

## Position Paper

## A novel Weibull-based dynamic model with application to streamflow time series

Débora Missio Bayer <sup>a,b</sup>, Júlia Konrad <sup>c</sup>, Bruna G. Palm <sup>b</sup>, Fábio M. Bayer <sup>b,d</sup><sup>a</sup> Department of Sanitary and Environmental Engineering, Universidade Federal de Santa Maria, Santa Maria, Brazil<sup>b</sup> Department of Mathematics and Natural Sciences, Blekinge Institute of Technology, Karlskrona, Sweden<sup>c</sup> Graduate Program in Civil Engineering, Universidade Federal de Santa Maria, and Fundação Cambirela do Meio Ambiente, Palhoça, Brazil<sup>d</sup> Department of Statistics and LACESM, Universidade Federal de Santa Maria, Santa Maria, Brazil

## ARTICLE INFO

Dataset link: <https://github.com/deborabayer-cpu/WEI-ARMA>

## Keywords:

ARMA models  
Hydrometeorological data  
Time series  
Trend test  
Vacacaí river

## ABSTRACT

Hydrometeorological time series are inherently stochastic and exhibit temporal dependence, commonly modeled using Gaussian autoregressive moving average (ARMA) models. However, the normality assumption is often too restrictive for environmental variables such as streamflow, which are nonnegative and right-skewed. We propose the Wei-ARMA model, a new class of ARMA models based on the Weibull distribution that incorporates ARMA components, external regressors, and a link function. A parametric trend test is also introduced, with parameters estimated via the conditional maximum likelihood method. Monte Carlo simulations assess finite-sample performance. An application to streamflow data from the Vacacaí River, Brazil, shows that the proposed model captures key statistical features, avoids unrealistic negative predictions, and outperforms the Gaussian ARMA. Mean absolute percentage errors in in-sample prediction are reduced by 23%, 34%, and 9% for mean, maximum, and minimum monthly streamflow, respectively, relative to the Gaussian ARMA model. The proposed trend test successfully detects monotonic trends.

## 1. Introduction

Time series found in the natural sciences, such as in hydrology, climatology, and environmental applications, hereafter referred to as hydrometeorological data, usually show temporal dependence, meaning that the observations are correlated over time. In addition, these series are often influenced by random factors, which makes them inherently stochastic (Papalexiou, 2018). Because of these characteristics, stochastic models are particularly useful for analyzing and forecasting such data. A well-known and widely applied class of models for this purpose is the autoregressive moving average (ARMA) models (Box et al., 2008), which are capable of capturing the autocorrelation structure in the data and using it, for example, to make forecasts (Mishra and Desai, 2005).

For inference on their parameters, several stochastic models assume that the underlying variable follows a normal distribution (Salas, 1992). This is the case for classical ARMA models (Box et al., 2008), which are widely applied in environmental sciences and hydrology (Chandler and Scott, 2011; Salas, 1992), and are considered among the main tools for hydrological forecasting (Valipour et al., 2013). However, the normality assumption can be too restrictive in many real-world applications (Li and McLeod, 1988; Tiku et al., 2000; Bayer et al., 2017; Ribeiro et al.,

2026). Many variables in the hydrometeorological field follow non-Gaussian distributions, and this needs to be considered (Lawrance and Kottegod, 1977; Bayer et al., 2025). Variables such as precipitation, streamflow, temperature, and wind speed, for example, are naturally constrained to be nonnegative as they are bounded below by zero. As a result, assuming a Gaussian distribution for these variables, or any other symmetric distribution over the entire real line, may lead to inaccurate inferences and misleading interpretations, since such distributions allow the mean to lie anywhere on the real line, including negative values (Chandler and Scott, 2011). In addition, because of its symmetry, the normal distribution is not suitable for modeling extreme events, which often produce data with skewed histograms (Helsel et al., 2020; Machiwal and Jha, 2012), and require a specific treatment (Davis and Resnick, 1989; D'Arcy and Tawn, 2025).

In this context, several hydrometeorological variables can be more appropriately modeled using the Weibull distribution. Originally proposed by Weibull (1951), this distribution has since been widely applied in different fields, particularly in the modeling and analysis of extreme events (Clarke, 2002; Beirlant et al., 2004; Li et al., 2013; Bardsley, 2019; Langat et al., 2019). The Weibull distribution is defined on the positive real line and is characterized by only two parameters (Tuller and Brett, 1984; Justus et al., 1978), making it both

\* Corresponding author at: Department of Sanitary and Environmental Engineering, Universidade Federal de Santa Maria, Santa Maria, Brazil.  
E-mail address: [debora.bayer@ufsm.br](mailto:debora.bayer@ufsm.br) (D.M. Bayer).

flexible and interpretable. Its shape makes it especially suitable for modeling streamflow data (Moore and Clarke, 1981), whether focusing on average behavior, minima, or maxima (Clarke, 1994; Langat et al., 2019; Li et al., 2018; Stedinger et al., 1992; Mahdi and Ashkar, 2004; Rizwan et al., 2018).

The development of new dynamic models based on alternative probability distributions has become an important tool when the random variable of interest does not follow a normal distribution. This is particularly relevant in the hydrometeorological field, where the impacts of climate change and the growing importance of anticipating extreme events demand more flexible and accurate modeling approaches. Despite these needs, stochastic dynamic models that accommodate a wider range of distributional characteristics remain scarce, highlighting the potential for further advancements in this area. Although the Weibull distribution has been explored in some dynamic modeling proposals (Sarлак, 2008; Duca et al., 2021), to the best of our knowledge, no autoregressive moving average model based on the Weibull distribution has yet been formally developed.

In this work, our objectives are threefold. First, we propose a time series model tailored for non-Gaussian contexts, where the observed hydrometeorological data are continuous, asymmetric, and restricted to positive values. For the proposed Weibull-based autoregressive and moving average (Wei-ARMA) model, we present the parameter estimation procedure, asymptotic properties of the estimators, and adequacy measures of the model. Second, we introduce a statistical test designed to detect trends in non-Gaussian time series. Traditional trend detection methods, such as the Mann-Kendall test and Sen's Slope Estimator, are nonparametric and widely used due to their robustness. However, these methods still assume independence of the observations, which is a strong and often unrealistic assumption in time series data. When the underlying distribution of the data is well-characterized, such as with hydrometeorological variables that follow a positively skewed distribution, a non-Gaussian parametric approach can provide more accurate and powerful inference. By leveraging the flexibility of the Weibull distribution, the proposed test offers a more suitable framework for identifying monotonic trends in nonnegative, asymmetric, and autocorrelated environmental time series. Third, we provide empirical evidence of the effectiveness of the proposed methodology through a Monte Carlo simulation study and an application to real streamflow data. These experiments provide evidence of the adequacy of the theoretical developments, the reliability of the estimation procedures, and the practical advantages of the proposed model in comparison with classical time series approaches.

This article is structured as follows. In Section 2.1, the Wei-ARMA model is introduced. Section 2.2 presents the conditional likelihood inference, along with closed-form expressions for the conditional score vector and the trend test proposed. Section 2.3 discusses some diagnostic measures and the forecasting method. Section 3.1 presents a Monte Carlo simulation study to assess the finite sample performance of the conditional maximum likelihood estimators. Applications of the Wei-ARMA models to streamflow data are presented in Section 3.2. The conclusions can be found in Section 4.

## 2. Method

In this section, we introduce the new model, discuss the inference theory for its parameters, and explore diagnostic and forecasting tools.

### 2.1. Proposed model

Let  $Y$  be a random variable following the two-parameter Weibull distribution with scale parameter  $\varphi > 0$  and shape parameter  $\delta > 0$ . The probability density function (PDF) of the Weibull distribution, denoted by  $W(\varphi, \delta)$ , is given by Stedinger et al. (1992):

$$f(y; \varphi, \delta) = \frac{\delta}{\varphi^\delta} y^{\delta-1} \exp \left[ - \left( \frac{y}{\varphi} \right)^\delta \right],$$

where  $y > 0$  represents the observed value. The cumulative distribution function (CDF), expected value, and variance of a Weibull-distributed variable are given, respectively, by:

$$\begin{aligned} F(y; \varphi, \delta) &= 1 - \exp \left[ - \left( \frac{y}{\varphi} \right)^\delta \right], \\ E(Y) &= \mu = \varphi \Gamma \left( 1 + \frac{1}{\delta} \right), \\ \text{Var}(Y) &= \varphi^2 \left[ \Gamma \left( 1 + \frac{2}{\delta} \right) - \Gamma^2 \left( 1 + \frac{1}{\delta} \right) \right], \end{aligned}$$

where  $\Gamma$  denotes the Gamma function.

To define the proposed model, the PDF of the Weibull distribution is reparametrized in terms of the mean  $\mu$ , using the relationship  $\varphi = \frac{\mu}{\Gamma(1+\frac{1}{\delta})}$ , such that, in this case, we have  $Y \sim W(\mu, \delta)$ . This reparametrization is adopted because, in time series analysis, the mean is typically a parameter of interest and provides more interpretability compared to the original scale and shape parameters.

Let  $\{Y_t\}_{t \in \mathbb{Z}}$  be a stochastic process and  $\mathcal{F}_{t-1}$  be the sigma-field generated by the past observations  $\{\dots, Y_{t-2}, Y_{t-1}\}$ . We assume that each random variable  $Y_t$  is, conditioned to  $\mathcal{F}_{t-1}$ , Weibull distributed, with conditional mean  $\mu_t > 0$  and shape parameter  $\delta > 0$ . Then, the mean-based PDF of  $Y_t | \mathcal{F}_{t-1}$  is defined by:

$$\begin{aligned} f_Y(y_t | \mathcal{F}_{t-1}) &= \frac{\delta}{\mu_t} \Gamma \left( 1 + \frac{1}{\delta} \right) \left[ \frac{y_t}{\mu_t} \Gamma \left( 1 + \frac{1}{\delta} \right) \right]^{\delta-1} \\ &\quad \times \exp \left\{ - \left[ \frac{y_t}{\mu_t} \Gamma \left( 1 + \frac{1}{\delta} \right) \right]^\delta \right\}. \end{aligned} \quad (1)$$

The conditional mean-based CDF, expected value, and variance of  $Y_t$  are given, respectively, by:

$$\begin{aligned} F_Y(y_t | \mathcal{F}_{t-1}) &= 1 - \exp \left\{ - \left[ \frac{y_t}{\mu_t} \Gamma \left( 1 + \frac{1}{\delta} \right) \right]^\delta \right\}, \\ E(Y_t | \mathcal{F}_{t-1}) &= \mu_t, \end{aligned}$$

and

$$\text{Var}(Y_t | \mathcal{F}_{t-1}) = \left[ \frac{\mu_t}{\Gamma \left( 1 + \frac{1}{\delta} \right)} \right]^2 \left[ \Gamma \left( 1 + \frac{2}{\delta} \right) - \Gamma^2 \left( 1 + \frac{1}{\delta} \right) \right]. \quad (2)$$

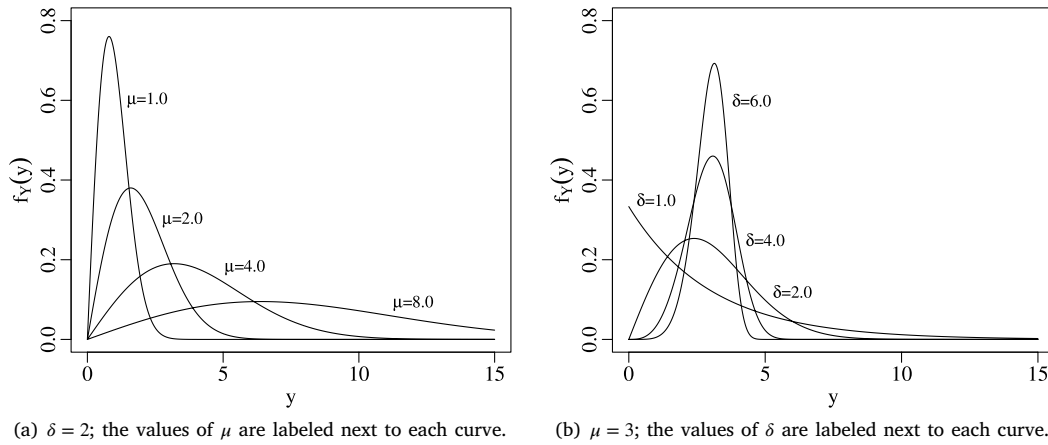
Fig. 1 displays the behavior of the mean-based Weibull PDF under varying values of the parameters  $\mu$  and  $\delta$ . The figure demonstrates that the distribution is flexible, capable of generating shapes that range from nearly symmetric to highly skewed. This adaptability allows the Weibull distribution to closely resemble other commonly used distributions, such as the normal and exponential, depending on the parameter configuration (Nimbal et al., 2012; Tuller and Brett, 1984).

The model is derived following the same approach that was used in GARMA (Benjamin et al., 2003),  $\beta$ ARMA (Rocha and Cribari-Neto, 2009), and KARMA (Bayer et al., 2017) models, but to model the conditional mean of the mean-based Weibull distribution. The linear predictor incorporates autoregressive (AR) and moving average (MA) components, as well as a set of covariates, such as seasonality, trend, or any other relevant time series, allowing for flexible modeling of complex dynamics. A link function  $g : (0, \infty) \rightarrow \mathbb{R}$  is used to map the conditional mean to the linear predictor. This function is assumed to be continuous, monotonic, and twice differentiable, with a continuous and twice differentiable inverse  $g^{-1} : \mathbb{R} \rightarrow (0, \infty)$ . In this work, the logarithmic function was employed as the link function.

Accordingly, the conditional mean of the Wei-ARMA( $p, q$ ) model is specified through the following dynamic structure:

$$\eta_t = g(\mu_t) = \zeta + \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^p \phi_i [g(y_{t-i}) - \mathbf{x}_{t-i}^\top \boldsymbol{\beta}] + \sum_{j=1}^q \theta_j r_{t-j}, \quad (3)$$

where  $\eta_t$  is the linear predictor,  $\zeta$  is the intercept,  $\mathbf{x}_t$  is a vector containing the covariates over time  $t$ ,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)^\top$  is an unknown



**Fig. 1.** The Weibull PDF parametrized in terms of the mean  $\mu$  and the shape parameter  $\delta$ . Panel (a) shows the effect of varying  $\mu$  with fixed  $\delta = 2$ , while panel (b) shows the effect of varying  $\delta$  with fixed  $\mu = 3$ .

parameter vector associated with the covariates,  $\phi = (\phi_1, \dots, \phi_p)^\top$  and  $\theta = (\theta_1, \dots, \theta_q)^\top$  are respectively the autoregressive and moving average parameters,  $p$  and  $q$  are the orders of the model, and  $r_t = g(y_t) - g(\mu_t)$  is the moving average error term. In practical applications, the parameter values are unknown and must be estimated. The inference procedures are discussed in the next section.

### 2.2. Conditional likelihood inference

One of the most widely used methods for parameter estimation is maximum likelihood estimation (MLE) (Chandler and Scott, 2011). MLE is known for its desirable statistical properties, particularly in large samples. In the context of hydrological applications, MLE has shown strong performance (Stedinger et al., 1992) and is also recommended for estimating parameters of the Weibull distribution (Clarke, 1994; Mahdi and Ashkar, 2004). In general, the likelihood function is defined as the product of the individual PDFs. Since the Wei-ARMA model is built on the conditional density function (Eq. (1)), the estimation procedure involves the conditional likelihood function. The conditional maximum likelihood estimation (CMLE) follows the same principle as traditional MLE and shares its asymptotic properties, such as consistency and asymptotic normality.

Given a sample  $y_1, \dots, y_n$  from a Wei-ARMA( $p, q$ ) process, as specified in Eqs. (1) and (3), the model parameter vector of dimension  $(p + q + k + 2)$  is defined as  $\gamma = (\zeta, \beta^\top, \phi^\top, \theta^\top, \delta)^\top$ , where  $k$  denotes the number of covariates included in the model. The CMLE are obtained by maximizing the logarithm of the conditional likelihood function for  $\gamma$ , which, conditional on the  $m = \max(p, q)$  preliminary observations, is given by:

$$\ell(\gamma; y_t) = \sum_{t=m+1}^n \log(f_Y(y_t | \mathcal{F}_{t-1})) = \sum_{t=m+1}^n \ell_t(\mu_t, \delta),$$

where

$$\ell_t(\mu_t, \delta) = \log(\delta) - \delta \log(\mu_t) + \delta \log \Gamma(1 + 1/\delta) + (\delta - 1) \log(y_t) - \left( \frac{y_t \Gamma(1 + 1/\delta)}{\mu_t} \right)^\delta.$$

#### 2.2.1. Conditional score vector

The conditional score vector is obtained by differentiating the conditional log-likelihood function with respect to each one of the parameters in the vector  $\gamma$ . For the  $j$ th elements of  $\gamma$ , where  $j = 1, \dots, (p + q + k + 1)$ , corresponding to the parameters of the dynamic structure ( $\zeta, \beta, \phi$ , and  $\theta$ ), we obtain the following expressions:

$$\frac{\partial \ell}{\partial \gamma_j} = \sum_{t=m+1}^n \frac{\partial \ell_t(\mu_t, \delta)}{\partial \mu_t} \frac{d\mu_t}{d\gamma_j} \frac{\partial \eta_t}{\partial \gamma_j},$$

being

$$\frac{\partial \ell_t(\mu_t, \delta)}{\partial \mu_t} = \frac{-\delta}{\mu_t} + \frac{\delta}{\mu_t} \left( \frac{y_t \Gamma(1 + 1/\delta)}{\mu_t} \right)^\delta = v_t, \tag{4}$$

and

$$\frac{d\mu_t}{d\eta_t} = \frac{1}{g'(\mu_t)},$$

where  $g'(\cdot)$  denotes the first derivative of the link function  $g(\cdot)$ . Therefore, we can write:

$$\frac{\partial \ell}{\partial \gamma_j} = \sum_{t=m+1}^n v_t \frac{1}{g'(\mu_t)} \frac{\partial \eta_t}{\partial \gamma_j}. \tag{5}$$

Note that, in the case of the logarithmic link, we have  $\frac{1}{g'(\mu_t)} = \mu_t$ .

The partial derivatives of  $\eta_t$  to  $\zeta$  and to each element of  $\beta, \phi$  and  $\theta$  are given respectively by:

$$\frac{\partial \eta_t}{\partial \zeta} = 1 - \sum_{j=1}^q \theta_j \frac{\partial \eta_{t-j}}{\partial \zeta},$$

$$\frac{\partial \eta_t}{\partial \beta_l} = x_{t,l} - \sum_{i=1}^p \phi_i x_{(t-i),l} - \sum_{j=1}^q \theta_j \frac{\partial \eta_{t-j}}{\partial \beta_l},$$

$$\frac{\partial \eta_t}{\partial \phi_i} = g(y_{t-i}) - \mathbf{x}_{t-i}^\top \beta - \sum_{j=1}^q \theta_j \frac{\partial \eta_{t-j}}{\partial \phi_i},$$

$$\frac{\partial \eta_t}{\partial \theta_j} = r_{t-j} - \sum_{j=1}^q \theta_j \frac{\partial \eta_{t-j}}{\partial \theta_j},$$

for  $l = 1, \dots, k, i = 1, \dots, p$  and  $j = 1, \dots, q$ .

Finally, the partial derivative of  $\ell_t$  with respect to parameter  $\delta$  is given by:

$$\frac{\partial \ell_t(\mu_t, \delta)}{\partial \delta} = \frac{1}{\delta} - \log(\mu_t) + \log(y_t) + b + c_t, \tag{6}$$

being

$$b = \log \Gamma(1 + 1/\delta) - \frac{1}{\delta} \psi(1 + 1/\delta),$$

and

$$c_t = \frac{\left( \frac{y_t \Gamma(1 + 1/\delta)}{\mu_t} \right)^\delta \left[ \psi(1 + 1/\delta) - \delta \log \left( \frac{y_t \Gamma(1 + 1/\delta)}{\mu_t} \right) \right]}{\delta},$$

where  $\psi$  is the digamma function.

Thus, we obtain the score vector  $\mathbf{U}(\gamma)$  in matrix form such as

$$\mathbf{U}(\gamma) = [\mathbf{U}_\zeta(\gamma), \mathbf{U}_\beta(\gamma)^\top, \mathbf{U}_\phi(\gamma)^\top, \mathbf{U}_\theta(\gamma)^\top, \mathbf{U}_\delta(\gamma)]^\top,$$

where

$$\mathbf{U}_\zeta(\gamma) = \mathbf{a}^\top \mathbf{T} \mathbf{v},$$

$$U_{\beta}(\gamma) = \mathbf{M}^T \mathbf{T} \mathbf{v},$$

$$U_{\phi}(\gamma) = \mathbf{P}^T \mathbf{T} \mathbf{v},$$

$$U_{\theta}(\gamma) = \mathbf{Q}^T \mathbf{T} \mathbf{v},$$

and

$$U_{\delta}(\gamma) = (n-m) \left( \frac{1}{\delta} + b \right) + \sum_{t=m+1}^n [\log(y_t) + c_t - \log(\mu_t)],$$

being

$$\mathbf{T} = \{1/g'(\mu_{m+1}), \dots, 1/g'(\mu_n)\},$$

$$\mathbf{v} = (v_{m+1}, \dots, v_n)^T,$$

$\mathbf{a} = \left( \frac{\partial \eta_{m+1}}{\partial \zeta}, \dots, \frac{\partial \eta_n}{\partial \zeta} \right)^T$  and  $\mathbf{M}$ ,  $\mathbf{P}$ , and  $\mathbf{Q}$  the matrices with dimensions  $(n-m) \times k$ ,  $(n-m) \times p$  and  $(n-m) \times q$ , respectively, for the  $(i, j)$ -th elements given by

$$M_{ij} = \frac{\partial \eta_{i+m}}{\partial \beta_j},$$

$$P_{ij} = \frac{\partial \eta_{i+m}}{\partial \phi_j} \text{ and}$$

$$Q_{ij} = \frac{\partial \eta_{i+m}}{\partial \theta_j}.$$

The condition  $U(\gamma) = \mathbf{0}$ , where  $\mathbf{0}$  is a vector with  $(p+q+k+2)$  zeros, leads to a system of equations with as many equations as unknown parameters, and its solutions correspond to the CMLE of  $\gamma$ . However, this system does not have a closed-form solution. Consequently, the CMLEs were obtained numerically, employing the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method (Press et al., 1992), with analytical first derivatives.

As with most iterative optimization routines, the algorithm requires initial values for the parameters. Following similar approaches in the literature (Bayer et al., 2017, 2020), the initial values for  $\zeta$ ,  $\beta$ , and  $\phi$  were obtained through least squares estimation of a linear regression on  $g(y)$ . The parameters in  $\theta$  were initialized to zero. Lastly, the initial value for  $\delta$  was computed using Eq. (2), replacing the conditional variance and mean by the sample variance and mean of the time series.

### 2.2.2. Parametric trend test

In this section, we propose a parametric trend test based on the asymptotic properties of the proposed Wei-ARMA model parameter estimators. Although this procedure can be applied to any parameter in the model —  $\zeta$ ,  $\beta$ ,  $\phi$ ,  $\theta$ , or  $\delta$  — to assess the relevance of its inclusion, our focus here is specifically on testing the significance of the trend-related parameter  $\beta_l$ , when a covariate  $x_l$  related to a trend is included in the model. This choice is motivated by the critical role that trend detection plays in the analysis of autocorrelated hydrometeorological time series. Trends are often difficult to identify through graphical inspection alone, which reinforces the need for formal statistical tests to support more reliable and objective assessments.

The hypotheses of the test are as follows:

$$\begin{cases} H_0 : \beta_l = 0 \text{ (null hypothesis—no trend),} \\ H_1 : \beta_l \neq 0 \text{ (alternative hypothesis),} \end{cases}$$

where  $\beta_l$  is the parameter related to a deterministic trend covariate  $x_l$  included in the model, such as  $x_{l,t} = t$ , for all  $t$ . The null hypothesis  $H_0$  states that there is no significant trend in the series, while the alternative hypothesis  $H_1$  suggests the presence of a statistically significant trend, either increasing or decreasing.

In likelihood-based inference, hypothesis testing for a specific parameter usually requires the Fisher information matrix. The inverse of this information matrix provides the asymptotic variance–covariance matrix of the estimators, which is fundamental for constructing both confidence intervals and test statistics. The derivation of the conditional Fisher information matrix, denoted by  $\mathbf{K}$ , for the Wei-ARMA model is presented in the Appendix A. The conditional Fisher information

also forms the basis for the normal approximation to the sampling distribution of the CMLE, yielding the following test statistic:

$$z = \frac{\hat{\beta}_l}{\text{se}(\hat{\beta}_l)},$$

where  $\hat{\beta}_l$  is the CMLE of the parameter  $\beta_l$  and  $\text{se}(\hat{\beta}_l)$  is the standard error of  $\hat{\beta}_l$ , given by the square root of the element corresponding to  $\beta_l$  in the diagonal of the inverse of  $\mathbf{K}$ . In large samples and under  $\mathcal{H}_0$ , the statistic  $z$  approximately follows a standard normal distribution. Therefore, the critical values for the test are given by the  $\alpha$ -quantiles of the standard normal distribution, denoted as  $z_{\alpha}$ . If  $|z| > z_{1-\alpha/2}$ , then we reject  $\mathcal{H}_0$  (there is a trend). On the other hand, if  $|z| < z_{1-\alpha/2}$ , then we do not have enough evidence to reject  $\mathcal{H}_0$  (there is no trend). We can also consider the  $p$ -value to evaluate the test. The  $p$ -value is given by  $p\text{-value} = 2 \times (1 - \Phi(|z|))$ , where  $\Phi$  is the cumulative distribution function of  $N(0, 1)$ . Thus, we reject  $\mathcal{H}_0$  if  $p$ -value is less than the significance level  $\alpha$ .

### 2.3. Diagnostic and forecasting

This section presents diagnostic tools and forecasting procedures for the proposed Wei-ARMA model. Once the model is fitted, diagnostic analysis is essential to assess whether it adequately captures the underlying data dynamics. A model that passes these diagnostic checks is considered appropriate for hypothesis testing and forecasting.

To select the most appropriate model from a set of fitted Wei-ARMA models with different orders, we adopt the Akaike Information Criterion (AIC) (Akaike, 1974), a widely recommended criterion in the literature (Chandler and Scott, 2011; Salas, 1992). For each candidate model, parameters are estimated using CMLE, and the one with the smallest AIC value is selected.

After parameter estimation, one of the key diagnostic steps involves analyzing the residuals of the fitted model, which provide insight into model adequacy (Kedem and Fokianos, 2005). In this work, we adopt the quantile residual (Dunn and Smyth, 1996) for diagnostic analysis, which offers several advantages over other types of residuals and has been commonly used in related studies (Bayer et al., 2017, 2020). When the model is correctly fitted, the quantile residuals are, by construction, approximately standard normally distributed and are defined as follows (Dunn and Smyth, 1996):

$$\text{res}_t^{(q)} = \Phi^{-1} (F_Y(y_t | \mathcal{F}_{t-1})),$$

being  $\Phi^{-1}$  the standard normal quantile function. If the model fits the data well, the index plot of the quantile residuals should exhibit white noise behavior, i.e., no visible autocorrelation or systematic pattern (Chandler and Scott, 2011; Kedem and Fokianos, 2005). To formally assess the adequacy of the fitted model, the Ljung–Box test (Ljung and Box, 1978) can be applied to the residuals as a diagnostic tool.

Finally, forecasts can be performed using classical time series forecasting techniques for ARMA models (Box et al., 2008). The first step involves computing the estimates  $\hat{\mu}_{m+1}, \dots, \hat{\mu}_n$  of the conditional mean, based on the CMLE  $\hat{\gamma}$ . To do this, it is necessary to reconstruct the moving average error terms  $\{r_t\}_{t=1}^n = \hat{r}_t$ , which are initialized as zero for  $t \in 1, \dots, m$ . The mean response estimate at  $t$ ,  $t = m+1, m+2, \dots, n$ , is given by:

$$\hat{\mu}_t = g^{-1} \left( \hat{\zeta} + \mathbf{x}_t^T \hat{\beta} + \sum_{i=1}^p \hat{\phi}_i [g(y_{t-i}) - \mathbf{x}_{t-i}^T \hat{\beta}] + \sum_{j=1}^q \hat{\theta}_j \hat{r}_{t-j} \right),$$

for  $\hat{r}_t = g(y_t) - g(\hat{\mu}_t)$ .

Let  $h_0$  denote the forecast horizon, and assume that the covariate values  $\mathbf{x}_t$  for  $t = n+1, \dots, n+h_0$  are either known or can be reasonably estimated for future time points. This assumption is common in cases where covariates follow a deterministic structure, such as sine and cosine functions in harmonic analysis, or polynomial trends. In other cases, covariate values may be forecast from external sources, such as

precipitation predicted by climate or hydrological models. When such covariate information is available or can be approximated, the forecast for the time  $n + h$ , for  $h = 1, \dots, h_0$ , can be computed as:

$$\hat{\mu}_{n+h} = g^{-1} \left( \hat{\zeta} + \mathbf{x}_{n+h}^\top \hat{\beta} + \sum_{i=1}^p \hat{\phi}_i [g(y_{n+h-i}) - \mathbf{x}_{n+h-i}^\top \hat{\beta}] + \sum_{j=1}^q \hat{\theta}_j \hat{r}_{n+h-j} \right),$$

where  $\hat{r}_t = 0$  for  $t > n$  and

$$\hat{g}(y_t) = \begin{cases} g(\hat{\mu}_t), & \text{if } t > n, \\ g(y_t), & \text{if } t \leq n. \end{cases}$$

### 3. Experiments and results

In this section, two computational experiments are conducted. First, a Monte Carlo simulation study is carried out to evaluate the finite-sample performance and applicability of the CMLE for the proposed Wei-ARMA models. Second, the proposed model and inference methodology are applied to a real time series dataset. The results of these experiments are then analyzed and discussed.

#### 3.1. Monte Carlo simulation study

A Monte Carlo simulation was performed to assess the performance of the CMLE for the proposed model. All simulations and analyses were implemented in the R programming language (R. Development Core Team, 2024). A total of 5000 Monte Carlo replications were performed for two model configurations: Wei-ARMA(1,0) and Wei-ARMA(2,1). The parameter values in each scenario were selected based on pilot experiments fitting the Wei-ARMA model to actual monthly streamflow and wind speed time series data. For the first scenario, the Wei-ARMA(1,0) model used the following parameter values:  $\zeta = 3.0$ ,  $\phi_1 = 0.3$ , and  $\delta = 1.2$ . In the second scenario, the Wei-ARMA(2,1) model was specified with  $\zeta = 2.8$ ,  $\phi_1 = -0.6$ ,  $\phi_2 = 0.2$ ,  $\theta_1 = 0.5$ , and  $\delta = 4.0$ . The sample sizes considered were  $n = 60, 120, 240, 480, 720$ , which correspond to simulated monthly observations over periods of 5, 10, 20, 40 and 60 years, respectively. Such monitoring periods are commonly found in the hydrometeorological field.

Finally, each scenario was simulated both with and without the inclusion of a seasonal covariate to assess whether its presence influences the inferential performance of the model parameters. The seasonal covariate was modeled using a cosine function,  $x_t = \cos(2\pi t/12)$ , with  $t = S, \dots, n + S$  (Chandler and Scott, 2011), where  $S$  is an arbitrary phase shift that adjusts the position of the peaks and troughs within a 12-period seasonal cycle. The coefficient  $\beta_1$  associated with the seasonal covariate was fixed in each scenario:  $S = 4$  and  $\beta_1 = 0.6$  for the Wei-ARMA(1,0) model, and  $S = 5$  and  $\beta_1 = 0.2$  for the Wei-ARMA(2,1) model.

Following the simulations, the mean, percentage relative bias (RB%), and mean squared error (MSE) of the CMLE were computed. These metrics are widely recognized as useful criteria for evaluating the performance of parameter estimators. The RB and MSE are defined as follows:

$$RB(\hat{\lambda}) = \frac{B(\hat{\lambda})}{\lambda} \times 100,$$

$$MSE(\hat{\lambda}) = B(\hat{\lambda})^2 + \text{Var}(\hat{\lambda}),$$

where  $\lambda$  is an arbitrary parameter,  $\hat{\lambda}$  is its estimator,  $B(\hat{\lambda}) = E(\hat{\lambda}) - \lambda$  represents the bias, and  $\text{Var}(\hat{\lambda})$  denotes the variance of the estimator. Ideally, these criteria, should be close to zero, indicating that the model parameter estimates are, on average, close to the true parameter values. The results from the Monte Carlo simulations are summarized in Tables 1 to 4.

The CMLE showed good performance across all model configurations, even for small sample sizes. The largest RB and MSE values were observed in the Wei-ARMA(2,1) model without a covariate and with  $n = 60$  (Table 3), where the relative bias for the estimator of

**Table 1**

Monte Carlo simulation results for the CMLE of the Wei-ARMA(1,0) model without a covariate.

Parameter	$\zeta$	$\phi_1$	$\delta$
Value	3.000	0.300	1.200
$n = 60$			
Mean	3.053	0.283	1.241
RB (%)	1.769	-5.669	3.435
MSE	0.155	0.011	0.019
SD	0.389	0.101	0.132
$n = 120$			
Mean	3.020	0.293	1.220
RB (%)	0.666	-2.169	1.706
MSE	0.074	0.005	0.009
SD	0.271	0.070	0.091
$n = 240$			
Mean	3.013	0.296	1.210
RB (%)	0.418	-1.315	0.840
MSE	0.035	0.002	0.004
SD	0.188	0.048	0.061
$n = 480$			
Mean	3.004	0.298	1.206
RB (%)	0.140	-0.526	0.489
MSE	0.017	0.001	0.002
SD	0.132	0.034	0.043
$n = 720$			
Mean	3.004	0.299	1.203
RB (%)	0.127	-0.420	0.283
MSE	0.012	0.001	0.001
SD	0.108	0.028	0.035

**Table 2**

Monte Carlo simulation results for the CMLE of the Wei-ARMA(1,0) model with a covariate.

Parameter	$\zeta$	$\phi_1$	$\delta$	$\beta_1$
Value	3.000	0.300	1.200	0.600
$n = 60$				
Mean	3.089	0.272	1.255	0.597
RB (%)	2.962	-9.483	4.594	-0.502
MSE	0.167	0.012	0.021	0.043
SD	0.398	0.104	0.135	0.207
$n = 120$				
Mean	3.036	0.288	1.227	0.601
RB (%)	1.216	-3.943	2.251	0.224
MSE	0.077	0.005	0.009	0.021
SD	0.275	0.071	0.092	0.145
$n = 240$				
Mean	3.020	0.294	1.213	0.598
RB (%)	0.666	-2.122	1.099	-0.360
MSE	0.036	0.002	0.004	0.010
SD	0.189	0.049	0.062	0.101
$n = 480$				
Mean	3.008	0.297	1.207	0.600
RB (%)	0.270	-0.949	0.621	-0.003
MSE	0.018	0.001	0.002	0.005
SD	0.132	0.034	0.043	0.073
$n = 720$				
Mean	3.006	0.298	1.204	0.601
RB (%)	0.204	-0.673	0.368	0.244
MSE	0.012	0.001	0.001	0.003
SD	0.108	0.028	0.035	0.058

the parameter  $\theta_1$  was  $RB = -16.605\%$ , and the mean squared error for  $\hat{\zeta}$  reached  $MSE = 0.621$ . In contrast, the Wei-ARMA(1,0) model without a covariate (Table 1), which has the fewest parameters, yielded estimates very close to the true parameter values. In this case, the largest RB was below 6%, which is lower than the typical values

**Table 3**  
Monte Carlo simulation results for the CMLE of the Wei-ARMA(2, 1) model without a covariate.

Parameter	$\zeta$	$\phi_1$	$\phi_2$	$\theta_1$	$\delta$
Value	2.800	-0.600	0.200	0.500	4.000
$n = 60$					
Mean	2.623	-0.512	0.200	0.417	4.223
RB (%)	-6.322	-14.595	-0.180	-16.605	5.585
MSE	0.621	0.095	0.018	0.104	0.272
SD	0.768	0.295	0.135	0.312	0.471
$n = 120$					
Mean	2.699	-0.554	0.204	0.457	4.105
RB (%)	-3.598	-7.709	1.989	-8.608	2.614
MSE	0.285	0.038	0.009	0.039	0.107
SD	0.525	0.190	0.092	0.193	0.310
$n = 240$					
Mean	2.745	-0.575	0.203	0.476	4.051
RB (%)	-1.964	-4.126	1.361	-4.876	1.269
MSE	0.139	0.017	0.004	0.017	0.045
SD	0.368	0.129	0.066	0.129	0.206
$n = 480$					
Mean	2.775	-0.589	0.201	0.490	4.028
RB (%)	-0.878	-1.787	0.749	-2.013	0.697
MSE	0.063	0.007	0.002	0.007	0.022
SD	0.251	0.086	0.046	0.084	0.146
$n = 720$					
Mean	2.779	-0.591	0.202	0.491	4.017
RB (%)	-0.736	-1.453	0.789	-1.714	0.417
MSE	0.041	0.005	0.001	0.005	0.014
SD	0.201	0.069	0.037	0.068	0.117

**Table 4**  
Monte Carlo simulation results for the CMLE of the Wei-ARMA(2, 1) model with a covariate.

Parameter	$\zeta$	$\phi_1$	$\phi_2$	$\theta_1$	$\delta$	$\beta_1$
Value	2.800	-0.600	0.200	0.500	4.000	0.200
$n = 60$						
Mean	2.684	-0.533	0.189	0.428	4.276	0.200
RB (%)	-4.130	-11.128	-5.628	-14.428	6.902	-0.008
MSE	0.633	0.091	0.020	0.105	0.310	0.003
SD	0.787	0.293	0.140	0.316	0.483	0.051
$n = 120$						
Mean	2.726	-0.562	0.198	0.460	4.130	0.201
RB (%)	-2.650	-6.390	-1.000	-8.081	3.238	0.354
MSE	0.302	0.041	0.009	0.043	0.115	0.001
SD	0.545	0.198	0.096	0.202	0.313	0.035
$n = 240$						
Mean	2.762	-0.581	0.200	0.479	4.062	0.200
RB (%)	-1.356	-3.207	-0.238	-4.248	1.539	-0.159
MSE	0.136	0.017	0.004	0.017	0.047	0.001
SD	0.367	0.128	0.067	0.127	0.207	0.024
$n = 480$						
Mean	2.780	-0.591	0.201	0.491	4.033	0.200
RB (%)	-0.701	-1.520	0.252	-1.896	0.825	0.005
MSE	0.065	0.008	0.002	0.007	0.022	0.000
SD	0.253	0.086	0.047	0.084	0.144	0.017
$n = 720$						
Mean	2.784	-0.593	0.201	0.492	4.020	0.200
RB (%)	-0.575	-1.221	0.324	-1.583	0.505	0.145
MSE	0.042	0.005	0.001	0.005	0.014	0.000
SD	0.204	0.070	0.038	0.069	0.118	0.014

reported in the literature for the comparable sample sizes (Bayer et al., 2017, 2020; Melo and Alencar, 2020). It is well documented in the literature on autoregressive and related time series models that the finite-sample performance of parameter estimators is adversely affected by the inclusion of additional parameters (Shaman and Stine, 1988; Abadir et al., 1999). The most complex configuration in this study, the Wei-ARMA(2,1) model with a covariate (Table 4), also showed the

highest relative bias (RB = -14.428% for  $\hat{\theta}_1$ ) and highest mean squared error (MSE = 0.633 for  $\hat{\zeta}$ ). For the scenario with covariate, shown in Table 2, the simulation results demonstrated good performance, with the largest relative bias being -9.483% for  $\hat{\phi}_1$ , and the MSE equal to 0.167 for  $\hat{\zeta}$ . Overall, the simulation results confirm that both RB and MSE decrease as the sample size  $n$  increases, which aligns with the asymptotic properties of the CMLE.

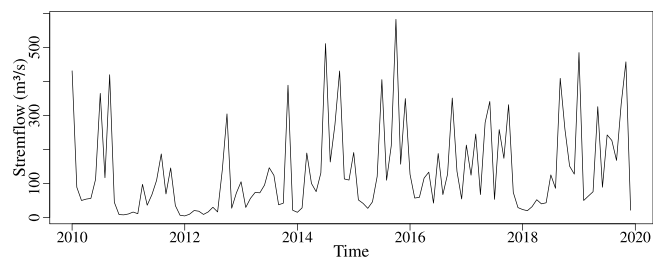


Fig. 2. Time series of the monthly mean streamflow of the Vacacaí River.

Table 5

Descriptive statistics of the monthly mean streamflow series of the Vacacaí River.

Non-conditional mean ( $\text{m}^3/\text{s}$ )	137.15
Standard deviation	131.75
Median ( $\text{m}^3/\text{s}$ )	92.67
Maximum ( $\text{m}^3/\text{s}$ )	583.21
Minimum ( $\text{m}^3/\text{s}$ )	5.11
Skewness	1.35
Kurtosis	4.02

### 3.2. Case study: streamflow data modeling

Accurate streamflow forecasting models are essential for effective water resource management, as they help predict floods and droughts and support efforts to mitigate socioeconomic impacts. Southern Brazil, particularly the state of Rio Grande do Sul (RS), has been increasingly affected by climate change (Piveta, 2024), experiencing both extreme flood events that have caused widespread damage and severe droughts that have significantly impacted agricultural productivity. In this context, we selected the Vacacaí River in RS for analysis of monthly mean streamflow. The data were obtained from the *Passo das Tunas* streamflow gauge station (ANA agency code 85600000), located at latitude  $-29.9258$  and longitude  $-53.4167$ , with a drainage area of  $6780 \text{ km}^2$  (ANA, 2025). Although records at this station date back to 1940, the historical series contains several gaps. For the purposes of this study, we used a time series continuously spanning January 2010 to December 2020.

Figs. 2 and 3 display, respectively, the time series and the main descriptive features of the monthly mean streamflow for the Vacacaí River. The unconditional mean streamflow over this period is  $137.15 \text{ m}^3/\text{s}$ , with a range of  $578.1 \text{ m}^3/\text{s}$  (Table 5). This wide range, with alternating peaks and near-zero values, is also reflected in the high standard deviation and kurtosis, the latter indicating a sharply peaked distribution.

The highest monthly mean streamflow,  $583.21 \text{ m}^3/\text{s}$ , occurred in October 2015, followed by  $511.88 \text{ m}^3/\text{s}$  in July 2014. These peak values are approximately four times higher than the unconditional mean. More recently, in May 2024, an exceptional flood event resulted in a mean monthly streamflow of  $819.44 \text{ m}^3/\text{s}$ , nearly six times the long-term mean. While this value is noteworthy, it could not be included in the present analysis due to data gaps. Nonetheless, we highlight this event here given its significant public impact and potential relevance to future studies. In contrast, two extended low-flow periods are evident: from October 2010 to May 2011, when the mean flow dropped to levels more than 16 times lower than the unconditional mean, and from November 2011 to August 2012, with values nearly 26 times lower than the mean.

Maximum and minimum streamflow values alternate throughout the year, reflecting a clear seasonal pattern, as illustrated in Figs. 3(a) and 4. The wet season spans from June to November, with October being the wettest month, while the dry season extends from December to May, with February showing the lowest flow.

Regarding the trend, Fig. 4 suggests an upward shift in mean streamflow over time. To formally assess this trend, we apply the

proposed trend test described in Section 2.2.2. Another characteristic is the high concentration of data at lower values. This is evident from the relationship between the median, mean, and overall data range, as well as the positive skewness, which indicates a right-skewed distribution (see Fig. 3(b)).

We modeled the data using the Box–Jenkins methodology (Box et al., 2008) to identify an appropriate time series model, which consists of three iterative stages: model identification, parameter estimation, and diagnostic checking. Seasonal behavior was captured through the covariate  $x_{t1} = \cos(2\pi t/12)$ , with  $t = S, \dots, n + S$  and  $S = 3$ , while a linear trend component was included as  $x_{t2} = 1.0, 1.1, 1.2, \dots, (1 + (n - 1) \times 0.1)$ . The total sample size is 120, with the final 12 observations reserved for forecast validation. Table 6 reports the estimated parameters of the selected Wei-ARMA(1,0) model, and residual diagnostics are shown in Fig. 5. All coefficients are statistically significant at the 5% level, including  $\beta_2$ , confirming the presence of a trend, as indicated by the proposed test. This finding is consistent with the visual inspection of Fig. 4, which suggests an increasing trend. The results indicate an increase in streamflow over time, as evidenced by the positive value of the parameter  $\beta_2$ . This upward trend may be associated with changes in precipitation patterns, temperature variations, or anthropogenic influences, including land use changes, agricultural and industrial activities, and the construction of dams. Previous studies conducted in different regions have shown that these factors can alter the hydrological cycle and, consequently, streamflow (Silva et al., 2020; Cecílio et al., 2021; Avcioglu et al., 2025; Yan, 2025; Petry et al., 2025). However, further studies are needed to accurately determine the underlying causes of this trend, although some regional studies have indicated an increase in temperature (Cardoso et al., 2022).

It is also worth emphasizing that the application of well-known nonparametric tests, such as the Mann–Kendall test, is not appropriate in this context, since these tests assume independence of the observations. This assumption is violated, as evidenced by the nonparametric Brock–Dechert–Scheinkman (BDS) test for independence, which yields a  $p$ -value of 0.001, as computed using the `serie` function from the `tsseries` R package (Trapletti and Hornik, 2020).

Fig. 5(a) shows residuals plotted over time, revealing no visible patterns. The normal Q–Q plot in Fig. 5(b) displays a nearly straight line, supporting the assumption of normally distributed residuals. The residual autocorrelation function (ACF) and partial ACF (PACF) plots (Figs. 5(c) and 5(d), respectively) support the white noise assumption, which was further verified using the Ljung–Box test (see Table 6). Overall, the residual diagnostics confirm that the fitted model is adequate and suitable for out-of-sample forecasting.

We also fitted an ARMA(2,2) model, which was selected as the best among the ARMA candidates based on the smallest AIC value. This model was fitted using the `Arima` function from the `forecast` R package (Hyndman et al., 2022). We adopted the same covariates as those used in the Wei-ARMA model. It is worth highlighting that the coefficient related to the trend was not statistically significant at the 5% level ( $p$ -value = 0.189). This result leads to the interpretation that there is no trend in the time series data, which contrasts with the findings from the previously presented test. This discrepancy may be due to an incorrect Gaussian assumption about the data distribution for inference over the ARMA model parameters.

Fig. 6 displays the fitted values and out-of-sample forecasts for both adjusted Wei-ARMA and ARMA models. We note that the Wei-ARMA model provided a closer representation of the observed dynamics, while the ARMA model exhibited an overly smooth behavior that failed to capture the variability of the series. To facilitate comparison, Table 7 presents a summary of the fitted and forecast values for both models. Notably, the Wei-ARMA model more accurately captured peak streamflow values, with its prediction summary statistics more closely aligned with the observed data than those of the ARMA model. In contrast, the ARMA model yielded some negative forecasts, which lack physical

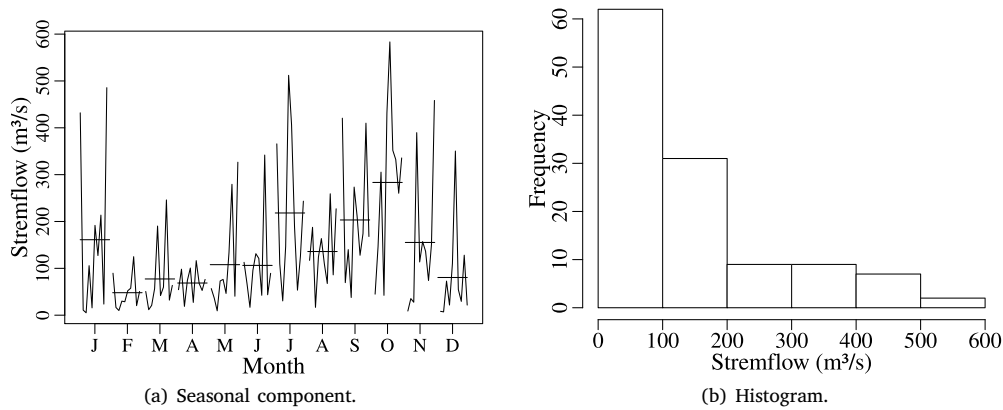


Fig. 3. Characterization of the monthly mean streamflow series of the Vacacaí River.

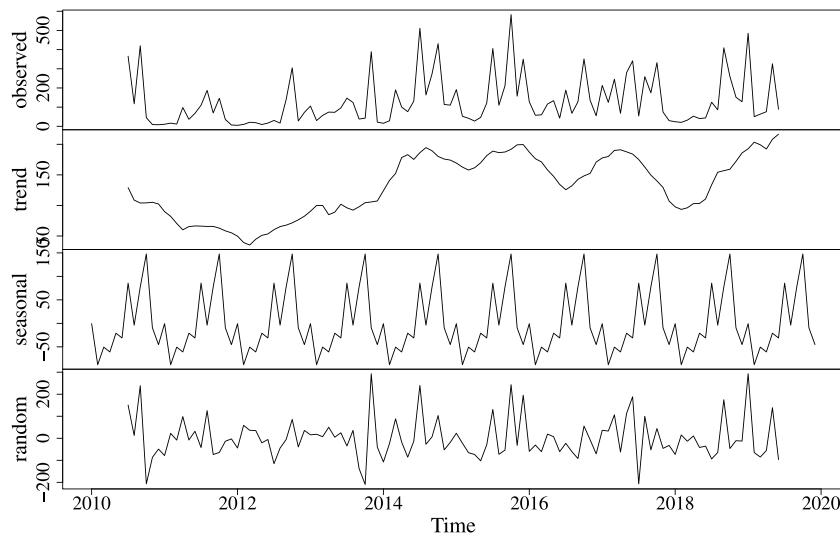


Fig. 4. Decomposition of the monthly mean streamflow series of the Vacacaí River.

Table 6

Coefficients,  $p$ -values, and diagnostic measures of the fitted Wei-ARMA model for monthly streamflow data.

Parameter	$\zeta$	$\phi_1$	$\delta$	$\beta_1$	$\beta_2$ (trend)
Estimate	2.8263	0.3446	1.2931	0.6825	0.0812
$p$ -valor ( $z$ stat.)	<0.001	<0.001	<0.001	<0.001	0.028

AIC = 1222.71

Ljung-Box:  $Q = 0.108$  ( $p$ -value = 0.742)

meaning in the context of streamflow. These unrealistic outcomes are highlighted with a circle in Fig. 6.

To quantitatively evaluate model accuracy, we used the mean absolute percentage error (MAPE) as a figure-of-merit. For the fitted values, the MAPE was 121.98% for the Wei-ARMA model and 145.32% for the ARMA model, corresponding to a reduction of 23.34%. In the forecast period, the MAPE was 134.58% for the Wei-ARMA model and 103.35% for the ARMA model. Although the ARMA model slightly outperformed the Wei-ARMA model in terms of MAPE during the forecasting phase, the Wei-ARMA model demonstrated superior performance in capturing the true streamflow behavior, particularly in representing extremes and avoiding unrealistic values such as negative forecasts. Figs. 7(a) and 7(b) display the ranked decreasing values for the fitted and forecast periods, respectively. It can be observed that, across the entire range of values, the proposed model yields predictions that are closer to the actual observations. These plots emphasize the superior performance of

the Wei-ARMA model in predicting extreme values when compared to the standard ARMA model.

It is also important to highlight that the selected Wei-ARMA model is more parsimonious than the selected ARMA model, as it achieves similar or better performance using fewer parameters. The adjusted Wei-ARMA model includes only three coefficients ( $\zeta$ ,  $\phi_1$ , and  $\delta$ ), while the ARMA model requires five ( $\zeta$ ,  $\phi_1$ ,  $\phi_2$ ,  $\theta_1$ , and  $\theta_2$ ).

In order to enrich the analysis and better explore the characteristics of the Weibull distribution for modeling extreme values, we also considered time series of the monthly maximum and minimum values over the same period, using the same covariates. For the monthly maximum streamflow data, the best-performing model was the Wei-ARMA(1,0), with the following estimated parameters:  $\hat{\zeta} = 3.6817$ ,  $\hat{\phi}_1 = 0.2846$ ,  $\hat{\delta} = 1.2490$ ,  $\hat{\beta}_1 = 0.6444$ , and  $\hat{\beta}_2 = 0.0908$ ; and ARMA(3,0), with estimates:  $\hat{\zeta} = 229.8129$ ,  $\hat{\phi}_1 = 0.0097$ ,  $\hat{\phi}_2 = 0.0905$ ,  $\hat{\phi}_3 = 0.2576$ ,  $\hat{\beta}_1 = 170.6109$ , and  $\hat{\beta}_2 = 14.5902$ . All parameters were statistically significant at the

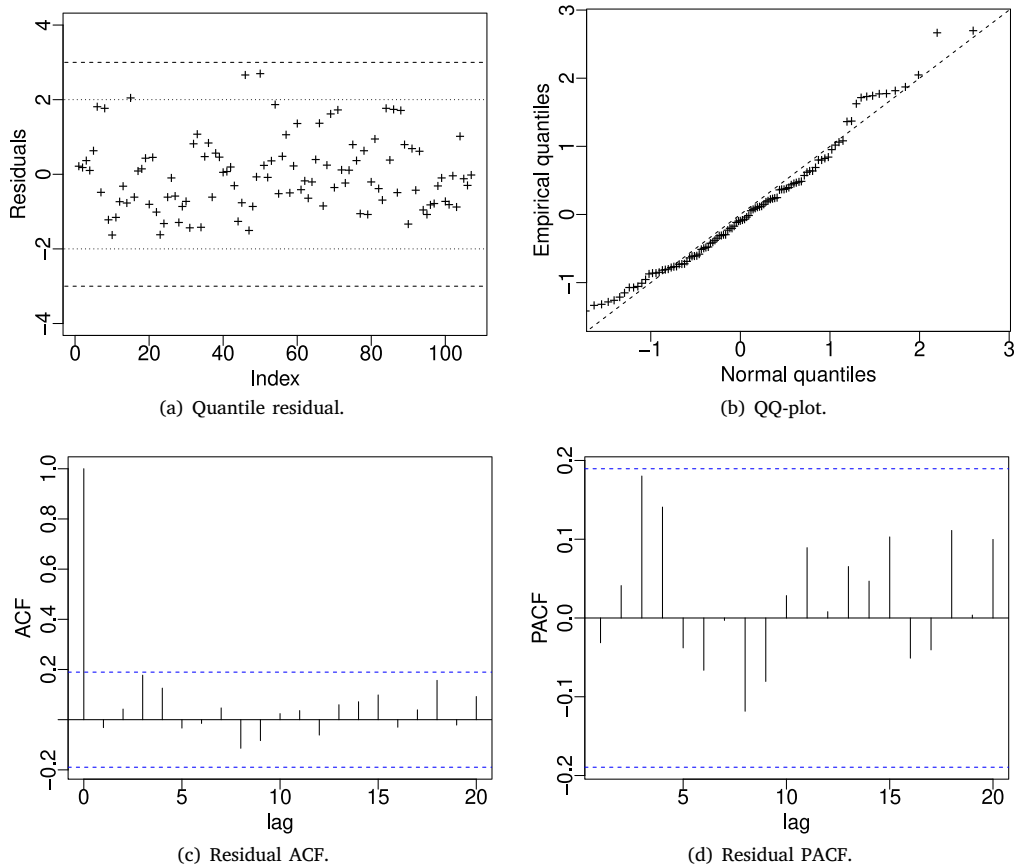


Fig. 5. Diagnostic plot of the fitted Wei-ARMA model for monthly mean streamflow data based on the quantile residual.

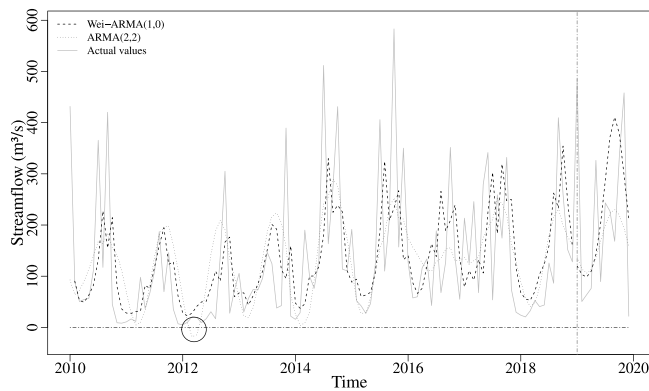


Fig. 6. Fitted and forecast values for Wei-ARMA and ARMA models of the monthly mean streamflow of Vacacaí River.

5% level, and the residuals passed the Ljung–Box diagnostic test. For the fitted values, the MAPE was 142.48% for the Wei-ARMA model and 176.72% for the ARMA model, representing a reduction of 34.24%; and for the forecasts, the MAPE increased to 182.90% for the Wei-ARMA model and 147.93% for the ARMA model. Fig. 8 presents the fitted and forecast values for the maximum streamflow time series. Besides some trade-offs in the accuracy metrics, visual inspection suggests that the Wei-ARMA model better captures the dynamic behavior of the extreme values when compared to the ARMA model.

The best models for the monthly minimum streamflow series were the Wei-ARMA(1,1) and ARMA(3,0) models. The adjusted Wei-ARMA model presented the following coefficients:  $\hat{\zeta} = 0.6418$ ,  $\hat{\phi}_1 = 0.7764$ ,

Table 7

Summary statistics of fitted and forecast values for different models.

Model	Observed	Wei-ARMA	ARMA
Fitted			
Mean (m <sup>3</sup> /s)	128.82	130.52	129.71
Standard deviation	126.26	79.45	72.97
Median (m <sup>3</sup> /s)	87.84	111.01	134.60
Maximum (m <sup>3</sup> /s)	583.21	354.29	289.71
Minimum (m <sup>3</sup> /s)	5.11	22.29	-18.37
Forecasting			
Mean (m <sup>3</sup> /s)	212.11	227.86	166.60
Standard deviation	160.90	117.19	45.68
Median (m <sup>3</sup> /s)	197.56	207.63	163.60
Maximum (m <sup>3</sup> /s)	485.53	410.87	230.10
Minimum (m <sup>3</sup> /s)	21.74	98.27	105.60

$\hat{\theta}_1 = -0.5310$ ,  $\hat{\delta} = 1.4593$ ,  $\hat{\beta}_1 = 0.4217$ , and  $\hat{\beta}_2 = 0.1084$ , while the ARMA model had  $\hat{\zeta} = 25.0308$ ,  $\hat{\phi}_1 = 0.2101$ ,  $\hat{\phi}_2 = 0.0382$ ,  $\hat{\phi}_3 = 0.2807$ ,  $\hat{\beta}_1 = 13.9006$ , and  $\hat{\beta}_2 = 1.2390$ . All parameters were statistically significant at the 5% level, and the residuals from both models passed the Ljung–Box test. For the fitted values, the Wei-ARMA model achieved an MAPE of 87.81%, while the ARMA model yielded an MAPE of 97.09%, corresponding to a reduction of 9.28%. In the forecasting period, however, the Wei-ARMA model did not perform well, with an MAPE of 105.82% compared to 50.53% for the ARMA model. Overall, the two models demonstrated comparable performance when modeling the minimum streamflow series, as illustrated in Fig. 9. While the ARMA model performed better in forecasting accuracy, both models captured the general trend and structure of the data similarly.

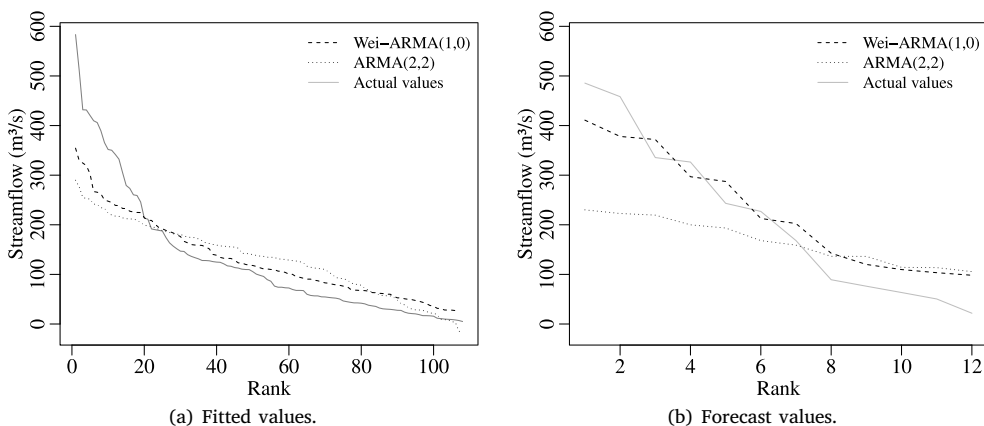


Fig. 7. Ranked decreasing fitted and forecast values from Wei-ARMA and ARMA models for the monthly mean streamflow of the Vacacaí River.

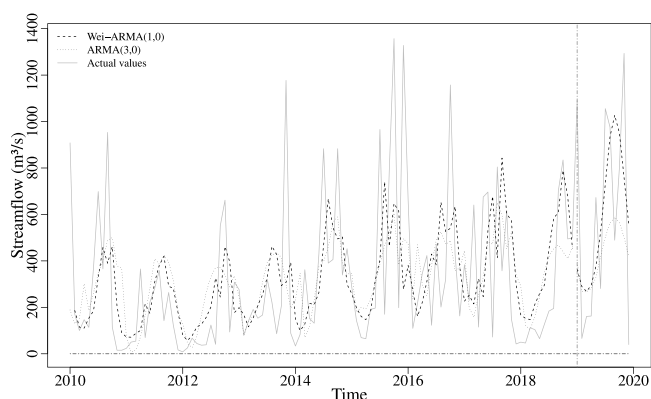


Fig. 8. Fitted and forecast values for Wei-ARMA and ARMA models of the monthly maximum streamflow of Vacacaí River.

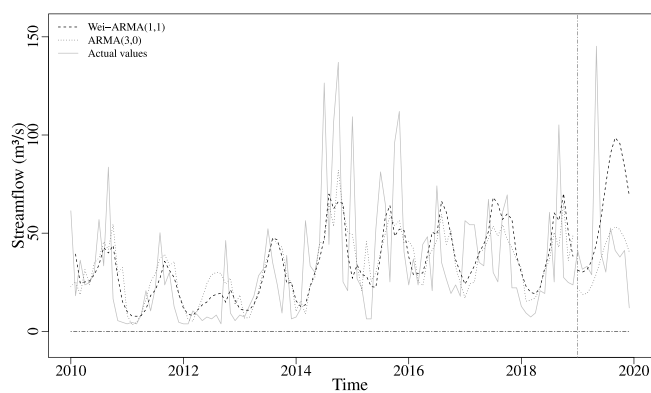


Fig. 9. Fitted and forecast values for Wei-ARMA and ARMA models of the monthly minimum streamflow of Vacacaí River.

Regarding the trend test, in the Wei-ARMA models fitted to both the monthly maximum and minimum streamflow time series, the parameters  $\beta_2$  associated with the trend covariate were statistically significant at the 5% level, with  $p$ -values equal to 0.010 and 0.017, respectively. This indicates the presence of a positive trend not only in the monthly mean streamflow, but also in the extreme values (maximum and minimum), as shown in Figs. 8 and 9. The presence of increasing trends in streamflow extremes in southern Brazil is consistent with findings reported in the literature (Souza and Reis, 2022; Costa et al., 2025; Laipelt et al., 2025). In contrast, when the trend covariates are tested in the Gaussian ARMA models, the results show no statistically significant trends at the 5% level in either the maximum ( $p$ -value = 0.218) or minimum ( $p$ -value = 0.579) monthly streamflow series.

#### 4. Conclusion

This work proposed a novel data-driven model that integrates the ARMA structure with the Weibull distribution, a distribution widely used for modeling hydrometeorological variables. The model was developed, theoretically formulated, numerically evaluated by simulations, and applied to observed streamflow time series. A parametric trend test was also introduced and discussed. The proposed model serves as a tool for analyzing and predicting continuous, autocorrelated, asymmetric, and nonnegative data, particularly in the hydrometeorological field, such as streamflow. Furthermore, this work aims to promote and contribute to the use of statistical methods for modeling and forecasting in hydrometeorological and environmental sciences.

Monte Carlo simulation results confirmed the reliability of the conditional maximum likelihood estimators of the Wei-ARMA parameters, which performed well in terms of relative bias and mean square error, even for small sample sizes. These findings support the theoretical foundations and computational implementation of the proposed model. Sample sizes smaller than 60 were not evaluated and may represent a lower bound for maintaining reliable parameter estimation. The application of the Wei-ARMA model to the analysis of monthly mean, minimum, and maximum time series highlights its practical relevance. The AIC criterion is recommended for model order selection, along with verification of the i.i.d. assumption in the model residuals using a portmanteau test. The proposed model effectively captured key features of the observed data and, in general, outperformed the conventional ARMA model. Additionally, the introduced parametric trend test was able to detect a positive trend, whereas the Gaussian-based ARMA models failed to do so. Importantly, by modeling the response variable with a distribution bounded at zero, the Wei-ARMA model naturally avoids unrealistic negative predictions, an important advantage when dealing with hydrometeorological variables.

Finally, it is worth noting that stochastic ARMA-based models, including the Wei-ARMA model, can be fitted and used to generate predictions based solely on the time series of the variable under study, an especially useful feature in regions with sparse hydrometeorological monitoring networks. Moreover, when additional information is available, the model is flexible enough to incorporate covariates, allowing for potentially improved analyses and interpretations.

In future work, prediction intervals, improvements in small-sample inference, and model selection strategies will be investigated. In addition, extensions of the Wei-ARMA model to accommodate long-memory structures will be explored.

### CRedit authorship contribution statement

**Débora Missio Bayer:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Júlia Konrad:** Writing – original draft, Validation, Methodology, Investigation, Formal analysis. **Bruna G. Palm:** Writing – review & editing, Supervision, Methodology, Formal analysis. **Fábio M. Bayer:** Writing – review & editing, Validation, Supervision, Methodology, Funding acquisition, Formal analysis, Conceptualization.

### Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT-4o to improve the readability and language of the manuscript. The authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

Financial support for this study was provided by the National Council for Scientific and Technological Development (CNPq) of the Ministry of Science, Technology and Innovation (MCTI) of Brazil; the Swedish-Brazilian Research and Innovation Center (CISB); Saab AB; and the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES), Ministry of Education of Brazil – Finance Code 001.

### Appendix A. Conditional Fisher information matrix

We begin by defining a set of expectations that will be used in the derivation of the conditional Fisher information matrix. These expectations form the basis for the analytical expressions of the matrix elements:

$$E(Y_t^\delta | \mathcal{F}_{t-1}) = \left( \frac{\mu_t}{\Gamma(1+1/\delta)} \right)^\delta = h_t^{-\delta}, \quad (\text{A.1})$$

$$\text{being } h_t = \frac{\Gamma(1+1/\delta)}{\mu_t},$$

$$E(Y_t^\delta \log(Y_t) | \mathcal{F}_{t-1}) = \frac{1}{\delta} h_t^{-\delta} (-\delta \log(h_t) - \kappa + 1), \quad (\text{A.2})$$

where  $\kappa = 0,5772156649 \dots$  is the Euler–Mascheroni constant (Gradshcheyn and Ryzhik, 2007), and

$$E(Y_t^\delta \log^2(Y_t) | \mathcal{F}_{t-1}) = h_t^{-\delta} \left( \log^2(h_t) + \frac{2}{\delta} (\kappa - 1) \log(h_t) + \frac{1}{\delta^2} \left( \kappa^2 - 2\kappa + \frac{\pi^2}{6} \right) \right). \quad (\text{A.3})$$

Thus, for the parameters of the dynamic structure, we have:

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \gamma_i \partial \gamma_j} &= \sum_{t=m+1}^n \frac{\partial^2 \ell_t(\mu_t, \delta)}{\partial \gamma_i \partial \gamma_j} = \sum_{t=m+1}^n \left( \frac{\partial \ell_t(\mu_t, \delta)}{\partial \mu_i} \frac{d\mu_t}{d\eta_i} \frac{\partial \eta_t}{\partial \gamma_j} \right) \frac{d\mu_t}{d\eta_i} \frac{\partial \eta_t}{\partial \gamma_i} \\ &= \sum_{t=m+1}^n \left[ \frac{\partial^2 \ell_t(\mu_t, \delta)}{\partial \mu_i^2} \frac{d\mu_t}{d\eta_i} \frac{\partial \eta_t}{\partial \gamma_j} + \frac{\partial \ell_t(\mu_t, \delta)}{\partial \mu_i} \frac{\partial}{\partial \mu_i} \left( \frac{d\mu_t}{d\eta_i} \frac{\partial \eta_t}{\partial \gamma_j} \right) \right] \frac{d\mu_t}{d\eta_i} \frac{\partial \eta_t}{\partial \gamma_i}, \end{aligned}$$

with  $i, j = 1, \dots, (p + q + r + 1)$ . However, we can prove that  $E \left( \frac{\partial \ell_t(\mu_t, \delta)}{\partial \mu_i} \middle| \mathcal{F}_{t-1} \right) = 0$ , so, after applying the conditional expectation in the equation above, we have:

$$E \left( \frac{\partial^2 \ell_t(\mu_t, \delta)}{\partial \gamma_i \partial \gamma_j} \middle| \mathcal{F}_{t-1} \right) = E \left( \frac{\partial^2 \ell_t(\mu_t, \delta)}{\partial \mu_i^2} \middle| \mathcal{F}_{t-1} \right) \left( \frac{d\mu_t}{d\eta_i} \right)^2 \frac{\partial \eta_t}{\partial \gamma_i} \frac{\partial \eta_t}{\partial \gamma_j},$$

where  $\frac{\partial \eta_t}{\partial \gamma_j}$  was already calculated to obtain the score vector,  $\left( \frac{d\mu_t}{d\eta_i} \right)^2 = \left( \frac{1}{g'(\mu_t)} \right)^2$ , and the expectation of the second-order derivative of the conditional log-likelihood with respect to  $\mu_i$  is obtained by differentiating Equation (4), again, yielding:

$$\frac{\partial^2 \ell_t(\mu_t, \delta)}{\partial \mu_i^2} = \frac{\delta}{\mu_t^2} \left[ 1 - (\delta + 1) \left( \frac{y_t \Gamma(1+1/\delta)}{\mu_t} \right)^\delta \right].$$

Replacing the expectation of Eq. (A.1), the expected value of the expression above becomes:

$$E \left( \frac{\partial^2 \ell_t(\mu_t, \delta)}{\partial \mu_i^2} \middle| \mathcal{F}_{t-1} \right) = \frac{\delta}{\mu_t^2} \left[ 1 - (\delta + 1) E(Y_t^\delta | \mathcal{F}_{t-1}) \left( \frac{\Gamma(1+1/\delta)}{\mu_t} \right)^\delta \right] = -\frac{\delta^2}{\mu_t^2}.$$

For presentation purposes, we consider:

$$E \left( \frac{\partial^2 \ell_t(\mu_t, \delta)}{\partial \mu_i^2} \middle| \mathcal{F}_{t-1} \right) \left( \frac{d\mu_t}{d\eta_i} \right)^2 = \left( -\frac{\delta^2}{\mu_t^2} \right) \left( \frac{1}{g'(\mu_t)} \right)^2 = w_t,$$

and thus, the conditional expectation of the second-order derivative of the conditional log-likelihood with respect to the parameters of the dynamic structure of the model can be written as the general expression:

$$\sum_{t=m+1}^n w_t \frac{\partial \eta_t}{\partial \gamma_i} \frac{\partial \eta_t}{\partial \gamma_j}.$$

Next, we need to determine the second-order derivatives of the conditional log-likelihood with respect to  $\delta$ , the shape parameter of the Weibull distribution. The second-order derivative of the log-likelihood with respect to the set of parameters  $\gamma$  and to  $\delta$  is obtained by differentiating (5) with respect to  $\delta$ , resulting in:

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \gamma_j \partial \delta} &= \sum_{t=m+1}^n \left\{ \frac{1}{\mu_t} [-1 + (y_t h_t)^\delta (\delta \log(y_t h_t) - \psi(1 + \frac{1}{\delta}) + 1)] \right\} \frac{1}{g'(\mu_t)} \frac{\partial \eta_t}{\partial \gamma_j}. \end{aligned}$$

Similarly, by substituting the expectations of Eq. (A.1) and (A.2), the expectation of the equation above reduces to:

$$\sum_{t=m+1}^n d_t \frac{1}{g'(\mu_t)} \frac{\partial \eta_t}{\partial \gamma_j},$$

$$\text{where } d_t = \frac{1}{\mu_t} \left( -\psi \left( 1 + \frac{1}{\delta} \right) - \kappa + 1 \right).$$

Finally, the second-order derivative of the conditional log-likelihood with respect to  $\delta$  was obtained by differentiating again Eq. (6):

$$\begin{aligned} \frac{\partial^2 \ell_t(\mu_t, \delta)}{\partial \delta^2} &= -\frac{1}{\delta^2} + \frac{\psi'(1+1/\delta)}{\delta^3} - (y_t h_t)^\delta \left[ \log^2(y_t h_t) - \frac{2\psi(1+1/\delta) \log(y_t h_t)}{\delta} + \frac{\psi(1+1/\delta)^2}{\delta^2} + \frac{\psi'(1+1/\delta)}{\delta^3} \right]. \end{aligned}$$

By substituting the expectations from Eqs. (A.2) and (A.3) into the equation above, the conditional expectation simplifies to:

$$\begin{aligned} E \left( \frac{\partial^2 \ell_t(\mu_t, \delta)}{\partial \delta^2} \middle| \mathcal{F}_{t-1} \right) &= \frac{1}{\delta^2} \left[ \psi \left( 1 + \frac{1}{\delta} \right) \left( -\psi \left( 1 + \frac{1}{\delta} \right) - 2\kappa + 2 \right) + \kappa^2 + 2\kappa - \frac{\pi^2}{6} - 1 \right] = s. \end{aligned}$$

Thus, the second-order derivative of  $\ell$  with respect to  $\delta$  is given by  $(n - m)s$ .

The conditional Fisher information matrix,  $\mathbf{K}(\boldsymbol{\gamma})$ , is given by the negative value of the conditional expectation of the second-order derivative of the log-likelihood with respect to each parameter of the model. In its matrix form, the conditional Fisher information matrix can be written as follows:

$$\mathbf{K} = \mathbf{K}(\boldsymbol{\gamma}) = \begin{pmatrix} \mathbf{K}(\zeta, \zeta) & \mathbf{K}(\zeta, \beta) & \mathbf{K}(\zeta, \phi) & \mathbf{K}(\zeta, \theta) & \mathbf{K}(\zeta, \delta) \\ \mathbf{K}(\beta, \zeta) & \mathbf{K}(\beta, \beta) & \mathbf{K}(\beta, \phi) & \mathbf{K}(\beta, \theta) & \mathbf{K}(\beta, \delta) \\ \mathbf{K}(\phi, \zeta) & \mathbf{K}(\phi, \beta) & \mathbf{K}(\phi, \phi) & \mathbf{K}(\phi, \theta) & \mathbf{K}(\phi, \delta) \\ \mathbf{K}(\theta, \zeta) & \mathbf{K}(\theta, \beta) & \mathbf{K}(\theta, \phi) & \mathbf{K}(\theta, \theta) & \mathbf{K}(\theta, \delta) \\ \mathbf{K}(\delta, \zeta) & \mathbf{K}(\delta, \beta) & \mathbf{K}(\delta, \phi) & \mathbf{K}(\delta, \theta) & \mathbf{K}(\delta, \delta) \end{pmatrix},$$

where:

$$\begin{aligned} \mathbf{K}(\zeta, \zeta) &= -\mathbf{a}^\top \mathbf{W} \mathbf{a}, \\ \mathbf{K}(\zeta, \beta) &= \mathbf{K}(\beta, \zeta) = -\mathbf{a}^\top \mathbf{W} \mathbf{M}, \\ \mathbf{K}(\zeta, \phi) &= \mathbf{K}(\phi, \zeta) = -\mathbf{a}^\top \mathbf{W} \mathbf{P}, \\ \mathbf{K}(\zeta, \theta) &= \mathbf{K}(\theta, \zeta) = -\mathbf{a}^\top \mathbf{W} \mathbf{Q}, \\ \mathbf{K}(\zeta, \delta) &= \mathbf{K}(\delta, \zeta) = -\mathbf{a}^\top \mathbf{T} \mathbf{d}, \\ \mathbf{K}(\beta, \beta) &= -\mathbf{M}^\top \mathbf{W} \mathbf{M}, \\ \mathbf{K}(\beta, \phi) &= \mathbf{K}(\phi, \beta) = -\mathbf{M}^\top \mathbf{W} \mathbf{P}, \\ \mathbf{K}(\beta, \theta) &= \mathbf{K}(\theta, \beta) = -\mathbf{M}^\top \mathbf{W} \mathbf{Q}, \\ \mathbf{K}(\beta, \delta) &= \mathbf{K}(\delta, \beta) = -\mathbf{M}^\top \mathbf{T} \mathbf{d}, \\ \mathbf{K}(\phi, \phi) &= -\mathbf{P}^\top \mathbf{W} \mathbf{P}, \\ \mathbf{K}(\phi, \theta) &= \mathbf{K}(\theta, \phi) = -\mathbf{P}^\top \mathbf{W} \mathbf{Q}, \\ \mathbf{K}(\phi, \delta) &= \mathbf{K}(\delta, \phi) = -\mathbf{P}^\top \mathbf{T} \mathbf{d}, \\ \mathbf{K}(\theta, \theta) &= -\mathbf{Q}^\top \mathbf{W} \mathbf{Q}, \\ \mathbf{K}(\theta, \delta) &= \mathbf{K}(\delta, \theta) = -\mathbf{Q}^\top \mathbf{T} \mathbf{d}, \text{ and} \\ \mathbf{K}(\delta, \delta) &= -(n-m)s, \text{ with } \mathbf{d} = (d_{m+1}, \dots, d_n)^\top \text{ and } \mathbf{W} = \text{diag}(w_{m+1}, \dots, w_n)^\top. \end{aligned}$$

The conditional Fisher information matrix is not block diagonal, and hence the parameters are not orthogonal (Cox and Reid, 1987).

## Appendix B. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.envsoft.2026.107027>.

## Data availability

The sources of software and data are cited in the manuscript. The code, developed by the authors, is available at: <https://github.com/deborabayer-cpu/WEI-ARMA>.

## References

Abadir, K.M., Hadri, K., Tzavalis, E., 1999. The influence of VAR dimensions on estimator biases. *Econometrica* 67, 163–181.

Akaike, H., 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control* 19, 716–723. <http://dx.doi.org/10.1109/TAC.1974.1100705>.

ANA, 2025. Portal hidroweb. URL <https://www.snirh.gov.br/hidroweb>. (Acesso em 03 March 2025).

Avcioglu, A., Vandromme, R., Grangeon, T., Minella, J.P.G., Evrard, O., Tassano, M., Scariot, N., Cerdan, O., 2025. Hydrological responses to land use changes and precipitation variability in southern Brazil. *J. Hydrol.: Reg. Stud.* 61, 102705. <http://dx.doi.org/10.1016/j.ejrh.2025.102705>.

Bardsley, E., 2019. The Weibull distribution as an extreme value model for transformed annual maxima. *J. Hydrol. (New Zealand)* 58, 123–131.

Bayer, F.M., Bayer, D.M., Marinoni, A., Gamba, P., 2020. A novel Rayleigh dynamical model for remote sensing data interpretation. *IEEE Trans. Geosci. Remote Sens.* 58, 4989–4999. <http://dx.doi.org/10.1109/TGRS.2020.2971345>.

Bayer, F.M., Bayer, D.M., Pumi, G., 2017. Kumaraswamy autoregressive moving average models for double bounded environmental data. *J. Hydrol.* 555, 385–396. <http://dx.doi.org/10.1016/j.jhydrol.2017.10.006>.

Bayer, F.M., Rosa, C.M., Cribari-Neto, F., 2025. A novel data-driven dynamic model for inflated doubly-bounded hydro-environmental time series. *Appl. Math. Model.* 137, 115680. <http://dx.doi.org/10.1016/j.apm.2024.115680>.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2004. *Statistics of Extremes*. John Wiley & Sons.

Benjamin, M.A., Rigby, R.A., Stasinopoulos, D.M., 2003. Generalized autoregressive moving average models. *J. Amer. Statist. Assoc.* 98, 214–223. <http://dx.doi.org/10.1198/016214503388619238>.

Box, G., Jenkins, G.M., Reinsel, G., 2008. *Time series analysis*. In: *Wiley series in probability and statistics*, fourth ed. John Wiley & Sons.

Cardoso, I.P., Siqueira, T.M., Timm, L.C., Rodrigues, A.A., Nunes, A.B., 2022. Analysis of average annual temperatures and rainfall in southern region of the state of rio grande do sul, Brazil. *Rev. Bras. Ciências Ambient.* 57, 58–71. <http://dx.doi.org/10.5327/Z2176-94781204>.

Cecilio, R.A., Gonçalves, C.J., Zanetti, S.S., Abreu, M.C., de Almeida, L.T., 2021. Trends in monthly and annual streamflow related to rainfall and land use at the atlantic rainforest biome. *J. South Am. Earth Sci.* 112, 103600. <http://dx.doi.org/10.1016/j.jsames.2021.103600>.

Chandler, R., Scott, E.M., 2011. *Statistical Methods for Trend Detection and Analysis in the Environmental Sciences*, first ed. John Wiley & Sons.

Clarke, R.T., 1994. *Statistical Modelling in Hydrology*. John Wiley & Sons.

Clarke, R.T., 2002. Estimating trends in data from the Weibull and a generalized extreme value distribution. *Water Resour. Res.* 38, 25–1–25–10. <http://dx.doi.org/10.1029/2001WR000575>.

Costa, J.M.F.d., Silveira, C.d.S., Costa, A.C., Marcos Junior, A.D., Gonçalves, S.T.N., 2025. Trend analysis of precipitation extremes in Brazil: the role of atmospheric temperature. *Rev. Bras. Ciências Ambient.* 60, e2123. <http://dx.doi.org/10.5327/Z2176-94782123>.

Cox, D.R., Reid, N., 1987. Parameter orthogonality and approximate conditional inference. *J. R. Stat. Soc. Ser. B* 49, 1–39. <http://dx.doi.org/10.1111/j.2517-6161.1987.tb01422.x>.

D'Arcy, E., Tawn, J.A., 2025. Extremal properties of max-autoregressive moving average processes for modelling extreme river flows. *Extremes* 28, 677–711. <http://dx.doi.org/10.1007/s10687-025-00515-6>.

Davis, R.A., Resnick, S.I., 1989. Basic properties and prediction of max-ARMA processes. *Adv. in Appl. Probab.* 21, 781–803. <http://dx.doi.org/10.2307/1427767>.

Duca, V.E.L.A., Fonseca, T.C.O., Oliveira, F.L.C., 2021. A generalized dynamical model for wind speed forecasting. *Renew. Sustain. Energy Rev.* 136, <http://dx.doi.org/10.1016/j.rser.2020.110421>.

Dunn, P.K., Smyth, G.K., 1996. Randomized quantile residuals. *J. Comput. Graph. Statist.* 5, 236–244. <http://dx.doi.org/10.2307/1390802>.

Gradshteyn, I.S., Ryzhik, I.M., 2007. *Table of Integrals, Series, and Products, Seventh ed.* Academic Press, San Diego.

Helsel, D., Hirsch, R., Ryberg, K., Archfield, S., Gilroy, E., 2020. *Statistical Methods in Water Resources*. USGS, p. 458, chapter A3, book 4, *Techniques of Water-Resources Investigations of the United States Geological Survey*.

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., Yassmeen, F., 2022. *Forecast*. URL <https://CRAN.R-project.org/package=forecast>. R package version 8.16.

Justus, C.G., Hargraves, W.R., Mikhail, A., Graber, D., 1978. Methods for estimating wind speed frequency distributions. *J. Appl. Meteorol.* 17, 350–353. [http://dx.doi.org/10.1175/1520-0450\(1978\)017<0350:MFEWSF>2.0.CO;2](http://dx.doi.org/10.1175/1520-0450(1978)017<0350:MFEWSF>2.0.CO;2).

Kedem, B., Fokianos, K., 2005. *Regression Models for Time Series Analysis*, vol. 488. John Wiley & Sons.

Laipelt, L., de Paiva, R.C.D., Fan, F.M., Collischonn, W., Papa, F., Ruhoff, A., 2025. SWOT reveals how the 2024 disastrous flood in south Brazil was intensified by increased water slope and wind forcing. *Geophys. Res. Lett.* 52, e2024GL111287. <http://dx.doi.org/10.1029/2024GL111287>.

Langat, P.K., Kumar, L., Koech, R., 2019. Identification of the most suitable probability distribution models for maximum, minimum, and mean streamflow. *Water* 11, 734. <http://dx.doi.org/10.3390/w11040734>.

Lawrance, A.J., Kottegoda, N.T., 1977. Stochastic modelling of riverflow time series. *J. Roy. Statist. Soc. Ser. A* 140, 1. <http://dx.doi.org/10.2307/2344516>.

Li, Z., Brissette, F., Chen, J., 2013. Finding the most appropriate precipitation probability distribution for stochastic weather generation and hydrological modelling in Nordic watersheds. *Hydrol. Process.* 27, 3718–3729. <http://dx.doi.org/10.1002/hyp.9499>.

Li, J., Lei, Y., Tan, S., Bell, C.D., Engel, B.A., Wang, Y., 2018. Nonstationary flood frequency analysis for annual flood peak and volume series in both univariate and bivariate domain. *Water Resour. Manag.* 32, 4239–4252. <http://dx.doi.org/10.1007/s11269-018-2041-2>.

Li, W.K., McLeod, A.I., 1988. ARMA modelling with non-Gaussian innovations. *J. Time Series Anal.* 9, 155–168. <http://dx.doi.org/10.1111/j.1467-9892.1988.tb00461.x>.

Ljung, G.M., Box, G.E.P., 1978. On a measure of lack of fit in time series models. *Biometrika* 65, 297–303. <http://dx.doi.org/10.2307/2335207>.

Machiwal, D., Jha, M.K., 2012. *Hydrologic Time Series Analysis*. Springer, New Delhi.

Mahdi, S., Ashkar, F., 2004. Exploring generalized probability weighted moments, generalized moments and maximum likelihood estimating methods in two-parameter Weibull model. *J. Hydrol.* 258, 62–75. <http://dx.doi.org/10.1016/j.jhydrol.2003.08.012>.

Melo, M., Alencar, A., 2020. Conway–Maxwell–Poisson autoregressive moving average model for equidispersed, underdispersed, and overdispersed count data. *J. Time Series Anal.* 41, 830–857. <http://dx.doi.org/10.1111/jtsa.12550>.

Mishra, A.K., Desai, V.R., 2005. Drought forecasting using stochastic models. *Stoch. Environ. Res. Risk Assess.* 19, 326–339. <http://dx.doi.org/10.1007/s00477-005-0238-4>.

Moore, R.J., Clarke, R.T., 1981. A distribution function approach to rainfall runoff modeling. *Water Resour. Res.* 17, 1367–1382. <http://dx.doi.org/10.1029/WR017i005p01367>.

- Nimbal, J., Naik, R.L., Jangamshetti, S.H., 2012. Wind data analysis: A case study. In: 2012 International Conference on Power, Signals, Controls and Computation. IEEE, Thrissur, p. 5.
- Papalexiou, S.M., 2018. Unified theory for stochastic modelling of hydroclimatic processes: Preserving marginal distributions, correlation structures, and intermittency. *Adv. Water Resour.* 115, 234–252. <http://dx.doi.org/10.1016/j.advwatres.2018.02.013>.
- Petry, I., Miranda, P.T., Paiva, R.C.D., Collischonn, W., Fan, F.M., Fagundes, H.O., Araujo, A.A., Souza, S., 2025. Changes in flood magnitude and frequency projected for vulnerable regions and major wetlands of south america. *Geophys. Res. Lett.* 52, e2024GL112436. <http://dx.doi.org/10.1029/2024GL112436>.
- Piveta, M., 2024. Climate change and El Niño increased the frequency and intensity of rainfall in southern Brazil. *Rev. Pesqui. FAPESP* 341.
- Press, W., Teukolsky, S., Vetterling, W., Flannery, B., 1992. *Numerical Recipes in C, second ed.* Cambridge University Press, Cambridge.
- R. Development Core Team, 2024. R foundation for statistical computing. URL <http://www.R-project.org>.
- Ribeiro, T.F., Bayer, F.M., Peña-Ramírez, F.A., Guerra, R.R., Alencar, A.P., de Santana-e Silva, J.J., 2026. A dynamical regression model for double-bounded time series based on the reflected unit Burr XII distribution. *Environ. Ecol. Stat.* 33, 195–220. <http://dx.doi.org/10.1007/s10651-026-00703-y>.
- Rizwan, M., Guo, S., Xiong, F., Yin, J., 2018. Evaluation of various probability distributions for deriving design flood featuring right-tail events in Pakistan. *Water* 10, 1603. <http://dx.doi.org/10.3390/w10111603>.
- Rocha, A.V., Cribari-Neto, F., 2009. Beta autoregressive moving average models. *Test* 18, 529–545. <http://dx.doi.org/10.1007/s11749-008-0112-z>.
- Salas, J.D., 1992. Analysis and modeling of hydrologic time series. In: Maidment, D.R. (Ed.), *Handbook of Hydrology*. McGraw-Hill, pp. 19.1–19.72.
- Sarlak, N., 2008. Annual streamflow modelling with asymmetric distribution function. *Hydrol. Process.* 22, 3403–3409. <http://dx.doi.org/10.1002/hyp.6949>.
- Shaman, P., Stine, R.A., 1988. The bias of autoregressive coefficient estimators. *J. Amer. Statist. Assoc.* 83, 842–848. <http://dx.doi.org/10.1080/01621459.1988.10478672>.
- Silva, F.B., de Almeida, L.T., de Oliveira Vieira, E., da Silva, D.D., Maciel, I.P., Parma Júnior, F., 2020. Pluviometric and fluviometric trends in association with future projections in areas of conflict for water use. *J. Environ. Manag.* 271, 110991. <http://dx.doi.org/10.1016/j.jenvman.2020.110991>.
- Souza, S.A.d., Reis, Jr., D.S., 2022. Trend detection in annual streamflow extremes in Brazil. *Water* 14, <http://dx.doi.org/10.3390/w14111805>.
- Stedinger, J.R., Vogel, R.M., Foufoula-Georgiou, E., 1992. Frequency analysis of extreme events. In: Maidment, D.R. (Ed.), *Handbook of Hydrology*. McGraw-Hill, pp. 18.1–18.66.
- Tiku, M.L., Wong, W.K., Vaughan, D.C., Bian, G., 2000. Time series models in non-normal situations: Symmetric innovations. *J. Time Series Anal.* 21, 571–596. <http://dx.doi.org/10.1111/1467-9892.00199>.
- Trapletti, A., Hornik, K., 2020. Tseries: Time series analysis and computational finance. URL <https://CRAN.R-project.org/package=tseries>.
- Tuller, S.E., Brett, A.C., 1984. The characteristics of wind velocity that favor the fitting of a Weibull distribution in wind speed analysis. *J. Appl. Meteorol. Clim.* 23, 124–134. [http://dx.doi.org/10.1175/1520-0450\(1984\)023<0124:TCOWVT>2.0.CO;2](http://dx.doi.org/10.1175/1520-0450(1984)023<0124:TCOWVT>2.0.CO;2).
- Valipour, M., Banihabib, M.E., Behbahani, S.M.R., 2013. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of dez dam reservoir. *J. Hydrol.* 476, 433–441. <http://dx.doi.org/10.1016/j.jhydrol.2012.11.017>.
- Weibull, W.A., 1951. A statistical distribution of wide applicability. *J. Appl. Mech.* 18, 293–297.
- Yan, H., 2025. Hydrological regime shifts driven by climate and anthropogenic interference in the upper Paraná River Basin, Brazil. *J. South Am. Earth Sci.* 168, 105834. <http://dx.doi.org/10.1016/j.jsames.2025.105834>.