TEST

A simple and useful regression model for bimodal extreme data]{A simple and useful regression model for bimodal extreme data.

--Manuscript Draft--

Manuscript Number:	SEIO-D-25-00070
Full Title:	A simple and useful regression model for bimodal extreme data]{A simple and useful regression model for bimodal extreme data.
Article Type:	Original research paper
Corresponding Author:	CIRA ETHEOWALDA GUEVARA OTINIANO University of Brasilia: Universidade de Brasilia Brasília, DF BRAZIL
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Brasilia: Universidade de Brasilia
Corresponding Author's Secondary Institution:	
First Author:	CIRA ETHEOWALDA GUEVARA OTINIANO
First Author Secondary Information:	
Order of Authors:	CIRA ETHEOWALDA GUEVARA OTINIANO
	Mathews N.S. Lisboa, Master in Statistics
	Terezinha K.A. Ribeiro, Ph.D.
	Juliana B. Fachini-Gomes, Ph.D.
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	Extreme weather events result from the interaction of extreme values of two or more random climate variables. When at least one of these variables exhibits bimodal characteristics, climate systems become more complex. The generalized bimodal distribution of extreme values emerges as an alternative method for modeling univariate extremes in complex climate systems. In this paper, we present a new regression model for extreme data from complex systems, where the distribution of the response variable is a reparameterization at the median of the generalized distribution of bimodal extreme values. In the context of extreme value theory, this is the first regression model for bimodal extreme data that incorporates both heavy and light tails. The main advantage of our new approach is the straightforward interpretation of the regression coefficients in terms of the median. We employ the maximum likelihood method to estimate the model parameters and conduct simulation studies to empirically demonstrate the behavior of the estimators. The proposed model was applied to climate data, and its adequacy was validated through the analysis of quantile residuals.
Suggested Reviewers:	Fábio Prataviera, P.h.D. Professor, USP: Universidade de Sao Paulo fabio_prataviera@usp.br One of his research areas is regression models.
	Guilherme Pumi, P.h.D. Professor, UFRGS: Universidade Federal do Rio Grande do Sul guipumi@gmail.com One of his research areas is regression models.

A simple and useful regression model for bimodal extreme data

Cira E. G. Otiniano[1*], Mathews N. S. Lisboa[1], Terezinha K. A. Ribeiro[1], Juliana B. Fachini-Gomes[1]

[1]Statistics Department, University of Brasília, Brasília, 70910-900, DF, Brazil.

*Corresponding author(s). E-mail(s): ciragotiniano@gmail.com; Contributing authors: mnsl.meiter@gmail.com; terezinha.ribeiro@unb.br; jfachini@unb.br;

Abstract

Extreme weather events result from the interaction of extreme values of two or more random climate variables. When at least one of these variables exhibits bimodal characteristics, climate systems become more complex. The generalized bimodal distribution of extreme values emerges as an alternative method for modeling univariate extremes in complex climate systems. In this paper, we present a new regression model for extreme data from complex systems, where the distribution of the response variable is a reparameterization at the median of the generalized distribution of bimodal extreme values. In the context of extreme value theory, this is the first regression model for bimodal extreme data that incorporates both heavy and light tails. The main advantage of our new approach is the straightforward interpretation of the regression coefficients in terms of the median. We employ the maximum likelihood method to estimate the model parameters and conduct simulation studies to empirically demonstrate the behavior of the estimators. The proposed model was applied to climate data, and its adequacy was validated through the analysis of quantile residuals.

Keywords: Bimodal model, Complex systems, Median regression, Extremes.

1 Introduction

Several extreme value datasets (maximums and minimums) in the financial and climate areas exhibit heterogeneity characteristics and heavy tails. In most cases, this heterogeneity can result in bimodality in the data. Examples include the mix of maximum with minimum values of financial returns (Otiniano et al. 2023), minimum temperature (Simões et al. 2024), maximum temperature at the oval point (Otiniano et al. 2023), maximum pressure (Otiniano et al. 2023), and combinations such as maximum temperature, minimum humidity, and wind speed (Otiniano et al. 2024). Thus, the extreme distributions of Fréchet, Weibull, or Gumbel (Fisher and Tippett 1928; Gnedenko 1943) and the generalized extreme value (GEV) distribution (Jenkinson 1955) are not able to correctly model this type of heterogeneous extreme data. The bimodal GEV (BGEV) distribution, initially proposed in Otiniano et al. (2023) and redefined by Otiniano et al. (2024), offers a suitable alternative for modeling heterogeneous stationary extreme data. Furthermore, extremes of climate variables are expressed by changes in the location parameter of the distribution. Kharin and Zwiers (2004) simulated transient climate changes in temperature and precipitation extremes using the second-generation coupled global climate model from the Canadian Climate Center. The return values of annual extremes were estimated using a GEV distribution. One of the study's conclusions was that changes in temperature extremes are largely associated with shifts in the location of the annual extreme distribution, without substantial changes in its shape throughout the year.

On the other hand, climate variables such as temperature, relative humidity, and precipitation, among others, depend on other variables such as radiation, pressure, and wind speed (Schlatter 1987). In statistics, a regression model can show whether changes observed in the dependent variable (response) are associated with changes in one or more independent variables (covariates).

Recent regression models for bimodal data include a new generalized odd log-logistic flexible Weibull regression model (Prataviera et al. 2018), the Gaussian exponential log-logistic regression model (Vasconcelos et al. 2021b), the bimodal gamma regression model (Vila et al. 2020), the extended Weibull regression model for censored data (Rodrigues et al. 2022), and the regression model based on the log-normal log-logistic log-odd (OLLLN) distribution (Vasconcelos et al. 2021a). Although the authors refer to this model as "a regression model for extreme events in the presence of bimodality", the OLLLN distribution does not originate from extreme value theory. Furthermore, none of these models can capture heavy-tailed events, which are common in extreme events.

We propose a new parameterization of the BGEV distribution, in which one of its parameters is the median. With this reparameterization, we introduce a regression model for unimodal and bimodal extreme data, structuring the regression on the median of the reparameterized BGEV distribution. This approach allows for a more direct and intuitive interpretation of the results. Since the BGEV distribution includes data with heavy and light tails, the new model becomes flexible enough to be applied to unimodal and bimodal data from various areas. Estimation and inference of the model parameters were performed by maximum likelihood, and expressions for the score function were derived. In addition, due to the simple closed form of the quantile function of the BGEV distribution, the behavior of the estimators can be easily evaluated through Monte Carlo simulations. The computational procedures were implemented in the R software. Finally, we applied the BGEV regression model to climate data collected by a meteorological station located in a city near Brasília, the capital of Brazil.

The remainder of the paper is organized as follows. In Section 2, we begin with the reparameterization of the BGEV distribution, where one of its parameters is the median. We then formulate the BGEV regression model and discuss parameter inference, diagnostics, and model selection. In Section 3, we present some numerical results on the performance of the estimators. In Section 4, we illustrate the applicability of the new regression model to climate data. Finally, in Section 5, we present our conclusions.

2 Main results

According to Otiniano et al. (2023), the BGEV distribution serves as a natural model in extreme value theory and effectively accommodates data with two modes. In this section, we introduce a new class of regression models based on the reparameterized BGEV distribution. This new class is called the BGEV regression model.

2.1 Reparameterization of the BGEV distribution

From Otiniano et al. (2024), a random variable Y has BGEV distribution with four parameters if its cumulative distribution function (CDF); $Y \sim \mathrm{BG}(\xi, \mu, \sigma, \delta)$, is given by

$$F(y; \xi, \mu, \sigma, \delta) = F(T(y); \xi, 0, \sigma), \tag{1}$$

where

$$F(y; \xi, 0, \sigma) = \begin{cases} \exp\left\{-\left[1 + \xi\left(\dfrac{y}{\sigma}\right)\right]^{-1/\xi}\right\}, & \text{if } \xi \neq 0, \\ \exp\left\{-\exp\left[-\left(\dfrac{y}{\sigma}\right)\right]\right\}, & \text{if } \xi = 0, \end{cases} \tag{2}$$

is the CDF of the GEV distribution with zero location parameter, shape $\xi \in \mathbb{R}$, and scale $\sigma > 0$, and

$$T(y) = (y - \mu)\left|y - \mu\right|^{\delta}, \quad \delta > -1, \quad \mu \in \mathbb{R} \tag{3}$$

is an invertible transformation with inverse function

$$T^{-1}(y) = \mathrm{sgn}(y)|y|^{1/(\delta+1)} + \mu. \tag{4}$$

Furthermore, its derivative is

$$T'(y) = (\delta + 1)|y - \mu|^{\delta}. \tag{5}$$

By deriving the function (1), we have that the probability density function (PDF) of Y, in terms of transformation (3) is

$$f(y;\xi,\mu,\sigma,\delta)=\begin{cases}\dfrac{1}{\sigma}\left[1+\xi\left(\dfrac{T(y)}{\sigma}\right)\right]^{(-1/\xi)-1}\exp\left[-\left[1+\xi\left(\dfrac{T(y)}{\sigma}\right)\right]^{-1/\xi}\right]T'(y), & \xi\neq 0\\[2ex] \dfrac{1}{\sigma}\exp\left(-\dfrac{T(y)}{\sigma}\right)\exp\left[-\exp\left(-\dfrac{T(y)}{\sigma}\right)\right]T'(y), & \xi=0,\end{cases}\tag{6}$$

whose support is

$$\text{Support}(f(\cdot;\xi,\mu,\sigma,\delta))=\begin{cases}\left[\mu-\left(\dfrac{\sigma}{\xi}\right)^{1/(\delta+1)},\ +\infty\right), & \text{if } \xi>0,\\[2ex] \left(-\infty,\ \mu+\left|\dfrac{\sigma}{\xi}\right|^{1/(\delta+1)}\right], & \text{if } \xi<0,\\[2ex] (-\infty,\ +\infty), & \text{if } \xi=0.\end{cases}\tag{7}$$

We highlight that in model (6), ξ, δ and σ are shape parameters, while μ is a location parameter. It is important to note that σ in the base GEV distribution is a scale parameter, but in the BGEV distribution, it does not satisfy the condition $f(y;\xi,\mu,\sigma,\delta)=(1/\sigma)f(y/\sigma;\xi,\mu,1,\delta)$, since $T(y)/\sigma\neq T(y/\sigma)$. The proof that μ is a location parameter follows from the fact that $T(y;\mu,\delta)=T(y-\mu;0,\delta)$ and $F(y;\xi,\mu,\sigma,\delta)=F(y-\mu;\xi,0,\sigma,\delta)$.

From (2), (1), and (4) we have that the τ-th quantile of the of BGEV distribution, with $\tau\in(0,1)$, is given by

$$q_\tau=\begin{cases}\operatorname{sgn}\left\{\frac{\sigma}{\xi}\left[(-\log(\tau))^{-\xi}-1\right]\right\}\left|\frac{\sigma}{\xi}[(-\log(\tau))^{-\xi}-1]\right|^{1/(\delta+1)}+\mu, & \text{if } \ \xi\neq 0,\\ \operatorname{sgn}\left[-\sigma\log\left(-\log\left(\tau\right)\right)\right]\ \left|-\sigma\log\left(-\log\left(\tau\right)\right)\right|^{1/(\delta+1)}+\mu, & \text{if } \ \xi=0.\end{cases}\tag{8}$$

In the formulation of the BGEV distribution (6), μ is a location parameter. However, μ is neither the mean nor any other simple characteristic of Y, and therefore does not have a direct interpretation.

In regression modeling, it is more common to directly model the mean or median parameter of the distribution. Hence, in this work, we introduce a median-based reparameterization of the BGEV distribution.

When $\tau=0.5$, q_τ corresponds to the median (m) of BGEV distribution. That is, $m=q_{0.5}$. By isolating the parameter μ from (8), in terms of the median, we have that

$$\mu^*=\begin{cases}m-\left[\dfrac{\sigma}{\xi}\left(c_1^{-\xi}-1\right)\right]^{1/(\delta+1)}, & \xi\neq 0,\\[3ex] m-\left(c_2\sigma\right)^{1/(\delta+1)}, & \xi=0,\end{cases}\tag{9}$$

where $c_1=-\log(0.5)$ and $c_2=-\log(-\log(0.5))$ are constants.

The idea proposed here is to replace the parameter μ in equation (1) by the function μ^* from (9). Thus, the parameters of the reparameterized BGEV distribution are ξ, m, σ and δ. This approach will result in a simpler and more interpretable model based on the BGEV distribution, with the regression structure assigned to the median parameter m.

Thus, a random variable Y has a reparameterized BGEV distribution if its CDF; $Y \sim \text{BGEV}(\xi, m, \sigma, \delta)$, is given by

$$F^*(y;\xi,m,\sigma,\delta) = \begin{cases} \exp\left\{-\left[1+\xi\left(\dfrac{T^*(y)}{\sigma}\right)\right]^{-1/\xi}\right\}, & \xi \neq 0, \\ \exp\left\{-\exp\left[-\dfrac{T^*(y)}{\sigma}\right]\right\}, & \xi = 0, \end{cases} \tag{10}$$

where

$$T^*(y) = (y-\mu^*)\,|y-\mu^*|^{\delta}. \tag{11}$$

The PDF $f^*(\cdot;\xi,m,\sigma,\delta)$ of reparameterized BGEV distribution is given by expression (6) replacing T by T^*, and its support is given by (7) replacing μ by μ^*.

Figure 1 shows the flexibility of the PDF of the BGEV distribution when varying one parameter at a time, keeping the other three fixed. Depending on the combination of parameters ξ, m, σ, δ, the PDF can be unimodal or bimodal, symmetric or asymmetric, and exhibit either a heavy or light tail. It is evident that the parameters ξ, δ, and σ alter the shape of the distribution as expected. The parameter m, which represents the median of the distribution, shifts the curve of the BGEV distribution. Furthermore, we observe that values of $\delta > 0$ result in bimodal behavior.

2.2 BGEV regression model

To establish the regression structure, consider the following. Let $Y_1, \ldots, Y_n$ be n independent random variables representing n response variables, where $Y_i \sim \text{BGEV}(\cdot;\xi, m_i, \sigma, \delta)$. The model is obtained by assuming that the median of Y_i can be written as

$$g(m_i) = \eta_i = \sum_{j=1}^{p} x_{ij}\beta_j = \boldsymbol{X}_i^\top \boldsymbol{\beta}, \tag{12}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^p$ is a vector of unknown regression parameters, $\boldsymbol{X}_i = (x_{i1}, \ldots, x_{ip})^\top$ is a vector of covariates of length p ($p < n$), which are assumed fixed and known. In addition, $g(\cdot)$ is a strictly monotonic and twice differentiable link function. There are several possible choices for the link function $g(\cdot)$. For instance, one can use the identity, inverse, and log link function defined as $g(m) = m$, $g(m) = 1/m$, and $g(m) = \log(m)$, respectively. In the BGEV regression model, the median is the only

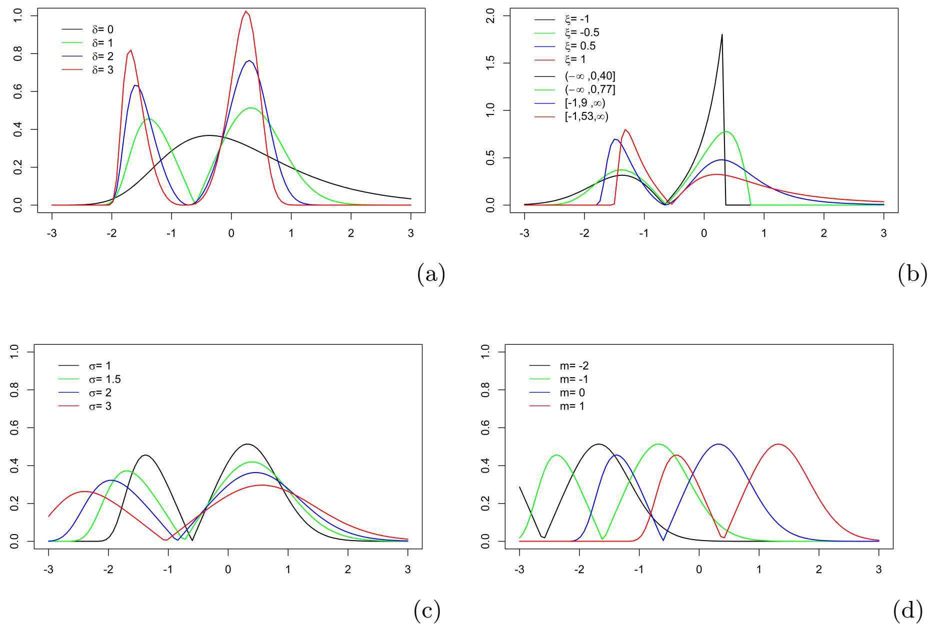


Fig. 1 Density of $X \sim \mathrm{BGEV}(\xi, m, \sigma, \delta)$ varying the parameters as panel legend: (a) $\xi = 0.5, m = 0, \sigma = 1$; (b) $m = 0, \sigma = 1, \delta = 1$; (c) $\xi = 0.5, m = 0, \delta = 1$; (d) $\xi = 0.5, \sigma = 1, \delta = 1$.

parameter allowed to vary with the observations. Therefore, the parameters ξ, σ, and δ are unknown constants since we do not assign regression structures to them.

2.3 Inference, diagnostics and model selection

The vector of unknown parameters for the BGEV regression model is represented as $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \xi, \sigma, \delta)^\top \in \mathbb{R}^{p+3}$ when $\xi \neq 0$ and as $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma, \delta)^\top \in \mathbb{R}^{p+2}$ when $\xi = 0$. Therefore, $\boldsymbol{\theta}$ must be estimated. Let k represent the number of parameters in the model. The inference procedure adopted here will be based on maximum likelihood estimation. The log-likelihood function based on a sample of n independent observations $y_1, \ldots, y_n$ is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}), \tag{13}$$

where $\ell_i(\boldsymbol{\theta}) = \log[f^*(y_i; \xi, m_i, \sigma, \delta)]$ with $f^*(\cdot; \xi, m_i, \sigma, \delta)$ denoting the PDF of reparameterized BGEV distribution. The maximum likelihood estimator (MLE) is obtained by maximizing the log-likelihood function in (13) with respect to $\boldsymbol{\theta}$. The MLE for $\boldsymbol{\theta}$ is denoted by $\widehat{\boldsymbol{\theta}}$.

The contribution of the i-th observation in the log-likelihood function is given by

$$\ell_i(\boldsymbol{\theta}) = \begin{cases} -\log(\sigma) + \log(\delta+1) + \delta \log|y_i - \mu_i^*| - \left(1+\frac{1}{\xi}\right)\log(\Psi_i(\boldsymbol{\theta})) - \Psi_i^{-1/\xi}(\boldsymbol{\theta}), \ \xi \neq 0, \\ \\ -\log(\sigma) + \log(\delta+1) + \delta \log|y_i - \mu_i^*| - \dfrac{T_i^*}{\sigma} - \mathrm{e}^{-T_i^*/\sigma}, \qquad \xi = 0, \end{cases}$$

where

$$\Psi_i(\boldsymbol{\theta}) = 1 + \frac{\xi T_i^*}{\sigma}, \quad T_i^* = (y_i - \mu_i^*)\,|y_i - \mu_i^*|^{\delta}.$$

The score function, obtained by differentiating the log-likelihood function given in (13) with respect to $\boldsymbol{\theta}$, is given by

$$\boldsymbol{U}(\boldsymbol{\theta}) = \begin{cases} (\boldsymbol{U}_{\boldsymbol{\beta}}^\top, U_\xi, U_\sigma, U_\delta)^\top, & \text{if } \ \xi \neq 0, \\ (\boldsymbol{U}_{\boldsymbol{\beta}}^\top, U_\sigma, U_\delta)^\top, & \text{if } \ \xi = 0, \end{cases}$$

with

$$\begin{aligned} \boldsymbol{U}_{\boldsymbol{\beta}} &= X^\top H \boldsymbol{a}, \\ U_\xi &= \boldsymbol{b}^\top \mathbf{1}_n, \\ U_\sigma &= \boldsymbol{d}^\top \mathbf{1}_n, \\ U_\delta &= \boldsymbol{e}^\top \mathbf{1}_n, \end{aligned}$$

where X is an $n \times p$ matrix whose i-th row is $\boldsymbol{X}_i^\top$, $H = \mathrm{diag}\{h_1, \ldots, h_n\}$ is an $n \times n$ diagonal matrix with $h_i = 1/g'(m_i)$, and $\boldsymbol{a} = (a_1, \ldots, a_n)^\top$, $\boldsymbol{b} = (b_1, \ldots, b_n)^\top$, $\boldsymbol{d} = (d_1, \ldots, d_n)^\top$, $\boldsymbol{e} = (e_1, \ldots, e_n)^\top$, $\mathbf{1}_n = (1, \ldots, 1)^\top$ are n-dimensional vectors with

$$a_i = \frac{1}{\sigma(y_i - \mu_i^*)}[(\delta+1)T_i^*\Omega_i(\boldsymbol{\theta}) - \delta\sigma];$$

$$\begin{aligned} b_i &= \frac{a_i}{\xi(\delta+1)}(m_i - \mu_i^*)\left(1 + \frac{\xi c_2 c_1^{-\xi}}{1 - c_1^{-\xi}}\right) + \frac{1}{\xi^2}\log(\Psi_i(\boldsymbol{\theta}))[1 - \Psi_i^{-1/\xi}(\boldsymbol{\theta})] - \frac{T_i^*\Omega_i(\boldsymbol{\theta})}{\xi\sigma}; \\ d_i &= -\frac{1}{\sigma} + \frac{\delta}{\sigma(\delta+1)}\frac{(m_i - \mu_i^*)}{(y_i - \mu_i^*)} + \frac{1}{\sigma^2}\Omega_i(\boldsymbol{\theta})(y_i - m_i)|y_i - \mu_i^*|^{\delta}; \\ e_i &= \log|y_i - \mu_i^*|\left(1 - \frac{T_i^*\Omega_i(\boldsymbol{\theta})}{\sigma}\right) + \frac{1}{\delta+1}\left[1 + a_i(m_i - \mu_i^*)\log(m_i - \mu_i^*)\right]; \end{aligned}$$

$$\Omega_i(\boldsymbol{\theta}) = \begin{cases} \Psi_i^{-1}(\boldsymbol{\theta})[1 + \xi - \Psi_i^{-1/\xi}(\boldsymbol{\theta})], & \text{if } \ \xi \neq 0, \\ \\ 1 - \mathrm{e}^{-T_i^*/\sigma}, & \text{if } \ \xi = 0. \end{cases}$$

The MLE of $\boldsymbol{\theta}$ is obtained by solving the system $\boldsymbol{U}(\boldsymbol{\theta}) = \mathbf{0}$, where $\mathbf{0}$ represents the null vector in $\mathbb{R}^k$. However, this system cannot be solved analytically, and iterative numerical methods must be employed to obtain an approximate solution. In this work, we use the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method (Nocedal and Wright 1987), available in the R Project for Statistical Computing (R Core Team 2022).

Under the usual regularity conditions for maximum likelihood estimation, when the sample size is large,

$$\widehat{\boldsymbol{\theta}} \overset{\text{a}}{\sim} \mathrm{N}_k(\boldsymbol{\theta}, J^{-1}(\boldsymbol{\theta})),$$

where $\overset{\text{a}}{\sim}$ denotes approximate distribution, $\mathrm{N}_k(\boldsymbol{\theta}, J^{-1}(\boldsymbol{\theta}))$ denotes the k-dimensional normal distribution with mean $\boldsymbol{\theta}$ and variance–covariance matrix $J^{-1}(\boldsymbol{\theta})$. The matrix $J^{-1}(\boldsymbol{\theta})$ is also the inverse of observed information matrix. In this work, the observed information matrix is obtained numerically.

Consider the interest in testing H_0: $\theta_j = \theta_j^0$ versus H_1: $\theta_j \neq \theta_j^0$, where θ_j^0 is a specific value for the unknown parameter θ_j, $j = 1, \ldots, k$. Under H_0,

$$z_j = \frac{\widehat{\theta}_j - \theta_j^0}{\mathrm{se}(\widehat{\theta}_j)} \overset{\text{a}}{\sim} \mathrm{N}(0, 1),$$

where $\widehat{\theta}_j$ is the MLE of θ_j, and $\mathrm{se}(\widehat{\theta}_j) = \sqrt{J_{jj}}$ is the asymptotic standard error of $\widehat{\theta}_j$, with J_{jj} the j-th element of the diagonal of $J^{-1}(\boldsymbol{\theta})$. The test statistic z_j is the square root of Wald's statistic (Wald 1943) which is widely used in practical applications using regression models. In Section 4, this test will be employed to evaluate the significance of the parameters in the BGEV regression model, with $\theta_j^0 = 0$.

To assess the goodness of fit under BGEV regression models, the residuals introduced by Dunn and Smyth (1996) can be employed. The quantile residual r_i, $i = 1, \ldots, n$, is given by

$$r_i = \begin{cases} \Phi^{-1}(F^*(y_i; \widehat{\xi}, \widehat{m_i}, \widehat{\sigma}, \widehat{\delta}), & \xi \neq 0, \\ \\ \Phi^{-1}(F^*(y_i; 0, \widehat{m_i}, \widehat{\sigma}, \widehat{\delta}), & \xi = 0, \end{cases}$$

where $\Phi^{-1}(\cdot)$ is the CDF of the standard normal distribution. If the BGEV regression model is well-fitted, r_i is approximated distributed as a standard normal distribution. To evaluate the empirical distribution of residuals, a common practice is to utilize normal probability plots of residuals with simulated envelope (Pereira 2019; Espinheira and Silva 2020; Tsuyuguchi et al. 2020). This is the approach we will adopt in Section 4 to assess the goodness of fit from BGEV regression models.

For model selection, we will utilize the Akaike Information Criterion (AIC) (Akaike 1974), defined as $\mathrm{AIC} = -2\ell(\widehat{\boldsymbol{\theta}}) + 2k$. Given a set of candidate models for the data, the preferred BGEV regression model is the one that has the lowest AIC value.

3 Monte Carlo simulation results

To evaluate the performance of the MLE under BGEV regression model, Monte Carlo simulations were carried out. The sample sizes considered are $n = 50, 100, 500$. For all the scenarios we employ identity link function in the median regression structure (12). The models include intercepts, i.e. $x_{i1} = 1$, for all $i = 1, 2, \ldots, n$, and the other covariates are taken as random draws from a standard uniform distribution and from Bernoulli distribution with probability of success 0.4. The covariates were kept constant throughout the simulated samples. All the results are based on 10000 Monte Carlo replications and were carried out using the R software environment (R Core Team 2022).

Through 10000 Monte Carlo replicates, we obtain for each θ_t, $t = 1, \ldots, k$, the parameter estimate and calculate, from their simulated sample distribution, the bias $\mathrm{B}(\widehat{\theta}_t)$, and the square root of the mean square error $\mathrm{RMSE}(\widehat{\theta}_t)$, defined, respectively, by

$$\mathrm{B}(\widehat{\theta}_t) = \frac{1}{10000}\sum_{j=1}^{10000}\left(\widehat{\theta}_t^j - \theta_t\right) \quad \text{and} \quad \mathrm{RMSE}(\widehat{\theta}_t) = \sqrt{\frac{1}{10000}\sum_{j=1}^{10000}\left(\widehat{\theta}_t^j - \theta_t\right)^2},$$

where $\widehat{\theta}_t^j$ is the MLE of the t-th element of the parameter vector $\boldsymbol{\theta}$ in the j-th Monte Carlo replicate, with $j \in \{1, \ldots, 10000\}$. Furthermore, we also obtain the mean of the asymptotic standard error (ASE) for each θ_t defined by

$$\mathrm{ASE}(\widehat{\theta}_t) = \frac{1}{10000}\sum_{j=1}^{10000}\mathrm{se}(\widehat{\theta}_t^j),$$

where $\mathrm{se}(\widehat{\theta}_t^j)$ is the asymptotic standard error of the t-th element of $\boldsymbol{\theta}$ in the j-th replicate.

For the parameters, four different configurations were considered, two with $\xi = 0$ and two with $\xi \neq 0$, as described below.

Scenario 1: BGEV regression with $\xi = 0$.

The parameter values were $\beta_1 = 1.2, \beta_2 = 2.0, \beta_3 = 1.5$, which yield $m_i \in (1.20, 4.68)$ with median$(m) \approx 2.8$, $\xi = 0$, $\sigma = 1.0$, and $\delta = 0.5$.

Scenario 2: BGEV regression with $\xi = 0$.

The parameter values were $\beta_1 = 4.0, \beta_2 = 2.0, \beta_3 = 3.0$, which yield $m_i \in (4.00, 8.99)$ with median$(m) \approx 5.6$, $\xi = 0$, $\sigma = 1.0$, and $\delta = 1.0$.

Scenario 3: BGEV regression with $\xi > 0$.

The parameter values were $\beta_1 = 1.2, \beta_2 = 2.0, \beta_3 = 1.5$, which yield $m_i \in (-4.65, 7.77)$ with median$(m) \approx 2.3$, $\xi = 0.25$, $\sigma = 1.0$, and $\delta = 0.5$.

Scenario 4: BGEV regression with $\xi < 0$.

The parameter values were $\beta_1 = 4.0, \beta_2 = 2.0, \beta_3 = 3.0$, which yield $m_i \in (4.00, 8.99)$ with median$(m) \approx 5.6$, $\xi = -0.25$, $\sigma = 1.0$, and $\delta = 1.0$.

Note that the primary difference between Scenarios 1 and 3 is that the first scenario considers $\xi = 0$, while the third scenario accounts for a positive value of ξ. Additionally, Scenario 1 only considers positive medians, whereas Scenario 3 includes both positive and negative medians. Similarly, the difference between Scenarios 2 and 4 is that the second scenario considers $\xi = 0$, while the fourth scenario considers a negative value of ξ.

After computing the MLE of θ_t^j, we obtain the bias, RMSE, and ASE for different sample sizes under scenarios 1-4. In Table 1, these results showed that, in most cases, the bias, RMSE, and ASE of the MLE for the parameters of the BGEV regression models decrease as the sample size n increases. These results indicate the consistency and asymptotic efficiency of the maximum likelihood estimators for all the scenarios considered. We also note that the bias, RMSE, and ASE of the MLE for σ and δ are slightly lower when $\xi = 0$ (Scenarios 1 and 2) compared to when $\xi \neq 0$ (Scenarios 3 and 4). On the other hand, we do not identify a clear pattern in the bias, RMSE, and ASE of the MLE for β_1, β_2, and β_3 when comparing the scenarios with $\xi = 0$ and $\xi \neq 0$.

4 Application

In this section, to demonstrate the applicability of the model (10), we use extreme weather data from the capital of the state of Goiás, Goiânia. This region is located 209 km from Brasília, the capital of Brazil. The data from this region were chosen to illustrate the applicability of the BGEV regression model since Goiás contains one of the most devastated biomes in Brazil, the cerrado vegetation.

Correctly modeling this data can help us monitor biomes, and agricultural production, and study climate change — topics that currently occupy a prominent place in the mainstream media.

Of all the extreme weather events that occur in the world, drought is the most punishing in the Goiânia region. According to Merladete (2024), de Agricultura (2024), and Pereira (2024), the prolonged forecast in this region has caused serious impacts on agricultural production in 2024.The La Niña phenomenon, which intensified in 2024, worsened the lack of rainfall and increased temperatures, affecting the cultivation of soybeans, corn, and vegetables, among other products. Livestock also suffers from the lack of pasture and the need for supplementary feed, which increases production costs. For these reasons, in this work, we chose climate variables that best explain the drought.

The data used here correspond to the period from January 1, 2011 to December 31, 2022 and come from the automatic weather station A002 in Goiânia. They are made available by the National Institute of Meteorology (INMET) on the website https://portal.inmet.gov.br/. From the total of 19 climate variables in the database, we selected four available variables to explain the relationship between climate and drought. These variables are: dew point temperature, measured in degrees Celsius (C$^{\text{o}}$); relative humidity, calculated by the percentage of water vapor in the atmosphere; atmospheric pressure, measured in millibars (mb); wind speed at maximum gust, measured in meters per second.

Among the four climate variables mentioned above, the minimum temperature of the dew point is the one that best explains the relationship between climate and drought (Changnon et al. 2003). The temperature of the dew point is closely related

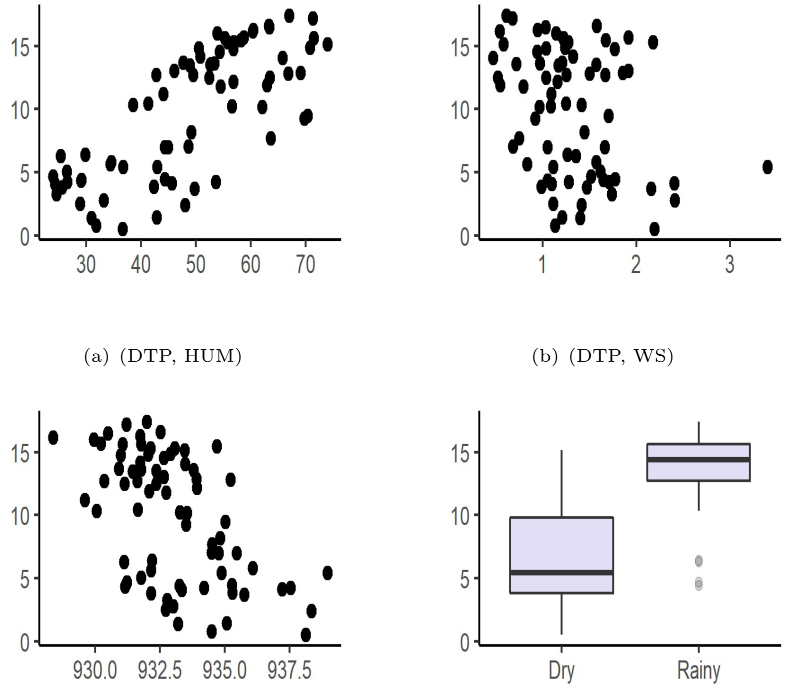
Table 1 Bias, RMSE and ASE under Scenarios 1-4.

	Scenario 1								
	$n = 50$			$n = 100$			$n = 500$		
parameters	Bias	RMSE	ASE	Bias	RMSE	ASE	Bias	RMSE	ASE
β_1	−0.048	0.273	0.136	−0.024	0.192	0.090	−0.029	0.098	0.035
β_2	0.047	0.376	0.184	0.055	0.283	0.126	0.044	0.147	0.046
β_3	0.038	0.236	0.119	0.036	0.183	0.081	0.018	0.086	0.028
σ	−0.025	0.137	0.117	−0.005	0.092	0.082	0.000	0.040	0.036
δ	0.047	0.219	0.177	0.015	0.143	0.122	−0.004	0.060	0.053
	Scenario 2								
	$n = 50$			$n = 100$			$n = 500$		
parameters	Bias	RMSE	ASE	Bias	RMSE	ASE	Bias	RMSE	ASE
β_1	−0.035	0.225	0.116	−0.019	0.147	0.083	−0.007	0.060	0.036
β_2	0.033	0.288	0.151	0.032	0.211	0.115	0.010	0.082	0.048
β_3	0.022	0.176	0.098	0.015	0.124	0.070	0.003	0.050	0.029
σ	−0.021	0.136	0.118	−0.007	0.089	0.083	−0.001	0.038	0.036
δ	0.048	0.315	0.235	0.023	0.198	0.163	0.003	0.078	0.071
	Scenario 3								
	$n = 50$			$n = 100$			$n = 500$		
parameters	Bias	RMSE	ASE	Bias	RMSE	ASE	Bias	RMSE	ASE
β_1	−0.066	0.239	0.095	−0.044	0.185	0.066	0.013	0.079	0.027
β_2	0.003	0.121	0.049	0.006	0.108	0.041	0.001	0.042	0.013
β_3	0.018	0.274	0.122	0.034	0.185	0.074	0.022	0.089	0.027
ξ	−0.086	0.236	0.151	−0.033	0.130	0.087	−0.001	0.043	0.030
σ	−0.032	0.170	0.123	−0.014	0.103	0.085	−0.001	0.039	0.036
δ	0.117	0.311	0.234	0.028	0.182	0.148	−0.003	0.073	0.061
	Scenario 4								
	$n = 50$			$n = 100$			$n = 500$		
parameters	Bias	RMSE	ASE	Bias	RMSE	ASE	Bias	RMSE	ASE
β_1	−0.060	0.241	0.119	−0.050	0.165	0.084	−0.036	0.070	0.037
β_2	0.025	0.309	0.144	0.030	0.215	0.113	0.013	0.079	0.047
β_3	0.017	0.188	0.094	0.013	0.120	0.068	0.004	0.045	0.027
ξ	−0.053	0.264	0.185	−0.021	0.140	0.112	−0.003	0.052	0.044
σ	−0.039	0.161	0.126	−0.016	0.104	0.086	−0.002	0.043	0.038
δ	0.115	0.452	0.299	0.048	0.256	0.195	0.008	0.090	0.082

to the relative humidity of the air and atmospheric pressure (Talaia and Vigário 2016; Costa Junior 2011).

Thus, using the BGEV regression model, the minimum dew point temperature (DPT) is the response variable considered to explain the drought, and the other variables as minimum humidity (HUM), maximum wind speed (WS), maximum pressure

(P), stations (S) are covariates. Figure 2 shows a strong dependence between the DTP variable and the HUM, WS, P and S variables.



(a) (DTP, HUM)

(b) (DTP, WS)

(c) (DTP, P)

(d) (DTP, S)

Fig. 2 Scatter plots of the variable DTP with the variables HUM, WS, P, and S, in panels (a)–(d), respectively.

Since the data exhibit seasonality possibly related to the two seasons of the year, we consider the covariate S, which is a binary variable. Here, S = 1 corresponds to the rainy season from October to April, while S = 0 corresponds to the dry season from May to September.

In the next subsection, we consider the minimum values of the dew point temperature as being independent and identically distributed according to the BGEV distribution with a common CDF $F(\cdot; \xi, \mu, \sigma, \delta)$. That is, the variable DTP as not influenced by $\mathrm{HUM}, \mathrm{P}, \mathrm{WS}, \mathrm{S}$.

4.1 Adjusting the dew point temperature minimum

The original data of DTP exhibit temporal dependence, so we first applied the minimum block technique (Jondeau et al. 2007) to obtain a subsample of independent minimum values. At a 5% level, the Ljung-Box test (Trapletti 2016) verified the serial independence of the subsample of minima for blocks of size $N = 1440$ hours (60 days).

In Figure 3, the histograms of the original data and the DTP minima are presented. The right panel shows that the histogram of the DTP minima exhibits bimodality. Therefore, we compared the fit of the GEV and BGEV distributions for the minimum DTP data.

The maximum likelihood estimators (MLE) for the parameters of the GEV and BGEV distributions were obtained using the evd (Stephenson 2002) and bgev (Otiniano et al. 2024) packages, as presented in R Core Team (2022). The MLE estimates of the parameters for each model, along with their corresponding standard errors, are shown in Table 2. We observe that the estimate of δ is 0.86, indicating that there is inherent bimodality in this data. The parameter $\delta > 0$ in the BGEV distribution is associated with the presence of bimodality, as previously noted in Subsection 2.1.

Table 2 Estimates and standard error under GEV and BGEV distributions for DTP data.

	BGEV		GEV	
Parameters	Estimates	Std. error	Estimates	Std. error
ξ	-0.38	0.09	-0.65	8.94
μ	8.72	0.23	0.90	0.72
σ	25.10	9.49	5.70	0.63
δ	0.86	0.20	-	-

To illustrate the goodness of fit of the minimum DTP data using the BGEV and GEV distributions, consider the right panel of Figure 3, which displays the histogram of the minimum DTP data alongside the adjusted density functions for both the GEV and BGEV distributions. It is evident that the fit with the BGEV distribution, characterized by a bimodal curve, is more suitable than the unimodal fit provided by the GEV distribution. This suggests that the BGEV model provides a superior fit compared to the GEV model.

4.2 BGEV regression for minimum dew point temperature

As shown in Subsection 4.1, the BGEV distribution provided an adequate fit to the minimum values of the DTP. In this Subsection, we utilize the BGEV regression model, as proposed in Subsection 2.2, to report how the median of the response variable DTP is affected by the covariates HUM, P, WS, S.

Considering the BGEV regression model, we estimate the vector of parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \xi, \sigma, \delta)^\top$, where $\boldsymbol{\beta}$ is the vector of regression coefficients corresponding to the covariates HUM, P, WS, and S, for the n independent observations of the response variable DTP, that is, $\text{DTP}_i \sim \text{BGEV}(\xi, m_i, \sigma, \delta)$.

To select the covariates for the BGEV regression model, we employed a forward selection method based on the AIC. The initial BGEV model started with a single covariate that had the lowest AIC value. We then added each covariate that improved

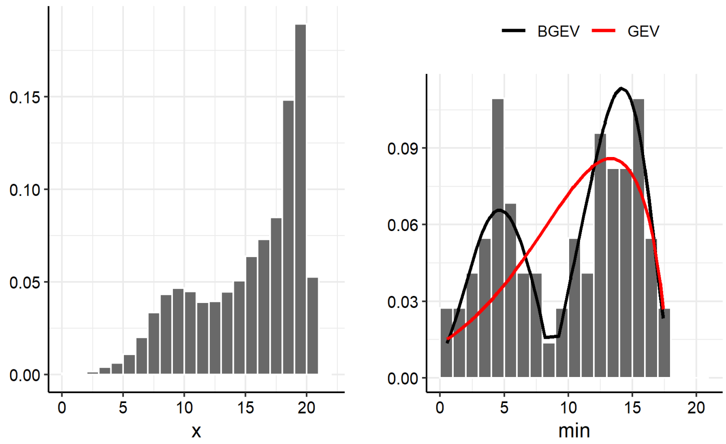


Fig. 3 Histogram left (original DTP data) and the histogram right (DTP minima). In the right panel, the curves represent the fitted densities GEV (red) and BGEV (black) distributions.

the model by further reducing the AIC value. The BGEV models under consideration have the following specifications:

- Specification 1: $m_i = \beta_1 + \beta_2 \text{HUM}_i^*$,
- Specification 2: $m_i = \beta_1 + \beta_2 \text{HUM}_i^* + \beta_3 \text{S}_i$,
- Specification 3: $m_i = \beta_1 + \beta_2 \text{HUM}_i^* + \beta_3 \text{S}_i + \beta_4 \text{P}_i^*$,
- Specification 4: $m_i = \beta_1 + \beta_2 \text{HUM}_i^* + \beta_3 \text{S}_i + \beta_4 \text{P}_i^* + \beta_5 \text{WS}_i^*$,

where $i = 1, \ldots, 73$, m_i is the i-th median of the response variable DTP, $\text{HUM}_i^* = \text{HUM}_i - \overline{\text{HUM}}$, $\text{P}_i^* = \text{P}_i - \overline{\text{P}}$, $\text{WS}_i^* = \text{WS}_i - \overline{\text{WS}}$ denote the i-th values of minimum humidity, maximum pressure, maximum wind speed, respectively, centered at their respective sample means, and S_i is the i-th season. The link function $g(\cdot)$ in (12) is set as the identity link function.

We fitted BGEV regression models with $\xi = 0$ and $\xi \neq 0$ for each specification using the maximum likelihood estimation procedure described in Subsection 2.3. Table 3 shows the AIC values for the fitted BGEV regression models for DTP data. According to the AIC results, when $\xi = 0$, the preferred BGEV regression model incorporates all four covariates (Specification 4), suggesting that all considered covariates are relevant for explaining the DTP median. On the other hand, when $\xi \neq 0$, the preferred BGEV model corresponds to Specification 3, which omits maximum wind speed.

To assess the goodness of fit, we consider normal probability plots of the quantile residuals defined in Subsection 2.3 with simulated envelopes for the fitted BGEV regression models, as shown in Figure 4. In the fitted BGEV models with $\xi = 0$, the residual plots for Specifications 1, 2, and 3 reveal some lack of fit, as several points fall outside the envelope. Differently, the residual plot for Specification 4 suggests a better fit for the data. Now, from the residual plots for BGEV models with $\xi \neq 0$ we observe an improvement in the fitted BGEV models for Specifications 1-3, illustrating that the inclusion of the parameter ξ typically results in a better fit to the data. However, under Specification 4, we do not observe an improvement in the goodness of

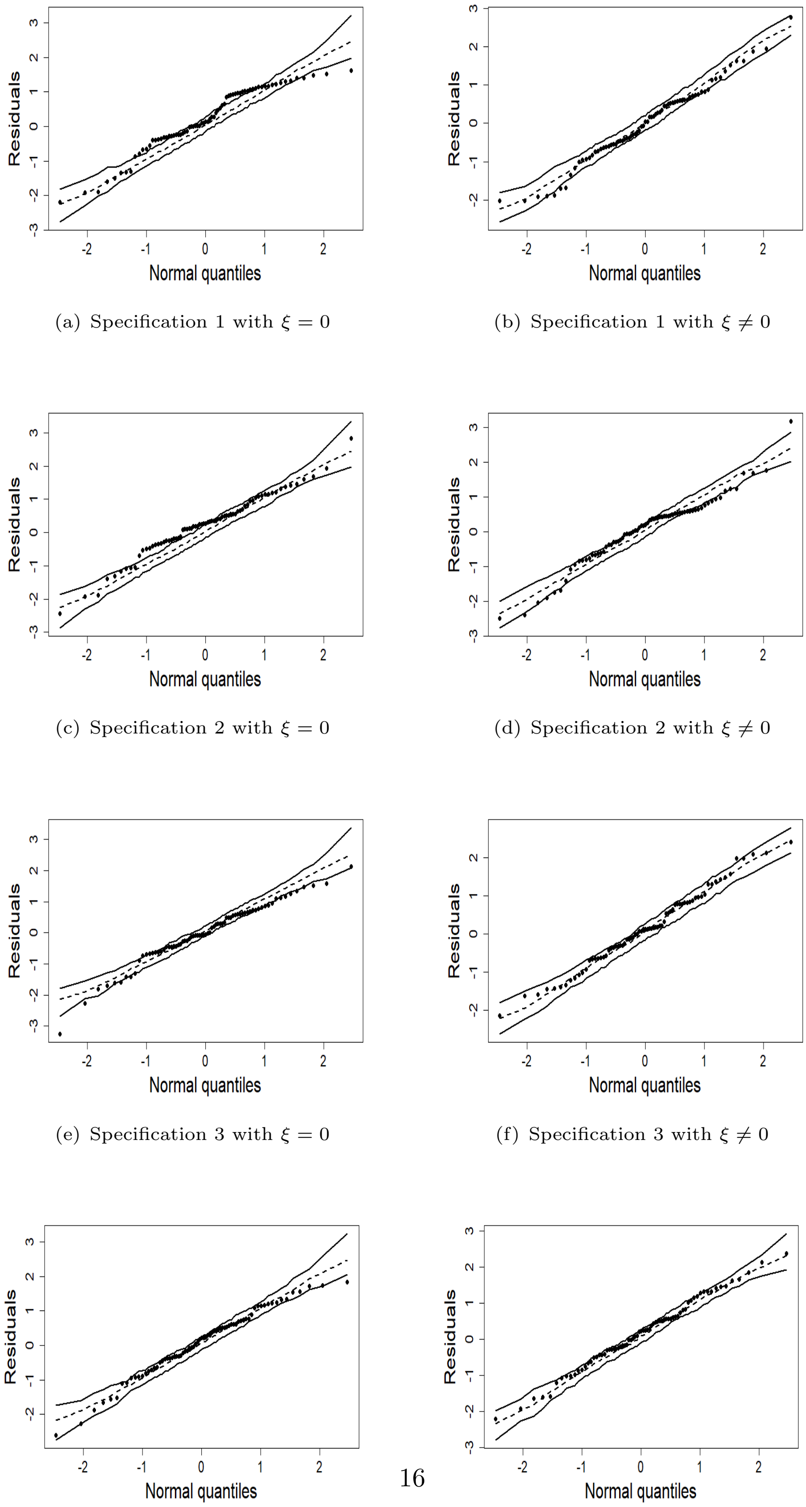
Table 3 AIC values for the fitted BGEV regression models for DTP data.

Specification	Model	AIC
1	$\xi = 0$	392.6367
	$\xi \neq 0$	371.7461
2	$\xi = 0$	316.7746
	$\xi \neq 0$	314.8961
3	$\xi = 0$	300.3726
	$\xi \neq 0$	289.2920
4	$\xi = 0$	298.3478
	$\xi \neq 0$	302.8041

fit with the additional parameter ξ. These results obtained from the residual plots are consistent with the AIC values presented in Table 3. Specifically, with the exception of Specification 4, the AIC value is lower when the model includes ξ. Besides, when comparing only the fitted models with $\xi \neq 0$, the residual plots for Specifications 3 and 4 suggest a good fit for the data, with Specification 3 showing a slightly better fit. This result also agrees with the AIC values.

To decide our final BGEV regression model for DTP data, we further examine the results presented in Table 4 that shows the estimates, standard errors, z-statistics, and p-values of the test of nullity of coefficients for the BGEV regression models. First, the fitted BGEV regression models indicate that the coefficients are statistically significant at a nominal level of 5%. This suggests that the covariates are relevant for explaining the median of DTP. Second, the p-value associated with the null hypothesis $H_0 : \xi = 0$ is 0.962 under the fitted BGEV model with Specification 4. This indicates that when the covariate WS* is included, the model with $\xi = 0$ is adequate. Third, the BGEV regression model with $\xi \neq 0$ and Specification 3 has the lowest AIC value. Fourth, the residual plots for the BGEV regression models with Specification 4, where $\xi = 0$, and Specification 3, where $\xi \neq 0$, which present better fit for the data, exhibit similar behavior. Hence, based on the analysis of AIC values, residual plots, and p-values, our final fitted BGEV model is the one with Specification 3 and $\xi \neq 0$.

Given the final BGEV regression model, we have the following conclusions. The DTP median increases by 3.262° Celsius during the rainy season (October to April) compared to the dry season (May to September). The DTP median estimate is 8.672° Celsius during the dry season, with a minimum humidity of 48.94% and a maximum pressure of 933.08 millibars (maintained at their respective sample means). We estimate that a 1% increase in minimum humidity results in a 0.234° Celsius increase in the DTP median. On the other hand, an increase of 1 millibar in maximum pressure results in a decrease of 0.423° Celsius in the DTP median.



(a) Specification 1 with $\xi = 0$

(b) Specification 1 with $\xi \neq 0$

(c) Specification 2 with $\xi = 0$

(d) Specification 2 with $\xi \neq 0$

(e) Specification 3 with $\xi = 0$

(f) Specification 3 with $\xi \neq 0$

(g) Specification 4 with $\xi = 0$

(h) Specification 4 with $\xi \neq 0$

Fig. 4 Normal probability plots of the quantile residuals with simulated envelope under BGEV regression models for DTP.

Table 4 Estimates, standard errors, z-stat and p-values for BGEV regression models for DTP data.

	Specification 1							
	$\xi = 0$				$\xi \neq 0$			
	Estimate	Std. error	z-stat	p-value	Estimate	Std. error	z-stat	p-value
parameters								
β_1	9.399	0.244	38.423	< 0.001	11.091	0.288	38.482	< 0.001
β_2	0.255	0.010	25.239	< 0.001	0.331	0.011	29.516	< 0.001
σ	8.275	2.243	-	-	6.796	1.690	-	-
δ	0.598	0.163	-	-	0.388	0.153	-	-
ξ	-	-	-	-	0.579	0.072	8.003	< 0.001
	Specification 2							
	$\xi = 0$				$\xi \neq 0$			
	Estimate	Std. error	z-stat	p-value	Estimate	Std. error	z-stat	p-value
parameters								
β_1	7.454	0.213	35.006	< 0.001	7.619	0.217	35.180	< 0.001
β_2	0.260	0.008	17.554	< 0.001	0.210	0.008	25.830	< 0.001
β_3	4.697	0.268	34.171	< 0.001	5.103	0.241	21.194	< 0.001
σ	6.065	1.562	-	-	3.891	0.836	-	-
δ	0.918	0.194	-	-	0.530	0.177	-	-
ξ	-	-	-	-	0.192	0.067	2.877	0.004
	Specification 3							
	$\xi = 0$				$\xi \neq 0$			
	Estimate	Std. error	z-stat	p-value	Estimate	Std. error	z-stat	p-value
parameters								
β_1	8.309	0.184	45.093	< 0.001	8.672	0.181	47.814	< 0.001
β_2	0.213	0.007	29.226	< 0.001	0.234	0.006	36.500	< 0.001
β_3	3.396	0.250	13.531	< 0.001	3.262	0.175	18.667	< 0.001
β_4	-0.737	0.056	-13.207	< 0.001	-0.423	0.036	-11.766	< 0.001
σ	2.606	0.433	-	-	3.237	0.613	-	-
δ	0.465	0.148	-	-	0.442	0.182	-	-
ξ	-	-	-	-	0.504	0.114	4.435	< 0.001
	Specification 4							
	$\xi = 0$				$\xi \neq 0$			
	Estimate	Std. error	z-stat	p-value	Estimate	Std. error	z-stat	p-value
parameters								
β_1	7.978	0.171	46.551	< 0.001	7.904	0.194	40.679	< 0.001
β_2	0.239	0.006	38.843	< 0.001	0.249	0.007	31.091	< 0.001
β_3	3.795	0.276	13.751	< 0.001	3.475	0.307	11.305	< 0.001
β_4	-0.690	0.050	-13.758	< 0.001	-0.954	0.068	-13.864	< 0.001
β_5	0.456	0.087	5.254	< 0.001	0.412	0.119	3.448	< 0.001
σ	2.450	0.380	-	-	2.154	0.341	-	-
δ	0.486	0.148	-	-	0.380	0.192	-	-
ξ	-	-	-	-	-0.006	0.144	-0.048	0.962

5 Conclusion

The GEV distribution is frequently used for modeling extreme data. However, it is not well-suited for datasets that exhibit bimodal behavior. To effectively address datasets that display both unimodal and bimodal characteristics, the BGEV distribution, initially proposed by Otiniano et al. (2023) and later redefined by Otiniano et al. (2024), can be employed.

In this work, we introduce a new class of regression models that utilize a new parameterization of the BGEV distribution, where one of the parameters represents the median. This approach allows for a clear interpretation of the relationships between the response variable and the covariates. The proposed model effectively accommodates both unimodal and bimodal extreme data encountered in various fields.

In short, the main contributions of this paper are as follows. First, it presents a new parameterization of the BGEV distribution that facilitates the interpretability of the parameters in regression models. Second, it proposes a new flexible class of regression models based on the reparameterized BGEV distribution, which offers a simple interpretation in terms of the median of the response variable. These models can be utilized to measure the relationship between variables and to predict new values of the response variable. Third, it presents a method for parameter estimation under BGEV regression models using maximum likelihood estimation, including the derivation of score functions, hypothesis tests, and quantile residuals. Fourth, it offers simulation results to evaluate the performance of the maximum likelihood estimators. Fifth, it illustrates the application of the BGEV regression model to minimum dew point temperature data, accompanied by a detailed analysis of its adequacy and interpretability. To ensure the applicability of the BGEV regression model, we initially employed the maximum block technique, ensuring that the data are independent.

As part of future research, we also plan to explore other estimation methods for the BGEV regression model beyond maximum likelihood, including robust and Bayesian approaches. Another natural extension of the current work is the development of BGEV regression models that assign regression structures to all four parameters of the BGEV distribution: ξ, m, σ, and δ. This approach will improve the fit of the BGEV model to the datasets.

Funding The authors of this paper would like to acknowledge the partial support of the University of Brasília, FAPDF, and FINATEC.

Declarations

Conflict of interest The authors declare that there are no conflict of interest regarding the publication of this article.

Ethical approval This research does not involve human or animal participants.

References

Akaike, H.: A new look at the statistical model identification. IEEE Transactions on Automatic Control **19**, 716–723 (1974)

Costa Junior, J.S.: Sensor de umidade microeletrônico fabricado em substrato de alumina e baseado na deteção do ponto de orvalho. Master's thesis, State University of Campinas. Faculty of Electrical and Computer Engineering. (2011). https://doi.org/10.47749/T/UNICAMP.2011.789974

Changnon, D., Sandstrom, M., Schaffer, C.: Relating changes in agricultural practices to increasing dew points in extreme chicago heat waves. Climate Research **24**(3), 243–254 (2003)

Agricultura, S.N.: Estiagem Causa Impactos na Produção e Nos Preços Dos Alimentos. Access in: 26 September 2024. https://sna.agr.br/estiagem-causa-impactos-na-producao-e-nos-precos-dos-alimentos/

Dunn, P.K., Smyth, G.K.: Randomized quantile residuals. Journal of Computational and Graphical Statistics **5**(3), 236–244 (1996)

Espinheira, P.L., Silva, A.O.: Residual and influence analysis to a general class of simplex regression. TEST **29**, 523–552 (2020)

Fisher, R.A., Tippett, L.H.C.: Limiting forms of the frequency distribution of the largest and smallest member of a sample. Mathematical Proceedings of the Cambridge Philosophical Society **24**(2), 180–190 (1928)

Gnedenko, B.V.: Sur la distribution limite du terme maximum d'une série aléatoire. Annals of Mathematics **44**(3), 423–453 (1943)

Jenkinson, A.F.: The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quarterly Journal of the Royal Meteorological Society **81**(348), 158–171 (1955)

Jondeau, E., Poon, S.H., Rockinger, M.: Financial Modeling Under Non-Gaussian Distributions. Springer, New York (2007)

Kharin, V.V., Zwiers, F.W.: Estimating extremes in transient climate change simulations. Journal of Climate **17**(8), 1156–1173 (2004)

Merladete, A.: Impacto da Estiagem em 2024 Agrava Crise No Campo. Acess in: 23 September 2024. https://www.agrolink.com.br/noticias/impacto-da-estiagem-em-2024-agrava-crise-no-campo_495186.html

Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, New York (1987)

Otiniano, C.G., Oliveira, Y.S., Maluf, Y.S.: Probability distribution of extreme events in complex systems: Application to climate data. Symmetry **16**, 1639 (2024)

Otiniano, C.E.G., Oliveira, Y.L.S., Sousa, T.R.: Bimodal GEV Distribution with Location Parameter. https://CRAN.R-project.org/package=bgev (2024)

Otiniano, C.E.G., Paiva, B.S., Vila, R., Bourguignon, M.: A bimodal model for extremes data. Environmental and Ecological Statistics **30**, 266–288 (2023)

Otiniano, C.G., Silva, E.B., Matsushita, R.Y., Silva, A.: Bridging extremes: The invertible bimodal gumbel distribution. Entropy **25**, 1598 (2023) https://doi.org/10.3390/e25121598

Otiniano, C.E.G., Vila, G.R., Brown, P.C., Pereira, M.B.: On the bimodal gumbel model with application to environmental data. Austrian Journal of Statistics **52**, 45–65 (2023)

Pereira, G.H.: On quantile residuals in beta regression. Communications in Statistics - Simulation and Computation **48**(1), 302–316 (2019) https://doi.org/10.1080/03610918.2017.1381740

Pereira, P.H.: Seca Prolongada Deve Impactar a Produção Agrícola e Elevar Custos. Acess in: 29 April 2025. https://edicaodobrasil.com.br/2024/10/04/seca-prolongada-deve-impactar-a-producao-agricola-e-elevar-custos

Prataviera, F.F., Ortega, E.M.M., Cordeiro, G.M., Pescim, R.R., Verssani, B.A.W.: A new generalized odd log-logistic flexible weibull regression model with applications in repairable systems. Reliability Engineering & System Safety **176**, 13–26 (2018)

R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2022). R Foundation for Statistical Computing. https://www.R-project.org/

Rodrigues, G.M., Ortega, E.M.M., Cordeiro, G.M., Vila, R.: An extended weibull regression for censored data: Application for covid-19 in campinas, brazil. Mathematics **10**(19), 3644 (2022)

Schlatter, T.W.: Temperature-humidity index. In: Climatology. Encyclopedia of Earth Science. Springer, Boston, MA (1987). https://doi.org/10.1007/0-387-30749-4_176

Simões, E.S.B., Otiniano, C.E.G., Nakano, E.Y.: The return period of heterogeneous climate data with a new invertible distribution. Stochastic Environmental Research and Risk Assessment **1**, 1–14 (2024)

Stephenson, A.G.: evd: Extreme value distributions. R News **2**(2), 31–32 (2002)

Tsuyuguchi, A.B., Paula, G.A., Barros, M.: Analysis of correlated birnbaum–saunders data based on estimating equations. TEST **29**, 661–681 (2020)

Trapletti, A.: Box-Pierce and Ljung-Box Tests. R Foundation for Statistical Computing, (2016). R Foundation for Statistical Computing. Documentation from the R statistical environment

Talaia, M., Vigário, C.: Temperatura de ponto de orvalho: um risco ou uma necessidade. Imprensa da Universidade de Coimbra. University Press. https://ucdigitalis.uc.pt/pombalina/item/69146 (2016)

Vasconcelos, J.C.S., Cordeiro, G.M., Ortega, E.M., Ribeiro, J.G.: A regression model for extreme events and the presence of bimodality with application to energy generation data. IET Renewable Power Generation **15**(2), 452–461 (2021)

Vasconcelos, J.C.S., Cordeiro, G.M., Ortega, E.M.M., Rezende, E.M.: A new regression model for bimodal data and applications in agriculture. Journal of Applied Statistics **48**(2), 349–372 (2021)

Vila, R., Ferreira, L., Saulo, H., Prataviera, F., Ortega, E.M.: A bimodal gamma distribution: Properties, regression model and applications. Statistics **54**(3), 469–493 (2020)

Wald, A.: Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society **54**(3), 426–482 (1943)