. **t-EV copula:** Its so-called t-Extreme Value. This model is derived from Student's distribution of the pairwise of (U_1, U_2) , with two parameters correlation coefficient $\rho \in [-1, 1]$, and degree of freedom $\nu > 0$. The Pickands dependence function of this model can write as

$$A(w) = (1 - w)\mathcal{T}_{\nu+1}(k(1 - w)) + w\mathcal{T}_{\nu+1}(k(w)), \quad (13)$$

where

$$k(w) = \sqrt{\frac{1 + \nu}{1 - \rho^2}} \left[\left(\frac{w}{1 - w} \right)^\nu - \rho \right], \quad (14)$$

and $\mathcal{T}_{\nu+1}$ is a univariate student's distribution function with a degree of freedom ν . If $\nu \rightarrow \infty$, then t-EV weakly converges to Hüsler-Reiss copula with parameter $\lambda = \sqrt{\frac{2}{\nu(1-\rho)}}$ [18].

Adaptation of extreme-value copula to spatial context

In the previous section, the extreme-value copula and corresponding Pickands dependence functions are presented in general concepts. In this part adaptation of these concepts to spatial extremes will be done. Let $\{Y^k(s)\}_{s \in \mathcal{S}}$, $\mathcal{S} \subset \mathbb{R}^d$, $d \geq 1$, and $k = 1, 2, \dots$ be an i.i.d replications of spatial process. Let $a_n(s) > 0$ and $b_n(s)$, $a, b \in \mathbb{R}^d$ be two continuous sequences, if there exist

$$\left\{ \max_{k=1, \dots, n} \frac{Y^k(s) - b_n(s)}{a_n(s)} \right\}_{s \in \mathcal{S}} \rightarrow \{X(s)\}_{s \in \mathcal{S}}, \quad \text{as } n \rightarrow \infty, \tag{15}$$

with non-degenerate marginals, then $X := \{X(s)\}_{s \in \mathcal{S}}$ is a max-stable process, such that $X \sim G(\mu(s), \sigma(s), \xi(s))$, where G refer to GEV distribution with location μ , scale σ , and shape ξ parameters [19]. The max-stable spatial process $X(s)$ is said to be strictly stationary, if $\forall s \in \mathcal{S}, \mathcal{S} \subset \mathbb{R}^d, \mu(s_1) = \dots = \mu(s_d) = \mu, \sigma(s_1) = \dots = \sigma(s_d) = \sigma$, and $\xi(s_1) = \dots = \xi(s_d) = \xi$. And $X(s)$ is isotropic if the covariance for each $(s, t) \in \mathcal{S}, \mathcal{S} \subset \mathbb{R}^2$ depends only on the distance, such that $cov(s, t) = \rho(h), h = \|s - t\|$. In what follow in this paper, the max-stable spatial process $X(s)$ will be under assumptions of stationarity and isotropy properties.

When the focus is on extreme values, it is necessary to use more suitable tools for analyzing the spatial dependence of extremes. Since our aim is modeling using the Extreme-Value Copula concept via Pickands dependence function, we present the concepts in a spatial context. Let $(U(s), U(s + h))$ be a pairwise of spatial process with unit uniform distribution separated by the distance h , such that for all $s \in \mathcal{S}, U(s) = G_{\mu(s), \sigma(s), \xi(s)}(X(s))$. The bivariate extreme-value copula corresponding to the pairwise is

$$\begin{aligned} C_h(u_1, u_2) &= \Pr(U(s) \leq u_1, U(s + h) \leq u_2) \quad u_1, u_2 \in [0, 1] \\ &= \exp\left(\log(u_1 u_2) A_h \left[\frac{\log(u_2)}{\log(u_1 u_2)} \right]\right). \end{aligned} \tag{16}$$

The Pickands dependence function $A_h(\cdot)$ is a function that evaluates the dependence strength between $(U(s), U(s + h))$ separated by distance h . Concerning Husler-Reiss Pickands function, the spatial aspect will be included, so that, the parameter λ in Equation (10), will be $\lambda_h = \sqrt{(1 - \rho(h)) \log n}$, where $\rho(h)$ is the spatial isotropic correlation function. Many models of correlation function exist, such as exponential, power exponential, and many others. As well as for ρ in t-EV copula model [20]. Concerning Gumbel and Galambos copula models, respectively with parameters θ and δ , the same consideration will be made. The fact that, the dependence strength of each pairwise in $(U(s), U(s + h))$ are varying, and since $U(s)$ is isotropic, which means this varying will be according to the distance. And most of the time this dependence strength decreases as h increases. Therefore, using this fact, we will consider the parameters θ and δ to be the trend across distance. Such that, $\theta_h = \beta h$, as well as for $\delta_h = \beta h$, where β is a coefficient of trending.

Composite Maximum Pseudo-Likelihood

A parametric estimation such as the Maximum Pseudo-Likelihood MPL method showed as a useful tool for estimating copula parameters, especially when the marginals are unknown [21]. Since just the bivariate extreme copula models exist, the composite likelihood is a reasonable method for estimating spatial extreme models [22, 23]. The combination of the two likelihood methods composites and pseudo was defined in [11], named Composite Pseudo-Likelihood CPL. This method is very suitable when using the copula concept in modeling spatial extremes. For that, this method was used in the Modeling of the extreme 2m air temperature event.

Given a max-stable dataset $\{X^k(s)\}_{s \in \mathcal{S}}$ with i.i.d $k = 1, \dots, n$ replicates, and let $\{\hat{U}^k(s)\}_{s \in \mathcal{S}} = G_{\hat{\mu}(s), \hat{\sigma}(s), \hat{\xi}(s)}(X^k(s))$ is pseudo max-stable spatial process. The Composite pseudo-likelihood function is given by

$$\mathcal{L}(\psi) = \sum_{k=1}^n \sum_{i < j}^d \log c(\hat{U}^k(s_i), (\hat{U}^k(s_j); \psi), \tag{17}$$

where c denoted to the likelihood contribution function of the pairwise (u_i, u_j) at the replication k . In this study c used as bivariate density of the corresponding extreme-value copula defined in Equation (16). Let the compact set of the parameters ψ of c is denoted by θ . The estimation of ψ can be achieved by maximizing \mathcal{L} , so that

$$\hat{\psi} = \max_{\psi \in \theta} \mathcal{L}(\psi). \tag{18}$$

Since the i.i.d achieved on copula when the marginals are known, such as in this case, the pairwise pseudo-likelihood estimator $\hat{\psi}$ has asymptotic normality as $n \rightarrow \infty$, with mean ψ and covariance matrix of sandwich form $J^{-1}(\psi)K(\psi)J^{-1}(\psi)$, where

$$K(\psi) = \mathbb{E} \left[\frac{\partial \mathcal{L}(\psi)}{\partial \psi} \frac{\partial \mathcal{L}(\psi)}{\partial \psi^T} \right], \quad \text{and} \quad J(\psi) = -\mathbb{E} \left[\frac{\partial^2 \mathcal{L}(\psi)}{\partial \psi \partial \psi^T} \right]$$

respectively are the variance of the score function, and the expected information matrix are computed from Equation (17). For more details about asymptotic behaviour, see [11], [24], and [25]. The estimation \hat{J} of $J(\psi)$ can be readily obtained from the Hessian provided by the optimization algorithm, and \hat{K} of $K(\psi)$ by the empirical variance of the score contribution of each observation [26]. The selected model will be according to the corresponding minimum of $CLIC^* = (D - 1)^{-1}CLIC$, where D is the number of locations in the dataset, and

$$CLIC = -2\{\mathcal{L}(\hat{\psi}) - tr(K(\hat{\psi})J(\hat{\psi})^{-1})\}, \tag{19}$$

is the Composite Likelihood Information Criterion. $CLIC^*$ is very closely to Akaike Information Criterion AIC, so for simplicity in computations, we shall use AIC instead of $CLIC^*$. See [23] and [26].

Modeling the 2m air temperature in Iraq dataset
Data description and pre-processing

The goal of this section is to model the extreme 2m air temperature in Iraq. The hourly 2m air temperature was collected from the fifth generation of atmospheric land and oceanic climate global dataset ERA5, produced by the European Centre for Medium-Range Weather Forecasts ECMWF. This dataset was collected for the region with a longitude range of 37.5 to 49 degrees, a latitude range of 27.5 to 38 degrees, and a grid spacing of 11 km during the summer season (June, July, and August) for the years from 1981 to 2022, at times from 11:00H to 17:00H. This collection of data resulted $41 \times 37 = 1517$ grids and 90528 hourly observations for each grid. Mathematically, let $\{Y^k(s)\}_{s \in \mathcal{S}}$, $S \subset \mathbb{R}^2$, $|\mathcal{S}| = 1517$, and $k = 1, \dots, 90528$ be a spatial process represent the 2m air temperature. To ensure the block maxima be i.i.d, a monthly block maxima $B_{7 \times 30}$ was proposed. So that for a non-overlapping replication

$$\{X(s)\}_{s \in \mathcal{S}} = \max_{k \in B_{7 \times 30}} \{Y^k(s)\},$$

where $\{X(s)\}_{s \in \mathcal{S}}$ is a spatial extreme process, for each marginal of s follows GEV distribution, and the amount 7×30 represents respectively the number of hours per day times the number of days per month.

To examine the stationarity of the dataset, the GEV's parameters $\mu(s), \sigma(s)$, and $\xi(s)$ are estimated for each location $s \in \mathcal{S}$ of $X(s)$ using the maximum likelihood estimation method. The grids in the three panels in Figure 1 represent the estimated location $\hat{\mu}(s)$, scale $\hat{\sigma}(s)$, shape $\hat{\xi}(s)$ parameters for each $s \in \mathcal{S}$, $|\mathcal{S}| = 1517$. Noting that, all the computational process in this study has been done by R program with main package 'copula' version 1.1-2, and others.

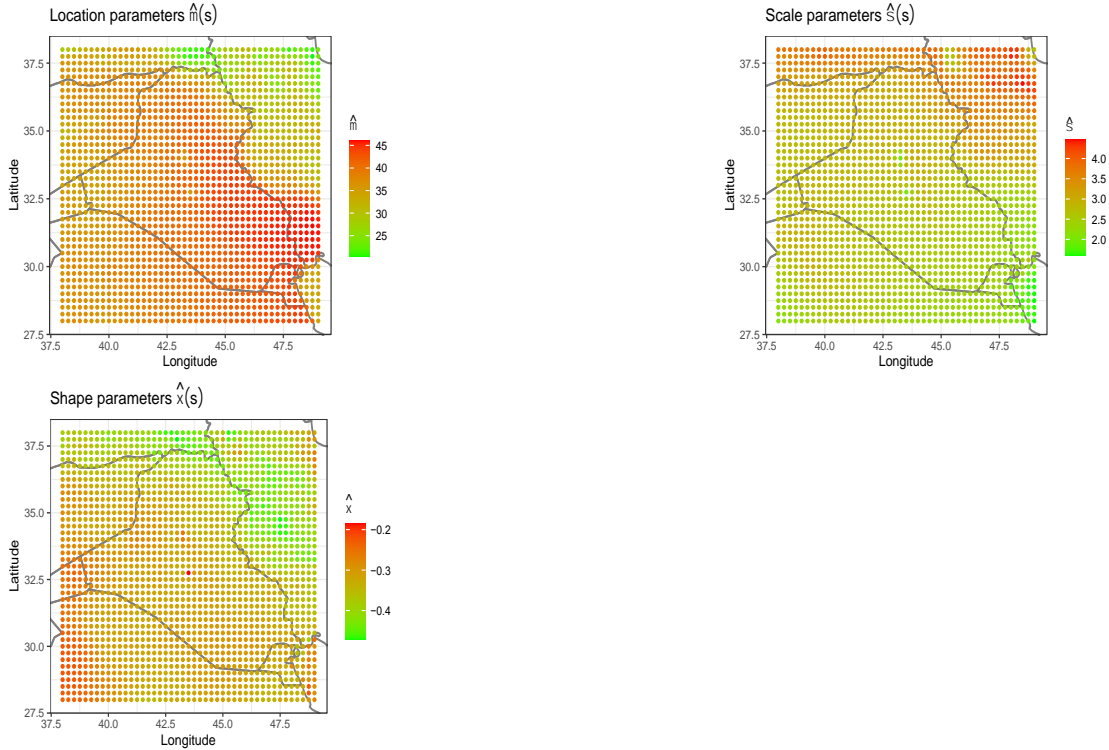


Figure 1: The panels respectively represent the estimated parameters $\hat{\mu}(s)$, scale $\hat{\sigma}(s)$, shape $\hat{\xi}(s)$ for each grid of the dataset $s \in \mathcal{S}$, $|\mathcal{S}| = 1517$

It is clear that each of the three estimated parameters for each grid is approximately equal, especially in the red region for $\hat{\mu}$, green in $\hat{\sigma}$; and red for $\hat{\xi}$ excluding the northeast of Iraq, due to the mountains. That means and from the definition of strictly spatial stationarity, we can consider the dataset has spatial stationary property. In the following step the spatial extreme dataset $\{X(s)\}_{s \in \mathcal{S}}$ will transform into $[0,1]$, such that for each $s \in \mathcal{S}$

$$\{\hat{U}(s)\}_{s \in \mathcal{S}} = G_{\hat{\mu}(s), \hat{\sigma}(s), \hat{\xi}(s)}(X(s)),$$

where G is the GEV distribution. In what follows, we shall deal with $\hat{U}(s)$ instead of $X(s)$. It is known the dependence structure pattern of the events is essential in modeling extreme events. This structure distinguishes between the models corresponding to asymptotic dependence/ independence structures. So, in the next step, examine the dataset for which the dependence structure belongs to asymptotic dependence or independence. This examination will be by empirical upper and lower tails dependence measures [27],

$$\tilde{\chi}(s, t) = \lim_{u \rightarrow 1^-} \frac{1 - 2u + \tilde{C}_{s,t}(u, u)}{1 - u}, \quad (20)$$

And

$$\tilde{\chi}(s, t) = \lim_{u \rightarrow 0^+} \frac{\tilde{C}_{s,t}(u, u)}{u}, \quad (21)$$

where $\tilde{C}_{s,t}$ is the empirical Copula, so that $\tilde{C}_{s,t}(u, u) = \mathbb{P}r(\hat{U}(s) \leq u, \hat{U}(t) \leq u)$, and $(s, t) \in \mathcal{S}$. For more details, see [10]. The threshold was chosen to be $u = 0.975$ to ensure there are data for computation. The pairwise evaluation of $\tilde{\chi}(s, t)$ and $\tilde{\chi}(s, t)$ of the dataset represented respectively by the first and second panels in Figure 2. From these panels, the dataset seems to have an upper tail dependence structure, due to $\tilde{\chi}(s, t) \in (0,1]$, as well as for $\tilde{\chi}(s, t)$. So, we can consider the block maxima $\hat{U}(s)$ has an asymptotic dependence structure, and this leads to consider extreme copula models. To verify if the dependence

structure of the dataset is present in extreme, furthermore, the exchangeability (have symmetric distribution) between the pairwise locations (s, t) , a test hypothesis corresponding to these two assumptions has been made, (see Figure 3). Once again, the northeastern region of Iraq does not appear a tail dependence structure, and this is clear on the top and right sides of the two panels. From this result, this region will be excluded definitely from the modeling.

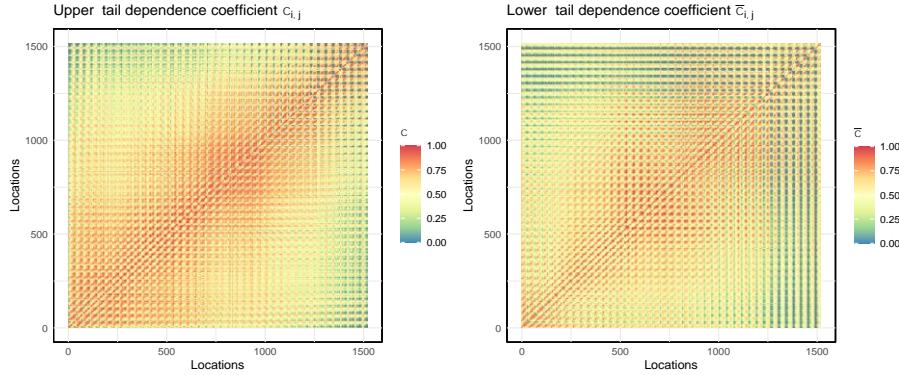


Figure 2: Empirical pairwise upper $\hat{\chi}(s, t)$ and lower $\hat{\bar{\chi}}(s, t)$ tail dependence measures, respectively represented by the left and right panels. Each grid in the panels represents the corresponding tail dependence strength between the pairwise $(\hat{U}(s), \hat{U}(t))$, $(s, t) \in \mathcal{S}$.

Regarding the pairwise extreme-value dependency, 100 locations sampled form $\mathcal{S} = 1517$ for this purpose. The p -value test with $\alpha = 0.05$ will be done between the empirical pairwise copula $\hat{C}(U(s), U(t))$ and extreme-value copula with the non-parametric estimate of Pickands dependence function $\hat{A}(\cdot)$ under the hypothesis

$$\mathcal{H}_0: C \in \mathcal{C}^* \quad \text{vs} \quad \mathcal{H}_1: C \notin \mathcal{C}^*,$$

where $\mathcal{C}^* := C(u, v) = (uv)^{A(w)}$, $A(w)$ is a Pickands dependence function of $w = \log(v)/\log(uv)$ [28]. The p -value statistics is illustrated in the left panel in Figure 3. The blue fill represents that the p -value test cannot reject \mathcal{H}_0 . In other words, the pairwise $(\hat{U}(s), \hat{U}(t))$, $(s, t) \in \mathcal{S}$ has extreme-value copula. As illustrated in the left panel, the test in most of the pairwise failed to reject \mathcal{H}_0 , so we can consider extreme-value models in modeling the dataset. To ensure this property exists in modeling, the pairwise locations rejected \mathcal{H}_0 will be excluded from the selection of locations for modeling.

To test the exchangeability between the pairwise $(\hat{U}(s), \hat{U}(t))$ (symmetry radial of the underlying multivariate copula), p -value statistics will be used based on empirical copula [29]. The assumption of the symmetry copula will be under the hypothesis

$$\mathcal{H}_0: C = \bar{C} \quad \text{vs} \quad \mathcal{H}_1: C \neq \bar{C},$$

where \bar{C} is the survival of C . The right panel in Figure 3 appears that the majority cannot reject \mathcal{H}_0 . Then we can consider the extreme-value and symmetric models.

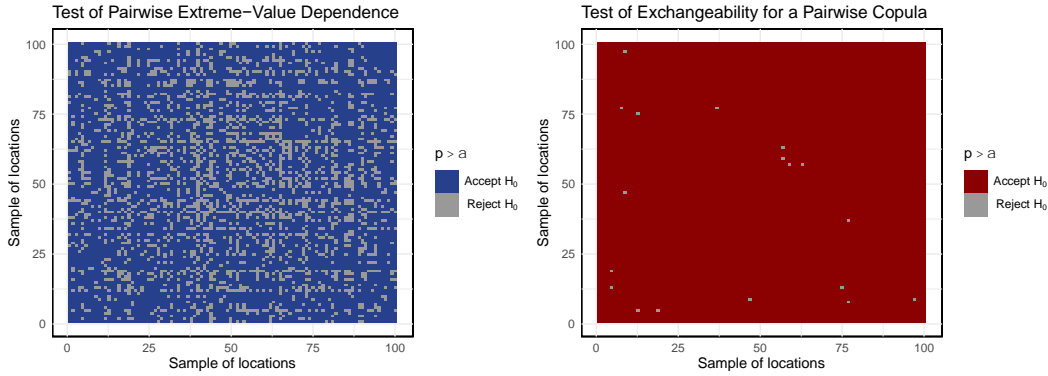


Figure 3: The panels represent the p -value test with $\alpha = 0.05$ for extreme value dependence and the exchangeability (symmetry) of the copula dataset. The left panel, represent the pairwise $(\hat{U}(s), \hat{U}(t))$, $(s, t) \in \mathcal{S}$ has extreme-value dependence after collecting randomly 100 locations among $\mathcal{S} = 1517$ location; while the right panel represents the which pairwise have symmetric copula model.

Copula models proposed for modeling 2m air temperature dataset

By adopting the results obtained in the pre-processing section, the locations implemented in the modeling will be randomly selected from $\mathcal{S} = 1517$ locations to ensure consistency with the outcomes of the previous section: The northeast region of Iraq will be excluded from the Modeling, due to the spatial non-stationary with the remaining region. In other words, the GEV marginals of this region have different behaviour from others; the process $\hat{U}(s)$ at location s which does not have extreme-value dependences; and also, does not have symmetric copula will be excluded from the modeling also. The 2m air temperature spatial process $\hat{U}(s)$ in 50 locations have been randomly selected, 40 for modeling and 10 for validation. The latter locations will not be used in modeling. Most of these locations are in Iraq and small numbers are sited around the border of the west and south of Iraq, which have the same behavior as the majority of locations. The coordinates of the selected locations are pointed out in Figure 4.

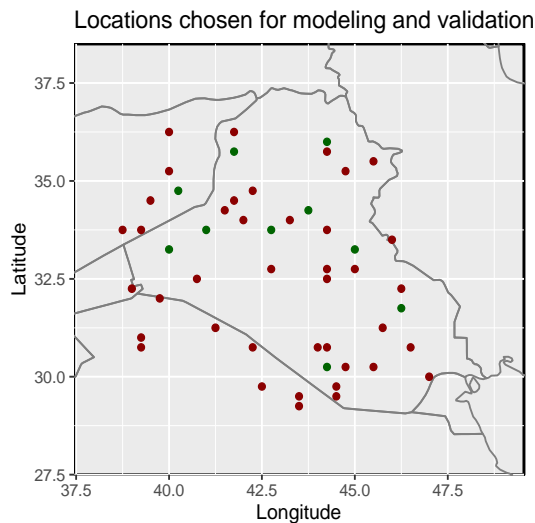


Figure 4: The coordinates of the selected locations of modeling and validation. The red dots represent the locations that will use in the modeling, while the green one will be for validation

According to the results of the preprocessing, a symmetric extreme-value Copula model is proposed with different Pickands dependence functions. One parameter extreme copula family (Hüsler-Reiss, Gumbel, and Galambos), and two parameters family, such as t-EV defined respectively in Equations (10), (12), (13) and (14) are chosen for modeling. Traditionally, in one-parameter models, the tail dependence strength between two extreme random variables is controlled by this parameter. To extend this concept to spatial context, should this parameter be varying across the distance $h = \|s - t\|$ between the pairwise locations

$(s, t) \in \mathcal{S}$, so that these parameters measure the tail dependence strength between $(X(s), X(s + h))$ or equivalently $(\tilde{U}(s), \tilde{U}(s + h))$ separated by the distance h , under isotropy assumption. Due to the dependence strength varying with each pairwise separated by a distance h , the parameters will consider as a function of distance. So, models A will consider the parameter varying across the distance with coefficient β ; model B, the parameter varying across h according to exponential dependence strength; models C will be according to power exponential; and model D with Cauchy. So that the proposed models will be

1. Hüsler-Reiss A with parameter, $\lambda = \beta h, \beta > 0$.
2. Hüsler-Reiss B with parameter, $\lambda^2 = (1 - \rho(h)) \log n$, where $\rho(h) = \exp(-h/\sigma)$, and, $\sigma > 0$.
3. Hüsler-Reiss C with parameter, $\lambda^2 = (1 - \rho(h)) \log n$, where $\rho(h) = \exp(-(h/\sigma)^\delta)$, and $\sigma, \delta > 0$.
4. Hüsler-Reiss D with parameter, $\lambda^2 = (1 - \rho(h)) \log n$, where, $\rho(h) = 1 - (1 + (h/\sigma)^2)^\delta$, and $\sigma, \delta > 0$.
5. Gumbel with parameter, $\theta = \beta h, \beta > 0$.
6. Galambos with parameter, $\theta = \beta h, \beta > 0$.
7. t-EV A with $\rho(h) = \exp(-h/\sigma)$, and, $\sigma > 0$.
8. t-EV B with $\rho(h) = \exp(-(h/\sigma)^\delta)$, and $\sigma, \delta > 0$.
9. t-EV C with $\rho(h) = 1 - (1 + (h/\sigma)^2)^\delta$, and $\sigma, \delta > 0$.

Due to no extreme-value models with $d > 2$ exist in spatial context, all models considered with $d = 2$, and therefore the estimation of the parameter's models will be by Composite Pseudo Likelihood CPL method previously introduced. Usually, Composite likelihood method is used for estimating parameters for the spatial extreme process. Table 1 shows the estimated parameters, log-likelihood amount, and model selecting criterion (Akaike information criterion AIC) corresponding to each model proposed.

Table 1: Show the estimated parameters, log-likelihood, and Akaike information criterion AIC for the models proposed

Copula model	Estimated parameters				log-likelihood	AIC
	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\delta}$	$\hat{\nu}$	$\log \ell(\hat{\theta})$	
Hüsler-Reiss A	1.406670				15.89644	-3.53219
Hüsler-Reiss B		0.839202			15.79983	-3.519998
Hüsler-Reiss C		71.5492	0.120000		16.00407	-1.545686
Hüsler-Reiss D		0.01508	0.105000		15.96269	-1.540508
Gumbel	1.888350				15.78417	-3.518015
Galambos	2.350493				15.74095	-3.512531
t-EV A		0.38663		2.838140	15.68348	-1.505216
t-EV B		0.15070	0.336934	3.219790	15.83652	0.4753627
t-EV C		0.10200	1.800157	1.083037	15.47631	0.5213791

The AIC information criterion in Table 1 indicates that the best four fitted models are Hüsler-Reiss A, B, Gumbel, and Galambos. The slight difference in the criterion values of AIC among the four candidate models made us consider the model which will represent the 2m air temperature has minimum divergence

between the non-parametric and parametric of the estimated bivariate extreme-value copulas. This divergence test will apply to the validation dataset using the Kuiback-Lebler KL method. For more information see, [10]. Each pairwise of the validation dataset, the KL divergence between the four estimated models and the model with non-parametric Pickands function $\hat{A}(\cdot)$ have been done. Figure 5 shows that the Hüsler-Reiss A model has a divergence density of the pairwise of the validation dataset with more skew to lift (to zero) than the three candidate models. In other words, the divergence between the non-parametric and estimated Hüsler-Reiss A copula is the minimum. For that, now we can consider the Hüsler-Reiss A copula as the best extreme-value copula model to represent the 2m air temperature event in Iraq, and they can adapt it in further studies.

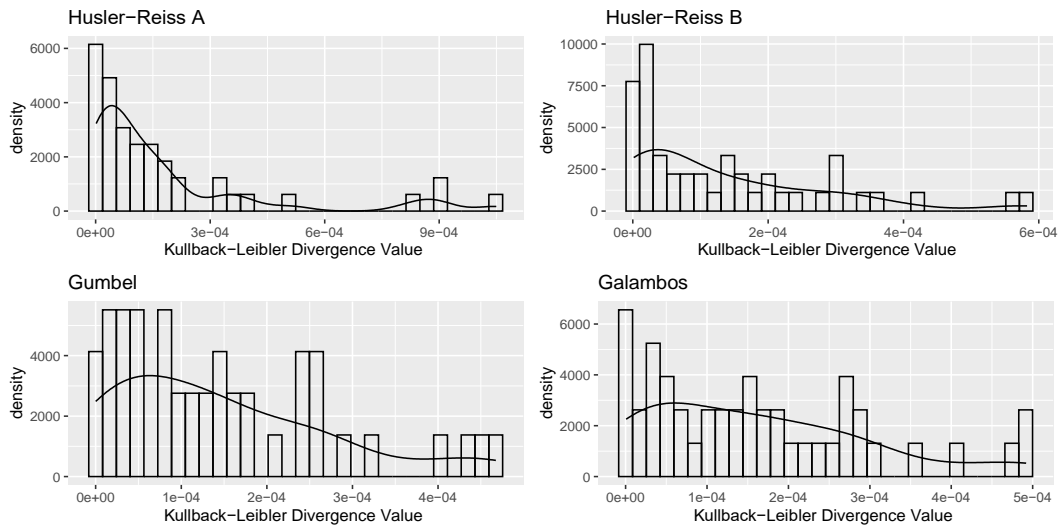


Figure 5: The densities of the Kuiback-Lebler divergence of the four candidate models (Hüsler-Reiss A, Hüsler-Reiss B, Gumbel, and Galambos) with the non-parametric extreme-value copula, which implemented on the validation dataset

Discussion and Conclusion

Iraq’s 2m extreme air temperature dataset $\{x(s)\}_{s \in \mathcal{S}}$ has been investigated to analyze its behaviour. The monthly block maxima have been adopted in this study, so that resulted from this block maxima procedure marginals having GEV distribution with location $\mu(s)$, $\sigma(s)$, and $\xi(s)$ parameters. The parameters of GEV for each marginal (location) $s \in \mathcal{S}$, $|\mathcal{S}| = 1571$ have been estimated by the maximum likelihood method. The estimated parameters $\hat{\mu}(s)$, $\hat{\sigma}(s)$, and $\hat{\xi}(s)$ appeared slightly varying in their amounts for most of the locations, except the northeast region. So, we can consider the events as stationary after excluding the northeast of Iraq region, due to the mountains which caused the spatial process $X(s)$ has non-stationary behaviour in this region. These estimated marginals are also used to transform $X(s)$ to a pseudo format, $\hat{U}(s) = G_{\hat{\mu}(s), \hat{\sigma}(s), \hat{\xi}(s)}(X(s))$, that will be used in the modeling. Examining the pseudo dataset has asymptotic dependence /independence behaviour has been done by the upper tail $\tilde{\chi}(s, t)$, and lower tail $\tilde{\chi}(s, t)$ measures using empirical copula. The dataset showed that has asymptotic dependence property. After examining the pseudo spatial dataset that has heavy tail dependence property, the next test is to see whether the dataset belongs to an extreme-value copula and it is symmetric (exchangeable). For the first one, p -value with $\alpha = 0.05$ test was used between the empirical copula and extreme-value copula with non-parametric estimated Pickands dependence function. This test was under hypotheses $\mathcal{H}_0: C \in \mathcal{C}^*$ vs $\mathcal{H}_1: C \notin \mathcal{C}^*$, where $\mathcal{C}^* := \mathcal{C}(u, v) = (uv)^{A(w)}$, $A(w)$ is a Pickands dependence function of $w = \log(v)/\log(uv)$. The statistics p -value falls to reject \mathcal{H}_0 . The symmetry (symmetry radial of the underlying multivariate copula) test is also done by p -value under the hypothesis $\mathcal{H}_0: C = \bar{C}$ vs $\mathcal{H}_1: C \neq \bar{C}$, where \bar{C} is the survival of C . The statistics p -value also falls to reject \mathcal{H}_0 .

The investigation of the 2m extreme air temperature in Iraq above, concluded that symmetric extreme-value copula models are suitable to consider in modeling. Nine Copula models were constructed from one

parameter family copulas (Hüsler-Reiss, Gumbel, and Galambos), and two parameters model, which is the t-EV copula after adopting the spatial context. Such that, and regarding the Hüsler-Reiss model, four models were considered. The first model A, when λ trend with h , such that $\lambda_h = \beta h$; the second B, third C, and fourth D, we used the spirit of its parameter $\lambda_h = \sqrt{(1 - \rho(h)) \log(n)}$, where $\rho(h)$ is a spatial and isotropic auto-correlation function, and $h = \|s - t\|$ is a distance between the pairwise locations $(s, t) \in \mathcal{S}$. The auto-correlation functions ρ have been chosen to be exponential, $\rho(h) = \exp(-h/\sigma)$, $\sigma > 0$; power exponential $\rho(h) = \exp(-(h/\sigma)^\delta)$, $\sigma, \delta > 0$, and Cauchy $\rho(h) = 1 - (1 + (h/\sigma)^2)^\delta$, $\sigma, \delta > 0$. These models of ρ used also in t-EV copula to construct three additional models, named t-EV A, B, and C. Finally, for Gumbel and Galambos models, the same consideration in Hüsler-Reiss A model has been used, such that their parameters respectively expressed by $\theta_h = \beta h$, and $\delta_h = \beta h$.

A sample of 50 locations was randomly selected from $|\mathcal{S}| = 1517$ locations divided into two parts, 40 for Modeling, and 10 for validation, after excluding the northeast region in Iraq from the sampling. The composite pseudo-likelihood estimation method is used to model the extreme event. According to the AIC information criterion, we selected 4 models as candidates. Which are: Husler-Reiss A and B; Gumbel, and Galambos. Due to the slight difference in the criterion values among the four candidate models, we implemented the KL divergence method between the four estimated models with the model having a non-parametric estimated Pickands dependence function. This divergence test was implemented on the validation dataset to choose the best-fitted model. The Hüsler-Reiss A model won in this competition, due to the density around zero of all the pairwise in the validation dataset being high.

Acknowledgement

The authors are very grateful to the University of Mosul, College of Computer Science and Mathematics for their provided facilities, which helped improve this work's quality.

References

- [1] Coles, S., Tawn, J. A., "Modelling extreme multivariate events," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 53, no. 2, pp. 377--392, 1991.
- [2] Sibuya, M., "Bivariate extreme statistics," *Annals of the Institute of Statistical Mathematics*, vol. 11, no. 2, pp. 195--210, 1960.
- [3] Gudendorf, G., Segers, J., "Copulas, Extreme-Value," in *Copula Theory and Its Applications*, Berlin, Heidelberg, Springer Berlin Heidelberg, 2010, pp. 127--145.
- [4] Palacios-Rodriguez, F., Bernardino, E. D., Mailhot, M., "smooth copula-based generalized extreme value model and spatial interpolation for extreme rainfall in Central Eastern Canada," *Environmetrics*, vol. 34, no. 3, p. e2795, 2023.
- [5] Pabaghi, Z., Bazrafshan, O., Zamani, H., Shekari, M., Singh, V.P., "Bivariate Analysis of Extreme Precipitation Using Copula Functions in Arid and Semi-Arid Regions," *Atmosphere*, vol. 14, no. 2, p. 275, 2023.
- [6] Krupskii, P., Hüser, R., Genton, M. G., "Factor copula models for replicated spatial data," *Journal of the American Statistical Association*, vol. 113, no. 521, pp. 467--47, 2018.
- [7] Rinaldi, A., Djuraidah, A., Wigena, A., Mangku, I. W., Gunawan, D., "Spatial extreme models with copula to determine extreme rainfall zone," *Applied Mathematical Sciences*, vol. 11, no. 27, pp. 1327--1336, 2017.
- [8] Cooley, D., Cisewski, J., Erhardt, R., Jeon, S., Mannshardt, E., Omolo, B. O., Sun, Y., "A survey of spatial extremes: Measuring spatial dependence and modeling spatial effects," *REVSTAT-Statistical Journal*, vol. 10, no. 1, pp. 135--165, 2012.
- [9] Ahmed, M., Maume-Deschamps, V., Ribereau, P., "Recognizing a spatial extreme dependence structure: A deep learning approach," *Environmetrics*, vol. 33, no. 4, p. e2714, 2022.
- [10] Eschenburg, P., "Properties of extreme-value copulas," *Technische Universität München, Fakultät für Mathematik*, 2013.
- [11] Górecki, J., Hofert, M., "Composite pseudo-likelihood estimation for pair-tractable copulas such as Archimedean, Archimax and related hierarchical extensions," *Journal of Statistical Computation and Simulation*, pp. 1-35, 2023.
- [12] Nelsen, R. B., *An introduction to copulas*, New York, NY: Springer New York, NY, 2006.
- [13] Sabbagh, M., *Copula models and pickands dependence functions*, Canada: McGill University, 2019.
- [14] Abu-Awwad, A., Maume-Deschamps, V., Ribereau, P., "Semiparametric estimation for space-time max-stable processes: an F-madogram-based approach," *Statistical Inference for Stochastic Processes*, vol. 24, pp. 241--276, 2021.
- [15] Pickands, J., "Multivariate extreme value distribution," *Proceedings 43th, Session of International Statistical Institution*, 1981.
- [16] Hüsler, J., Reiss, R.-D., "Maxima of normal random vectors: between independence and complete dependence," *Statistics & Probability Letters*, vol. 7, no. 4, pp. 283--286, 1989.
- [17] Gumbel, E. J., "Bivariate logistic distributions," *Journal of the American Statistical Association*, vol. 56, no. 249, pp. 335--349,

- 1961.
- [18] Nikoloulopoulos, A. K., Joe, H., Li, H., "Extreme value properties of multivariate t copulas," *Extremes*, vol. 12, pp. 129--148, 2009.
- [19] De Haan, L., Pereira, T. T., "Spatial extremes: Models for the stationary case," *The Annals of Statistics*, pp. 146--168, 2006.
- [20] Ahmed, M., Maume-Deschamps, V., Ribereau, P., Vial, C., "Spatial risk measures for max-stable and max-mixture processes," *Stochastics*, vol. 92, no. 7, pp. 1005--1020, 2020.
- [21] Genest, C., Ghoudi, K., Rivest, L.-P., "A semiparametric estimation procedure of dependence parameters in multivariate families of distributions," *Biometrika*, vol. 82, no. 3, pp. 543--552, 1995.
- [22] Bacro, J.-N., Gaetan, C., Toulemonde, G., "A flexible dependence model for spatial extremes," *Journal of Statistical Planning and Inference*, vol. 172, pp. 36--52, 2016.
- [23] Thibaud, E., Mutzner, R., Davison, A.C., "Threshold modeling of extreme spatial rainfall," *Water resources research*, vol. 49, no. 8, pp. 4633--4644, 2013.
- [24] Zhao, Y., Joe, H., "Composite likelihood estimation in multivariate data analysis," *Canadian Journal of Statistics*, vol. 33, no. 3, pp. 335--356, 2005.
- [25] Kent, J. T., "Robust properties of likelihood ratio tests," *Biometrika*, vol. 69, no. 1, pp. 19--27, 1982.
- [26] Varin, C., Reid, N., Firth, D., "An overview of composite likelihood methods," *Statistica Sinica*, pp. 5--42, 2011.
- [27] Coles, S., Heffernan, J., Tawn, J.A., "Dependence measures for extreme value analyses," *Extremes*, vol. 2, no. 4, pp. 339--365, 1999.
- [28] Kojadinovic, I., Yan, J., "Nonparametric rank-based tests of bivariate extreme-value dependence," *Journal of Multivariate Analysis*, vol. 101, no. 9, pp. 2234-2249, 2010.
- [29] Kojadinovi, I., "Some copula inference procedures adapted to the presence of ties," *Computational Statistics and Data Analysis*, vol. 112, pp. 24-41, 2017.

نمذجة درجة حرارة الهواء المتطرفة على ارتفاع مترين في العراق باستخدام نماذج كوبيولا للقيم المتطرفة

مناف حازم احمد مطرود و زيد طارق صالح الخالدي

قسم الإحصاء والمعلوماتية، كلية علوم الحاسوب والرياضيات، جامعة الموصل، الموصل، العراق
الخلاصة

أظهرت تقارير منظمة الهجرة الدولية (IOM) التابعة للأمم المتحدة أن العراق هو البلد الخامس الأكثر تأثراً بارتفاع درجات الحرارة. ولما لهذا الارتفاع في درجات الحرارة من تأثير سلبي على جميع قطاعات الحياة منها البايولوجية، الاقتصادية وغيرها فإن ذلك يتطلب معالجة المخاطر الناتجة من هذا الارتفاع من خلال فهم دقيق لسلوك القيم المتطرفة لهذا الحدث. في هذه الدراسة، تم إجراء تحقيق عميق حول سلوك حرارة الهواء المتطرفة عند مستوى 2 متر في العراق باستخدام نماذج كوبيولا للقيم المتطرفة بالاعتماد على دوال Pickands للارتباط. أذ تم افتراض ان معلمات هذه النماذج تتغير بتغير المسافة بين المواقع، اي تم اعتبارها دوال للمسافة بين المواقع. أسفر التحقيق في هذه الحدث أن نماذج كوبيولا المتماثلة للقيم المتطرفة هي أكثر ملائمة في تمثيل سلوك درجات حرارة الهواء المتطرفة. وعليه تم بناء تسع نماذج وهم (Hüsler-Reiss A and B; Gumbel, and Galambos) ذات معلمة واحدة، بالإضافة إلى نموذج يحتوي على معلمتين، وهو t -EV. تم اختيار 50 موقعاً عشوائياً من العراق من بين 1517 موقعاً بحيث تم تقسيم هذه المواقع إلى جزئين، 40 منها لغرض النمذجة و 10 الباقية لغرض التحقق. تم استخدام طريقة تقدير (Composite pseudo-likelihood) في النمذجة. وبناءً على المعيار AIC، تم اختيار 4 نماذج كنماذج مرشحة وهم (Hüsler-Reiss A و B؛ Gumbel، و Galambos). نظراً للفرق الطفيف في قيم المعيار AIC بين

النماذج الأربعة المرشحة، تم إجراء تقييم إضافي بين النماذج المرشحة الأربعة عن طريق المقياس Kullback–Leibler لقياس التقارب بين النموذج كوبيولا الغير معلمي مع النماذج الأربعة المقدرّة باستخدام بيانات التحقق. بعد الاختبار تبين ان نموذج Hüsler–Reiss A أكثر ملائمة، نظرًا لارتفاع كثافة KL حول الصفر لجميع الأزواج بين المواقع. الكلمات الدالة: لاحصاء المكاني للقيم المتطرفة، كوبيولا للقيم المتطرفة، 2 من درجات حرارة الهواء فوق الارض ، بيكاندز للارتباط