A Bayesian Hierarchical Spatial Copula Model: An Application to Extreme Temperatures in Extremadura (Spain)

J. Agustín García 1,†, Mario M. Pizarro 2,*,†, F. Javier Acero 1,† and M. Isabel Parra 2,†

Abstract: A Bayesian hierarchical framework with a Gaussian copula and a generalized extreme value (GEV) marginal distribution is proposed for the description of spatial dependencies in data. This spatial copula model was applied to extreme summer temperatures over the Extremadura Region, in the southwest of Spain, during the period 1980–2015, and compared with the spatial noncopula model. The Bayesian hierarchical model was implemented with a Monte Carlo Markov Chain (MCMC) method that allows the distribution of the model’s parameters to be estimated. The results show the GEV distribution’s shape parameter to take constant negative values, the location parameter to be altitude dependent, and the scale parameter values to be concentrated around the same value throughout the region. Further, the spatial copula model chosen presents lower deviance information criterion (DIC) values when spatial distributions are assumed for the GEV distribution’s location and scale parameters than when the scale parameter is taken to be constant over the region.

Keywords: Bayesian hierarchical model; extreme temperature; Gaussian copula; generalized extreme value distribution

1. Introduction

Extreme events tend to occur naturally as topics of importance in several sciences—climatology, hydrology, engineering, etc.—but also in finance (financial crisis studies) and in the insurance industry. Extreme Value Theory (EVT) is a widely used statistical tool with which to address their study. It is used in several particular scientific fields to model and predict extreme events of precipitation [1–5], temperature [6–8], solar climatology [9,10], and financial crises [11,12].

In climatology, given the spatial nature of extreme events, it is useful to apply a spatial theory as this will improve the accuracy when estimating parameter distributions by sharing information from similar sites [13,14]. Several theories are used to address the problem of spatial extremes, some examples are max-stable processes [15], Bayesian hierarchical models [16–18], and copula theory [19–22].

Copula theory is increasingly being used in multivariate extreme value models in climatology [23], since Sklar’s theorem [24] allows the construction of joint distributions using the desired univariate marginal distributions. Of the different copula families, there stand out the Archimedean copulas, extreme value copulas, and elliptical copulas (in particular, the Gaussian and the Student-t copulas). These copulas constitute a convenient tool for modeling dependence in a non-Gaussian and high-dimensional context [25]. Indeed, copula theory is based on describing the dependence structure regardless of any marginal distributions [26], which makes copulas an ideal candidate for modeling dependence.

Wikle et al. [27] proposed a general Bayesian hierarchical framework in which to describe the spatial variability of the distribution of some environmental variable. Their
model comprises various layers. For example, in the first layer, it is assumed that the data follow a distribution with unknown parameters, while in a second layer, the variability of these parameters is modeled spatially by regression techniques. This kind of model has been used in many extreme rainfall studies [28–31] and also in temperature studies [32,33]. In most of these works however, it is assumed that the data are spatially independent given the values of the parameters of their distribution.

As pointed out by Cooley et al. [14], the spatial dependence among the parameters is related to what the authors called spatial climate dependence since the climate of a location or region is described by the statistical distribution of the variable of interest—in our case, the temperature. For readers interested in a climate parameter such as the return period or the return level, the hierarchical model used in the previously cited papers is enough to estimate those parameters. However, if someone is interested in, for example, transferring information from a gauged site to an ungauged site, the spatial dependence among the data must be taken into account in what Renard [20] called spatial weather dependence, because both dependencies are involved in the transferring information process. The spatial climate dependence allows the transfer of information from one place to another. However, the correlation between observatories decreases the content of the information available in nearby observatories [20]. Therefore, there must be an increase in uncertainty in the information transferred to the ungauged site when the correlation between observatories is taken into account. This key point of considering the spatial dependence among the data can be accomplished by means of a copula model.

In this sense, the main goal of the present work is to address spatial dependence by incorporating a copula into this model—specifically, a Gaussian copula (GC). An application of the proposed model is presented with the maximum temperatures recorded in the Extremadura Region, in southwestern Spain, for the period 1980–2015.

This communication is structured as follows. Section 2 describes the hierarchical model with copula, and Section 3 describes the proposed posterior distribution model and its use when inferring the GEV distribution parameters. The selected data are briefly described in Section 4, and Section 5 presents the results of applying the model to extreme temperatures in the Extremadura Region (Spain). Finally, some conclusions are drawn in Section 6.

2. Statistical Model

A Bayesian approach was chosen to address the spatial extremes problem because of its flexibility, the possibility of adding further elements or layers, and its adaptability to situations in which complex variations in the parameters of the extremes values distribution appear. In addition, one of the disadvantages shown in [18] of assuming conditional independence among observed data is resolved by introducing a Gaussian copula to control the existence of dependence, with this fact being the main difference with that model.

The proposed theoretical Bayesian hierarchical framework can be factored into the three stages shown in Figure 1. The first stage (Section 2.1) allows one to model the observations’ joint distribution with a Gaussian copula and at-site marginal GEV distributions. The second stage (Section 2.2) models the GEV parameter variances with a latent process by means of a conditional probability. In the third stage (Section 2.3), prior distributions are given for the model’s parameters. Thus, the posterior distribution of the parameters is calculated using Bayes’ theorem (see Section 3). In the following, the model’s stages will be described in more detail.

Stage 1 (data level) P (data | process, parameters)
Stage 2 (process level) P (process | parameters)
Stage 3 (prior distribution) P (parameters)

Figure 1. Stages of the Bayesian hierarchical model.
We shall denote by \( Y_{st} \) the block maximum of the variable of interest, with \( s \in S = \{ s_1, \ldots, s_M \} \) and \( t \in T = \{ t_1, \ldots, t_N \} \). For simplicity and consistency with the application, we shall refer to \( s \) as the site and \( t \) as the time. Further, for each \( s \) and \( t \), \( Y_{st} \) follows a specific distribution whose parameters vary spatially.

### 2.1. Data Level

According to the theorems of Gnedenko [34] and Fisher and Tippett [35], asymptotically, \( Y_s \) has a GEV distribution

\[
Y_s \sim GEV(\mu_s, \sigma_s, \xi_s)
\]

with cumulative distribution function (cdf)

\[
P(Y_s \leq y) = \exp \left\{ - \left[ 1 + \xi_s \left( \frac{y - \mu_s}{\sigma_s} \right) \right]^{-1/\xi_s} \right\},
\]

where \( 1 + \xi_y (y - \mu_s) / \sigma_s \geq 0 \) and in which \( \mu, \sigma, \) and \( \xi \) are the location, scale, and shape parameters, respectively.

The location parameter explains the mean values of the extreme value distribution, the scale one is referred to the variability, and the shape parameter determines the rate of decay of the upper tail of the distribution. Shape parameter values below zero indicate that the distribution has an upper bound showing that the maximum values are not getting large, and values above or equal to zero indicate that the distribution has no upper limit showing that maximum values are getting infinitely large [36].

For any set of \( M \) sites, an \( M \)-dimensional Gaussian copula is assumed for the multivariate distribution of the data, with pairwise correlation matrix \( C \) and marginal distributions \( \{ GEV(\mu_s, \sigma_s, \xi_s) \}_{s \in S} \)

\[
(Y_{s_1 t}, \ldots, Y_{s_M t}) \sim GC_M(C, \{ \mu_s, \sigma_s, \xi_s \}_{s \in S}),
\]

for which the probability density function (pdf) is shown in Appendix A.

Geostatistical models use correlation matrices that capture the positive relationship between different stations through their distance. In this case, the Gaussian copula collects this information in the correlation matrix \( C \) which is positively defined. Then, it is assumed that the correlation between data from two sites depends on the distance between them. In particular, the elements of the correlation matrix are defined as

\[
C_{ij} = \begin{cases} 
   c_0 \cdot \exp \left\{ - \frac{\| \vec{x}(s_i) - \vec{x}(s_j) \|}{c_1} \right\}, & i \neq j \\
   1, & i = j 
\end{cases}
\]

where \( c_0 \) and \( c_1 \) are unknown parameters, \( \vec{x}(s) \) is the geographical position (longitude, latitude) of the site \( s \in S \), and \( \| \vec{x}(s_i) - \vec{x}(s_j) \| \) are the distances between sites \( s_i \) and \( s_j \), for \( i, j = 1, \ldots, M \). Moreover, independence is assumed between observations \( (Y_{s_1 t_1}, \ldots, Y_{s_M t_1}) \) and \( (Y_{s_1 t'}, \ldots, Y_{s_M t'}) \) for \( t, t' \in T \) such that \( t \neq t' \).

The multivariate distribution of the data thus comprises the following components: at-site distribution (1), Gaussian copula to model the spatial dependence (3) and correlation matrix (4).

### 2.2. Process Level

The second stage of the hierarchical model consists of describing the variance of the at-site GEV parameters (location—\( \mu \), scale—\( \sigma \), and shape—\( \xi \)) through a Gaussian spatial process whose mean depends on covariates that describe the site’s characteristics. A spatial regression model as described by Garcia et al. [18] is used:
\[ \begin{align*}
\mu(s) &= X(s) \cdot a_\mu + W_\mu(s) + \epsilon_\mu(s) \\
\sigma(s) &= X(s) \cdot a_\sigma + W_\sigma(s) + \epsilon_\sigma(s) \\
\xi(s) &= X(s) \cdot a_\xi + W_\xi(s) + \epsilon_\xi(s)
\end{align*} \]

where \( s = (s_1, \ldots, s_M) \) denotes a site vector and \( p_\mu, p_\sigma, \) and \( p_\xi \) denote the number of regression parameters in each case. To simplify use of the notation, the parameters \( \mu, \sigma, \) or \( \xi \) will be represented by \( k \), so that \( X(s) \) represents \( p_k \) spatial covariates (geographic coordinates), \( a_k \) is a set of \( p_k \) regression parameters (including the intercept), \( W_k(s) \) represents a spatial model that captures the dependencies between different sites, and \( \epsilon_k(s) \) is the noise not included in the spatial model.

In general, we shall denote the proposed models by BHGCM-\( p_\mu p_\sigma p_\xi \), and the noncopula models described by Garcia et al. [18] by BHM-\( p_\mu p_\sigma p_\xi \).

With regard to Equations (5), for \( p_k = 1 \), we shall assume that no covariate associated with the site characteristics is involved, i.e.,

\[ X(s) \cdot a_k = a_{k1}, \]

and for \( p_k = 2 \), the only covariate is \( h(s) \), the altitude of site \( s \), i.e.,

\[ X(s) \cdot a_k = a_{k1} + a_{k2} \cdot h(s). \]

As a particular case, for \( p_k = 0 \), we shall assume that the parameter \( k \) is constant, and hence, the spatial model does not intervene, i.e., \( W_k(s) = 0 \).

In addition, a position-independent Gaussian model, i.e., \( N(0, \tau_\epsilon^2) \), is adopted for the pure noise effect \( \epsilon_k(s) \). The spatial term \( W_k(s) \) was considered to be a random variable with an \( M \)-dimensional normal distribution \( N(0, \Sigma_k) \), where the covariance matrix \( \Sigma_k \) follows the exponential model

\[ \Sigma_k(i, j) = \beta_{k0} \cdot \exp\left\{ -\frac{\| x(s_i) - x(s_j) \|}{\beta_{k1}} \right\}, \quad i, j = 1, \ldots, M, \]

where \( \beta_{k0} \) (the sill) and \( \beta_{k1} \) (the range) are unknown parameters.

The random variables on the left-hand side of Equation (5) are assumed to have an \( M \)-dimensional normal distribution:

\[ \begin{align*}
P(\mu(s)) &= N \left( X(s) \cdot a_\mu + W_\mu(s), \tau_\mu^2 \cdot \text{Id}_M \right) \\
P(\sigma(s)) &= N \left( X(s) \cdot a_\sigma + W_\sigma(s), \tau_\sigma^2 \cdot \text{Id}_M \right) \\
P(\xi(s)) &= N \left( X(s) \cdot a_\xi + W_\xi(s), \tau_\xi^2 \cdot \text{Id}_M \right)
\end{align*} \]

2.3. Prior Distribution

The Bayesian framework requires the prior distributions of the parameters included in the model to be specified. The prior distribution for the spatial regression parameters \( a_k \) was a \( p_k \)-dimensional normal distribution with hyperparameters chosen such that the distribution was either non- or only weakly informative. Inverse gamma distributions were taken for the sill \( \beta_{k0} \) and the variance \( \tau_\mu^2 \) parameters, and a gamma distribution for the range \( \beta_{k1} \). For the Gaussian copula parameters, \( c_0 \) and \( c_1 \), uniform prior distributions were assumed—\( U(0, 1) \) and \( U(0, 1000) \), respectively. A normal distribution with mean 0 was taken for the shape parameter. The parameters were assumed to be mutually independent.

3. Estimation

We shall apply two proposed models to the extreme temperature data (see Section 5). In the first, denoted BHGCM-200, the scale and shape parameters are constant and the location parameter is modeled as in Equation (5). In the second, denoted BHGCM-210, the shape parameter is constant and the location and scale parameters are modeled as in
Equation (5), with $p_\mu = 2$ and $p_c = 1$, respectively. The model that best fits the data will be compared with the equivalent noncopula version.

As mentioned above, Bayes’ theorem allows one to calculate the posterior distribution of a proposed model as being proportional to the product of the probabilities described in Figure 1. To simulate the posterior distribution of each of the proposed models, a Markov Chain Monte Carlo (MCMC) method was applied, in particular, using a Gibbs sampler with embedded Metropolis–Hastings steps [37] and—as appropriate given the characteristics of these methods—the Gelman–Rubin diagonal convergence test [38]. For this last test, the CODA package [39] of the R language was used.

Four parallel chains with sizes of 30,000 values were constructed starting at different points. For each chain, 10,000 values were used as burn-in, leaving 20,000 values less 10 taken for thinning. The last 2000 values of each chain were combined to form a single chain of 8000 values with which to construct the posterior distribution.

The code used to carry out the simulations was written in FORTRAN, closely following the procedure set out. Maps were prepared using the R package ggplot2 [40], with the geographical coordinates provided by Spain’s National Centre for Geographic Information (Centro Nacional de Información Geográfica, CNIG) [41].

3.1. Posterior Distribution

For each GEV parameter that changes in the model, the hierarchical framework described in Section 2 estimates the following unknowns: Gaussian copula parameters ($c_0$ and $c_1$), regression parameters ($\alpha_k$), sill parameter ($\beta_{\mu0}$), range parameter ($\beta_{\mu1}$), and variance parameter ($\tau_c^2$).

For the BHGCM-200 model, the posterior distribution is

$$P\left(c_0, c_1, \mu(\mathbf{s}), \sigma, \xi, W_\mu(\mathbf{s}), \alpha_\mu, \beta_{\mu0}, \beta_{\mu1}, \tau_\mu^2 \mid Y, h(\mathbf{s})\right),$$  \hspace{1cm} (10)

and for the BHGCM-210 model, it is

$$P\left(c_0, c_1, \mu(\mathbf{s}), \sigma(\mathbf{s}), \xi, W_\mu(\mathbf{s}), W_\sigma(\mathbf{s}), \alpha_\mu, \alpha_\sigma, \beta_{\mu0}, \beta_{\mu1}, \beta_{\sigma0}, \beta_{\sigma1}, \tau_\mu^2, \tau_\sigma^2 \mid Y, h(\mathbf{s})\right).$$  \hspace{1cm} (11)

Assuming independence between the observations $(Y_{s1t}, \ldots, Y_{sMt})$ and $(Y_{s1t'}, \ldots, Y_{sMt'})$ $(t, t' \in T$ with $t \neq t')$, the likelihood function of the observations is given by

$$L(c_0, c_1, \mu(\mathbf{s}), \sigma(\mathbf{s}), \xi | Y) := P(Y | c_0, c_1, \mu(\mathbf{s}), \sigma(\mathbf{s}), \xi) = \prod_{t \in T} f_{GC}(y_{s1t}, \ldots, y_{sMt}).$$  \hspace{1cm} (12)

This shows the details of Equations (A3)–(A5) in Appendix B.

3.2. Assessment of the Models’ Goodness-Of-Fit

The deviance information criterion (DIC) described by Spiegelhalter et al. [42] was used to choose the model that best fits the observed data. The best model has the lowest DIC value. The parameter values needed to calculate the DIC were those determined through the MCMC procedure. The criterion is defined as

$$DIC = D_\theta + p_\theta,$$

where

(a) $\theta$ is the parameter vector of interest in the model (GEV parameters in a BHM model, and GEV and Gaussian copula parameters in a BHGCM model).

(b) $D_\theta = E[D(\theta)]$ measures the model’s goodness-of-fit, where the deviance $D(\theta) = -2 \cdot \ln L(\theta | Y)$, i.e., $-2$ times the logarithm of the likelihood of the random variable $Y$ under study. In a BHGCM model, the likelihood is defined by Equation (A5), and in a BHM model, by the GEV pdf.
Algorithm 1 Ungauged Site

Do for \( l = 1 \ldots nsim \):
1. Using the well-known formula for conditional Gaussian distributions, generate \( W_k(\tilde{s}) \) with a normal distribution of mean \( \mu_{cond} \) and standard deviation \( \sigma_{cond} \) given by

\[
\begin{align*}
\mu_{cond} &= \Omega^{(l)} \cdot (\Sigma^{(l)})^{-1} \cdot W_k^{(l)}(s) \\
\sigma_{cond} &= \beta_{40} - \Omega^{(l)} \cdot (\Sigma^{(l)})^{-1} \cdot (\Omega^{(l)})^t
\end{align*}
\]

(14)

where \( W_k^{(l)}(s) \) is the \( 1 \times M \) vector generated with the gauged sites \( s \), \( \Sigma_k \) is the \( M \times M \) covariance matrix of the spatial model for the gauged sites \( s \), and \( \Omega^{(l)} \) is the \( 1 \times M \) vector of covariances between \( \tilde{s} \) and the gauged sites \( s \).

2. Compute the GEV parameters for an ungauged site \( \tilde{s} \), \( (\mu^{(l)}_{\tilde{s}}, \sigma^{(l)}_{\tilde{s}}, \xi^{(l)}_{\tilde{s}}) \), from the regression model (5).

Algorithm 2 Observations

Do for \( l = 1 \ldots nsim \):
1. Calculate the GEV distribution’s parameters for an ungauged site \( \tilde{s} \) with Algorithm 1, taking into account the GEV parameters of the gauged sites \( s \) obtained with the MCMC generated sample.
2. Generate a value of the Gaussian copula \( (u_{s_1}^{(l)}, \ldots, u_{s_M}^{(l)}, u_{\tilde{s}}^{(l)}) \) with the \( (M + 1) \times (M + 1) \) correlation matrix, \( C^{(l)} \), between \( \tilde{s} \) and the gauged sites \( s \), given by Equation (4).
3. Invert the value of the copula in accordance with the relationship \( y_s^{(l)} = F_{\text{GEV}}^{-1}(u_s^{(l)}|\mu_s^{(l)}, \sigma_s^{(l)}, \xi_{\tilde{s}}^{(l)}) \) for all \( s = s_1, \ldots, s_M, \tilde{s} \), thus yielding a vector of observations \( y^{(l)} = (y_{s_1}^{(l)}, \ldots, y_{s_M}^{(l)}, y_{\tilde{s}}^{(l)}) \).

Thus, the sample \( \left(y^{(l)}\right)_{l=1\ldots nsim} \) is a realization of the posterior predictive distribution of observations.

4. Data

The data used in this study are annual maximum observed temperatures at a set of meteorological observatories distributed over the Extremadura Region (Spain), from 1980
to 2015. These time series were provided by Spain’s State Meteorological Agency (Agencia Estatal de Meteorología, AEMET). Figure 2 shows the location of this region within Spain and the spatial distribution of the observatories considered. In particular, there are $M = 28$ meteorological observatories, each providing a time series of $N = 36$ extreme temperatures.

For a site $s \in S$, temperature and altitude are correlated, as higher altitudes mean lower temperatures, i.e., temperature decreases with increasing altitude. This reason, together with the fact that Extremadura is not a large region, led us to take the altitude of the sites as being the only covariate in the regression model (5). The altitude was standardized as follows:

$$
\tilde{h}_s = \frac{h_{\text{obs}}^s - \min_s h_s}{\max_s h_s - \min_s h_s},
$$

where $h_{\text{obs}}^s$ is the altitude above mean sea level of site $s \in S$, and $\max_s h_s$ and $\min_s h_s$ are the maximum and minimum altitudes of all the sites, respectively.

5. Results

The Bayesian spatial copula model was applied to the described data set (see Section 4). In particular, the BHGCM-200 and BHGCM-210 models were compared using the DIC as a measure of the goodness-of-fit (see Section 3.2). The parameters of the model that best fits the data were compared with the equivalent noncopula model (BHM).

5.1. Evaluation of the Models

As noted above, the DIC was employed to compare the two candidate spatial copula models. The results are presented in Table 1. One observes that the copula model in which a spatial model intervenes in the location and scale parameters’ regressions, i.e., BHGCM-210, has a lower DIC value than the model in which the scale parameter is constant (BHGCM-200).

Table 1. Results of using the DIC for the copula models. Boldface indicates the better model.

| Model        | $\hat{D}_\theta$ | $D(\hat{\theta})$ | $p_{\theta}$ | DIC     |
|--------------|------------------|-------------------|--------------|---------|
| BHGCM-200    | 3697.57          | 3668.49           | 29.08        | 3726.64 |
| BHGCM-210    | 3589.41          | 3536.57           | 52.83        | 3642.24 |
Therefore, the spatial copula model chosen is BHGCM-210. Table 2 gives the results of applying the goodness-of-fit to the BHM-210 model. Recall that the parameter \( p_\theta \) indicates the complexity of the model, and, as can be observed in Table 2, it is greater in the BHGCM model, which was to be expected given that the proposed models are intrinsically more complex. However, this increase in complexity over the models proposed by García et al. [18] is acceptable since the variance in the models proposed in the present work is less than in the BHM cases.

| Model     | \( \hat{D}_\theta \) | \( D(\hat{\theta}) \) | \( p_\theta \) | DIC  |
|-----------|-----------------------|------------------------|----------------|------|
| BHM-210   | 4046.96               | 3997.51                | 49.46          | 4096.42 |
| BHGCM-210 | 3589.41               | 3536.57                | 52.83          | 3642.24 |

5.2. Parameter Estimates

In this subsection, the regression parameters of the BHM-210 model and the proposed BHGCM-210 model are compared.

Figure 3 shows the estimate of the posterior distribution density function from the selected BHGCM-210 model (red line) versus that estimated by the BHM-210 model (blue line) for the different regression coefficients \( \alpha \). The regression coefficient \( \alpha_{\mu 1} \) (lower left panel) of the scale parameter, \( \sigma \), has a mean value of 0.78 °C (SD: 0.51 °C) for the BHGCM model. Moreover, this model provides an estimate of the density function that is less concentrated than that given by the BHM model.

With regard to the regression coefficients for the location parameter, \( \mu \) (top row), the two models give qualitatively similar posterior density function estimates. In particular, the coefficient \( \alpha_{\mu 1} \) (top left panel) has a mean of 41.12 °C (SD: 0.84 °C), while the coefficient \( \alpha_{\mu 2} \) (top right panel) is clearly negative with a mean of −2.97 °C km\(^{-1}\) (SD: 1.15 °C km\(^{-1}\)). These negative values are consistent with the fact that temperature decreases with altitude.

Another interesting result is the covariance function of the GEV parameters given by Equation (8). Table 3 lists the medians and (2.5%, 97.5%) quantiles of the sill (\( \beta_{\mu 0} \)) and range (\( \beta_{\mu 1} \)) coefficients for the location and scale parameters in models BHM-210 and BHGCM-210. The values of these coefficients are of similar orders of magnitude for the two models. Since the range is a measure of the strength of spatial dependence for the location and scale parameters, one observes that, in both models, this dependence is weaker for the location parameter than for the scale parameter.
Table 3. Median (2.5%, 97.5%) of the sill and range coefficients for the location and scale parameters of models BHM-210 and BHGCM-210.

| Model     | Location Sill  | Location Range      | Scale Sill   | Scale Range         |
|-----------|----------------|---------------------|--------------|---------------------|
| BHM-210   | 0.52 (0.18, 2.08) | 395.07 (132.90, 885.62) | 0.25 (0.11, 0.62) | 554.84 (235.55, 1110.02) |
| BHGCM-210 | 0.53 (0.18, 2.15) | 389.82 (125.59, 872.49) | 0.27 (0.12, 0.71) | 562.47 (233.41, 1133.02) |

5.3. Validation of the Models

The posterior predictive distribution (PPD) provides temperature values for the measured observatories that can be compared to the observed temperatures. The error (E) and the absolute error (AE) were used to validate the BHGCM-210 model. These errors were calculated using the values obtained from the PPD. They define as

\[
E_{st} = \hat{Y}_{st} - Y_{st}
\]

\[
AE_{st} = |\hat{Y}_{st} - Y_{st}|
\]

where \(\hat{Y}_{st}\) is the predictive temperature, \(Y_{st}\) is the observed temperature, \(s \in S = \{s_1, \ldots, s_M\}\), and \(t \in T = \{t_1, \ldots, t_N\}\).

Figure 4 shows the density function of E (left panel) and AE (right panel) for the BHM-210 model (black color) and BHGCM-210 model (red color). The error’s density functions are symmetric around their mean values of 0.23 and 0.06 for the BHGCM-210 and BHM-210 models, respectively. Note that the mean values represent the bias of the models. The precision, measured as the absolute error, is slightly better in the copula model than in the noncopula model. This is related with the increase in uncertainty, previously mentioned in the Introduction.

Figure 5 shows the Q-Q plots for all observatories between the theoretical and empirical quantile of \(Y_{st}\) and \(\hat{Y}_{st}\), respectively, for BHM-210 (left panel) and BHGCM-210 (right panel) models. It can be seen that the ratio between both is close to the red 1:1 line.
5.4. Inference

The chosen model (BHGC-210) estimates the location, scale, and shape parameters of the at-site GEV distribution. These parameters are predicted at the ungauged sites with Algorithm 1. Figure 6 shows the estimated posterior distribution density function for the shape parameter. This parameter takes clearly negative values, with a symmetric and homogeneous distribution around the mean value $-0.38$ and a standard deviation of $0.03$, indicating that there is an upper bound on how high extreme temperatures can be. This leads to a quick decrease of the rate of decay of the extreme temperature distribution, so it does not increase infinitely. Moreover, it is relevant to highlight the low value of the standard deviation obtained when compared with that obtained for the shape parameter estimated in each location individually. Using the maximum-likelihood fit, the standard deviation of the shape parameter varies from 0.09 to 0.12—meaning more than three times that obtained with the spatial model. Therefore, pooling nearby data leads to a decrease in the statistical uncertainty of the parameters and implies an important advantage of using spatial models.

Figure 7 shows the spatial posterior distributions of the means and standard deviations of the location and scale parameters. The spatial posterior distribution of the location parameter shows its dependence on altitude, with mean values between 39.29 °C and 41.12 °C and standard deviations between 1.31 °C and 1.41 °C. The areas with low values of the location parameter correspond to those of higher altitude, since the temperature is highly determined by the orography (see Figure 2). The spatial posterior distribution of the scale parameter shows no spatial dependence, taking very similar values over the entire
region (between 2.96 °C and 3.24 °C), and the standard deviations are very concentrated throughout the region.

![Spacial posterior distributions of the mean and standard deviation](image)

**Figure 7.** Spatial posterior distributions of the mean (left) and the standard deviation (right) of the location (top row) and scale (bottom row) parameters.

Figure 8 shows the estimation of the maximum temperatures in Extremadura through the posterior predictive distribution. These estimations are consistent with the estimated values for the location parameter given in the previous figure.

![Spatial posterior predictive distribution of the maxima temperature](image)

**Figure 8.** Spatial posterior predictive distribution of the maxima temperature.

### 6. Conclusions

1. Bayesian hierarchical models, BHM, proposed by García et al. [18] present the problem of assuming spatial independence between observations at different sites. The present work has addressed this problem by introducing a copula.

2. A Gaussian copula is assumed as a joint distribution with at-site GEV marginal distributions. In this way, the spatial dependence of observations from different sites
is represented by a correlation matrix. In addition, spatial regression models of the GEV parameters are proposed.

3. Two BHGCM models are proposed: BHGCM-200 takes a spatial regression model for \( \mu \) while the parameters \( \sigma \) and \( \xi \) are constant; BHGCM-210 takes spatial models for \( \mu \) and \( \sigma \), while the parameter \( \xi \) is constant.

4. The BHGCM-210 model has a better DIC goodness-of-fit value than the BHGCM-200 model and the noncopula BHM-210 model.

5. For the GEV distribution’s location parameter, the BHGCM-210 and BHM-210 models give qualitatively similar estimates of the regression parameter posterior distributions.

6. For the GEV distribution’s scale parameter, the BHGCM-210 model gives a distribution with greater variance than that given by the BHM-210 model.

7. In the BHGCM-210 model, the GEV shape parameter takes negative values, and its posterior distribution is symmetrical and highly concentrated around \(-0.38\). Therefore, the extreme temperature distribution is not expected to increase too much.

8. The BHGCM-210 model gives a spatial posterior distribution for the location parameter that is strongly dependent on altitude, unlike the scale parameter. The location parameter’s mean values in the region lie between 39.29 °C and 41.12 °C.

9. In the BHGCM-210 model, the scale parameter’s spatial posterior distribution is very concentrated, taking very similar values throughout the region.

Author Contributions: All the authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by FEDER / Ministerio de Ciencia, Innovación y Universidades – Agencia Estatal de Investigación / Proyecto MTM2017-86875-C3-2-R, and by the Junta de Extremadura, FEDER Funds, GR18108, GR18097 and IB16063 (Consejería de Economía e Infraestructuras).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study belong to Spain’s State Meteorological Agency (Agencia Estatal de Meteorología: www.aemet.es) (accessed on 9 July 2021). The code is available upon request to the author.

Acknowledgments: Thanks are due to the Spanish State Meteorological Agency for providing the daily temperature time series used in this study.

Conflicts of Interest: The authors declare that there were no conflicts of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in writing the manuscript, or in the decision to publish the results.

Appendix A. Gaussian Copula

Definition A1. For every \( M \geq 2 \), an \( M \)-dimensional copula is an \( M \)-variate distribution function on \([0,1]^M\) whose marginals are uniformly distributed on \([0,1]\).

In copula theory, the most important theorem is Sklar’s Theorem, because it permits establishing a relation between multivariate distributions and copulas.

Theorem A1 (Sklar’s Theorem). Let \( F \) be a \( M \)-dimensional distribution function with univariate \( F_1, F_2, \ldots, F_M \). Let \( A_j := F_j(\mathbb{R}) \) denote the range of \( F_j \) \((j = 1, 2, \ldots, M)\). Then, there exists a copula \( C \) such that for all \((x_1, x_2, \ldots, x_M) \in \mathbb{R}^M\),

\[
F(x_1, x_2, \ldots, x_M) = C(F_1(x_1), F_2(x_2), \ldots, F_M(x_M)). \tag{A1}
\]

Such a \( C \) is uniquely determined on \( A_1 \times A_2 \times \cdots \times A_M \), hence, it is unique when \( F_1, F_2, \ldots, F_M \) are all continuous.
An elliptical copula is a family of copulas. Gaussian copula is one of them. This copula has probability density function (pdf)

\[
    f_{GC}(y_1, \ldots, y_M) = \frac{\prod_{i=1}^{M} f_{GEV}(y_i | \mu_i, \sigma_i, \xi_i)}{\prod_{i=1}^{M} f_N(u_i)} \cdot f_{NM}(u_1, \ldots, u_M | C),
\]

where \( f_{GEV} \) represents the GEV distribution's pdf (1), \( f_N \) the pdf of the standard normal distribution, \( f_{NM} \) the pdf of the \( M \)-dimensional normal distribution with correlation matrix \( C \), and \( u_i \) the quantile of \( P(Y_{t,i} \leq y_i) \) (2) of the standard normal distribution for \( i = 1, \ldots, M \).

**Appendix B. Estimation**

The posterior distributions of different models are

\[
    P(c_0, c_1, \mu(s), \sigma(s), W_p(s), \alpha, \beta_{\mu,0}, \beta_{\mu,1}, t_\mu^2, C | Y, h(s)) \propto P(Y | c_0, c_1, \mu(s), \sigma(s), \xi) \times P(\mu(s) | h(s), W_p(s), \alpha, \beta_{\mu,0}, \beta_{\mu,1}, t_\mu^2) \times P(\sigma(s) | C) \times P(\xi) \times P(W_p(s) | \beta_{\mu,0}, \beta_{\mu,1}) \times P(\alpha) \times P(\beta_{\mu,0}) \times P(\beta_{\mu,1}) \times P(t_\mu^2),
\]

for the BHGCM-200 model; for the BHGCM-210 model, it is

\[
    P(c_0, c_1, \mu(s), \sigma(s), W_p(s), W_r(s), \alpha, \beta_{\mu,0}, \beta_{\mu,1}, \beta_{\sigma,0}, \beta_{\sigma,1}, t_\mu^2, t_\sigma^2 | Y, h(s)) \propto P(Y | c_0, c_1, \mu(s), \sigma(s), \xi) \times P(\mu(s) | h(s), W_p(s), \alpha, \beta_{\mu,0}, \beta_{\mu,1}, t_\mu^2) \times P(\sigma(s) | W_r(s), \alpha, \beta_{\sigma,0}, \beta_{\sigma,1}, t_\sigma^2) \times P(\xi) \times P(W_p(s) | \beta_{\mu,0}, \beta_{\mu,1}) \times P(W_r(s) | \beta_{\sigma,0}, \beta_{\sigma,1}) \times P(\alpha) \times P(\beta_{\mu,0}) \times P(\beta_{\mu,1}) \times P(t_\mu^2) \times P(t_\sigma^2).
\]

Additionally, the likelihood function given in (12) can be written in logarithmic terms as

\[
    \ln L(c_0, c_1, \mu(s), \sigma(s), \xi | Y) = -N \cdot \sum_{i=1}^{M} \ln \sigma_i - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{M} \sum_{j=1}^{N} \ln \left(1 + \frac{y_{t,i} - \mu_i}{\sigma_i}\right) - \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(1 + \frac{y_{t,i} - \mu_i}{\sigma_i}\right)^{-1/\xi} \sum_{j=1}^{N} y_{t,j}^2 - \frac{1}{2} N \cdot \ln |C| - \frac{1}{2} \sum_{j=1}^{N} (u_{t,1j}, \ldots, u_{t,MT})^t C^{-1} (u_{t,1j}, \ldots, u_{t,MT})
\]

For the BHGCM-200 model, the above equation is simpler since \( \sigma_i = \sigma \) for \( i = 1, \ldots, M \).

**References**

1. García, J.; Gallego, M.C.; Serrano, A.; Vaquero, J. Trends in Block-Seasonal Extreme Rainfall over the Iberian Peninsula in the Second Half of the Twentieth Century. *J. Clim.* 2007, 20, 113–130. [CrossRef]
2. Re, M.; Barros, V.R. Extreme rainfalls in SE South America. *Clim. Chang.* 2009, 96, 119–136. [CrossRef]
3. Acero, F.J.; García, J.A.; Gallego, M.C. Peaks-over-Threshold Study of Trends in Extreme Rainfall over the Iberian Peninsula. *J. Clim.* 2011, 24, 1089–1105. [CrossRef]
4. Acero, F.J.; Parey, S.; Hoang, T.H.I.; Dacunha-Castelle, D.; García, J.A.; Gallego, M.C. Non-stationary future return levels for extreme rainfall over Extremadura (SW Iberian Peninsula). *Hydrol. Sci. J.* 2017, 62, 1394–1411. [CrossRef]
5. Wi, S.; Valdés, J.B.; Steinschneider, S.; Kim, T.W. Non-stationary frequency analysis of extreme precipitation in South Korea using peaks-over-threshold and annual maxima. *Stoch. Environ. Res. Risk Assess.* 2016, 30, 583–606. [CrossRef]
6. Nogaj, M.; Yiou, P.; Parey, S.; Malek, F.; Naveau, P. Amplitude and frequency of temperature extremes over the North Atlantic region. *Geophys. Res. Lett.* 2006, 33. [CrossRef]
7. Coelho, C.A.S.; Ferro, C.A.T.; Stephenson, D.B.; Steinskog, D.J. Methods for Exploring Spatial and Temporal Variability of Extreme Events in Climate Data. *J. Clim.* 2008, 21, 2072–2092. [CrossRef]

8. Acero, F.J.; Fernández-Fernández, M.I.; Carrasco, V.M.S.; Parey, S.; Hoang, T.T.H.; Dacunha-Castelle, D.; García, J.A. Changes in heat wave characteristics over Extremadura (SW Spain). *Theor. Appl. Climatol.* 2018, 133, 605–617. [CrossRef]

9. Ramos, A.A. Extreme value theory and the solar cycle. *Astron. Astrophys.* 2007, 472, 293–298. [CrossRef]

10. Acero, F.J.; Carrasco, V.M.S.; Gallego, M.C.; García, J.A.; Vaquero, J.M. Extreme Value Theory Applied to the Millennial Sunspot Number Series. *Astrophys. J.* 2018, 853, 80. [CrossRef]

11. Longin, F.M. From value at risk to stress testing: The extreme value approach. *J. Bank. Financ.* 2000, 24, 1097–1130. [CrossRef]

12. Castillo, E.; Hadi, A.S.; Balakrishnan, N.; Sarabia, J.M. *Extreme Value and Related Models with Applications in Engineering and Science*; Wiley: Hoboken, NJ, USA, 2004.

13. Casson, E.; Coles, S. Spatial regression models for extremes. *Extremes* 1999, 1, 449–468. [CrossRef]

14. Cooley, D.; Nychka, D.; Naveau, P. Bayesian spatial modeling of extreme precipitation return levels. *J. Am. Stat. Assoc.* 2007, 102, 824–840. [CrossRef]

15. Portero, J.; Acero, F.J.; Garcia, J.A. Analysis of Extreme Temperature Events over the Iberian Peninsula during the 21st Century Using Dynamic Climate Projections Chosen Using Max-Stable Processes. *Atmosphere* 2020, 11, 506. [CrossRef]

16. Davison, A.C.; Padoan, S.A.; Ribatet, M. Statistical modeling of spatial extremes of spatial extremes. *Stat. Sci.* 2012, 27, 161–186. [CrossRef]

17. Acero, F.J.; García, J.A.; Gallego, M.C.; Parey, S.; Dacunha-Castelle, D. Trends in summer extreme temperatures over the Iberian Peninsula using nonurban station data. *J. Geophys. Res. Atmos.* 2014, 119, 39–53. [CrossRef]

18. García, A.; Martín, J.; Naranjo, L.; Acero, F.J. A Bayesian hierarchical spatio-temporal model for extreme rainfall in Extremadura (Spain). *Hydrol. Sci. J.* 2018, 63, 878–894. [CrossRef]

19. Renard, B.; Lang, M. Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. *Adv. Water Resour.* 2007, 30, 897–912. [CrossRef]

20. Renard, B. A Bayesian hierarchical approach to regional frequency analysis. *Water Resour. Res.* 2011, 47, 11513. [CrossRef]

21. Liu, Y.R.; Li, Y.P.; Ma, Y.; Jia, Q.M.; Su, Y.Y. Development of a Bayesian-copula-based frequency analysis method for hydrological risk assessment—The Naryn River in Central Asia. *J. Hydrool.* 2020, 580, 124349. [CrossRef]

22. Beck, N.; Genest, C.; Jalbert, J.; Mailhot, M. Predicting extreme surges from sparse data using a copula-based hierarchical Bayesian spatial model. *Environmetrics* 2020, 31, e2616. [CrossRef]

23. Salvadori, G.; De Michele, C.; Kottekoda, N.T.M.; Rosso, R. *Extremes in Nature: An Approach Using Copulas*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2007; Volume 56.

24. Sklar, A. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* 1959, 8, 229–231.

25. Genest, C.; Favre, A.C.; Béliveau, J.; Jacques, C. Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resour. Res.* 2007, 43, W09401. [CrossRef]

26. Favre, A.C.; El Adlouni, S.; Perreault, L.; Thiémonge, N.; Bobée, B. Multivariate hydrological frequency analysis using copulas. *Water Resour. Res.* 2004, 40. [CrossRef]

27. Wikle, C.K.; Berliner, M.L.; Cressie, N. Hierarchical Bayesian space-time models. *Environ. Ecol. Stat.* 1998, 5, 117–154. [CrossRef]

28. Sun, X.; Thyer, M.; Renard, B.; Lang, M. A general regional frequency analysis framework for quantifying local-scale climate effects: A case study of ENSO effects on Southeast Queensland rainfall. *J. Hydrool.* 2014, 512, 53–68. [CrossRef]

29. Dyrrdal, A.V.; Lenkoski, A.; Thorarindsottir, T.L.; Stordal, F. Bayesian hierarchical modeling of extreme hourly precipitation in Norway. *Environmetrics* 2015, 26, 89–186. [CrossRef]

30. Ragulinia, G.; Reitan, T. Generalized extreme value shape parameter and its nature for extreme precipitation using long time series and Bayesian approach. *Hydrol. Sci. J.* 2017, 62, 863–879. [CrossRef]

31. Barlow, A.M.; Rohrbeck, C.; Sharkey, P.; Shooter, R.; Simpson, E.S. A Bayesian spatio-temporal model for precipitation extremes: STOR team contribution to the EVA2017 challenge. *Extremes* 2018, 21, 431–439. [CrossRef]

32. Craigmile, P.F.; Guttorp, P. Can a regional climate model reproduce observed extreme temperatures? *Statistica* 2013, 73, 103–122.

33. Daraio, J.A.; Amponsah, A.O.; Sears, K.W. Bayesian Hierarchical Regression to Assess Variation of Stream Temperature with Effects: A case study of ENSO effects on Southeast Queensland rainfall. *J. Hydrol.* 2014, 512, 53–68. [CrossRef]

34. Dyrrdal, A.V.; Lenkoski, A.; Thorarindsottir, T.L.; Stordal, F. Bayesian hierarchical modeling of extreme hourly precipitation in Norway. *Environmetrics* 2015, 26, 89–186. [CrossRef]

35. Fisher, R.A.; Tippett, L.H.C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Ann. Math.* pp. 180–190. [CrossRef]

36. Coles, S. *An Introduction to Statistical Modeling of Extreme Values*; Springer: London, UK, 2001; Volume 208.

37. Gilks, W.R.; Richardson, S.; Spiegelhalter, D.J. *Introducing Markov Chain Monte Carlo*. In *Markov Chain Monte Carlo in Practice*; Chapman & Hall: New York, NY, USA, 1996.

38. Cowles, M.K.; Carlin, B.P. *Markov chain Monte Carlo convergence diagnostics: A comparative review*. *J. Am. Stat. Assoc.* 1996, 91, 883–904. [CrossRef]

39. Martyn, P.; Nicky, B.; Kate, C.; Karen, V. *CODA: Convergence diagnosis and output analysis for MCMC*. *R News* 2006, 6, 7–11.

40. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Version 2 (1); Springer-Verlag: New York, NY, USA, 2016.
41. Centro Nacional de Información Geográfica. Modelo Digital del Terreno 2015 CC-BY 4.0. Available online: scne.es (accessed on 9 July 2021).

42. Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P.; van der Linde, A. The deviance information criterion: 12 years on. J. R. Stat. Soc. 2014, 76, 485–493. [CrossRef]

43. Gelman, A.; Carlin, J.; Stern, H.; Rubin, D. Bayesian Data Analysis, 2nd ed.; Texts in Statistical Science; Chapman and Hall: New York, NY, USA, 1995; 696p.