Revista Colombiana de Estadística
Special Issue: Colombian and Brazil Graduate Programs
December 2025, volume 48, issue 3, pp. 433–457
http://doi.org/10.15446/rce.v48n3.123621

A Mean-Parameterized Maxwell Time Series Model for Positive Data

Un modelo de series temporales Maxwell con parametrización en la media para datos positivos

Arthur H. da Rocha Hintz[1,a], Fernando A. Peña-Ramírez[1,2,b], Fábio M. Bayer[1,2,3,4,5,c]

[1]Department of Statistics, Centro de Ciências Naturais e Exatas, Universidade Federal de Santa Maria, Santa Maria, Brazil

[2]Graduate Program in Mathematics, Centro de Ciências Naturais e Exatas, Universidade Federal de Santa Maria, Santa Maria, Brazil

[3]Santa Maria Space Science Laboratory (LACESM), Universidade Federal de Santa Maria, Santa Maria, Brazil

[4]Graduate Program in Industrial Engineering, Centro de Tecnologia, Universidade Federal de Santa Maria, Santa Maria, Brazil

[5]Graduate Program in Statistics, Instituto de Matemática e Estatística, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

[6]Department of Mathematics and Natural Sciences, Blekinge Institute of Technology, Karlskrona, Sweden

Abstract

This paper introduces a new time series model for non-negative continuous data based on the Maxwell distribution. The proposed model employs a reparameterization of the Maxwell distribution in which its parameter directly represents the mean. In this formulation, the conditional mean is modeled through a dynamic structure that combines autoregressive and moving average components, linked by an appropriate link function. Parameter estimation is carried out using the conditional maximum likelihood method, for which closed-form matrix expressions of the conditional score vector and the conditional Fisher information matrix are derived. Based on the asymptotic properties of the estimators, procedures for interval estimation and hypothesis testing are presented. Monte Carlo simulations assess the finite-sample performance, and provide evidence of the estimators' convergence toward

[a]Undergraduate Student. E-mail: arthur.hintz@acad.ufsm.br
[b]Ph.D. E-mail: fernando.p.ramirez@ufsm.br
[c]Ph.D. E-mail: bayer@ufsm.br

the true parameter values as the sample size increases. An empirical application involving wind speed data from Brasília, the capital of Brazil, shows the practical relevance and effectiveness of the proposed model for real-world time series modeling and forecasting.

Keywords: ARMA models; Maxwell distribution; Mean parametrization; Positive time series.

Resumen

Este trabajo introduce un nuevo enfoque para modelar series temporales positivas y continuas basado en la distribución Maxwell. El modelo propuesto utiliza una reparametrización de dicha distribución, en la cual su parámetro de escala representa directamente la media. En esta formulación, la media condicional se modela mediante una estructura dinámica que combina componentes autorregresivos y de media móvil, vinculados a través de una apropiada función de enlace. La estimación de los parámetros se realiza mediante el método de máxima verosimilitud condicional, para el cual se derivan expresiones matriciales en forma cerrada del vector score condicional y de la matriz de información de Fisher. Con base en las propiedades asintóticas de los estimadores, se presentan procedimientos para la estimación por intervalos y pruebas de hipótesis. Mediante simulaciones de Monte Carlo se evalúa el desempeño en muestras finitas y se aporta evidencia de la convergencia de los estimadores hacia los valores verdaderos de los parámetros a medida que aumenta el tamaño muestral. Una aplicación empírica con datos de velocidad del viento en Brasilia, capital de Brasil, demuestra la relevancia práctica y la eficacia del modelo propuesto para el ajuste y predicción de series temporales reales.

Palabras clave: Distribución Maxwell; Modelos ARMA; Parametrización; Series de tiempo positivas.

1. Introduction

In time series analysis, several traditional models are widely employed to describe autocorrelated data observed over time. Among the most common approaches are the autoregressive integrated moving average (ARIMA) models (Box et al., 2015) and the family of exponential smoothing models, which includes the Holt–Winters (HW) method (Winters, 1960) and its state-space generalization through the Error–Trend–Seasonal (ETS) framework. However, these models assume that the variable of interest has support over the entire real line, which implies positive probability mass outside the domain of interest. As a consequence, they may generate forecasts that fall outside the feasible range of the observed data, for instance, negative predictions for quantities that are inherently non-negative. Furthermore, parametric models such as ARIMA typically rely on Gaussian assumptions for parameter inference, which may not hold in practice.

In this context, it is therefore useful to consider a new model that explicitly considers non-negative continuous support. The Maxwell distribution, also known as the Maxwell–Boltzmann distribution in reference to the pioneering works of

Maxwell (1867) and Boltzmann (1868), is a one-parameter probability distribution that describes the distribution of particle speeds in a thermal equilibrium gas, with a relatively simple functional form. The Maxwell distribution is defined exclusively for non-negative values and typically exhibits positive skewness, which makes it suitable for modeling positive-valued time series.

Since its introduction, the Maxwell distribution has been widely applied to a variety of physical and chemical phenomena. More recent studies have explored its use in statistical mechanics, as in Kazmi et al. (2011) and Kazmi et al. (2012). In survival analysis, the seminal work of Tyagi & Bhattacharya (1989) is considered a cornerstone. From a methodological standpoint, several authors have investigated inferential properties, parameter estimation methods, and reliability functions for this distribution, including Chaturvedi & Rani (1998), Bekker & Roux (2005), Dey & Maiti (2010), Radha & Venkatesan (2013), and Tomer & Panwar (2015). These studies have demonstrated that the Maxwell distribution is a flexible and versatile model, motivating its potential application in other domains. However, applications of the Maxwell distribution in time series modeling remain scarce, and this paper aims to provide an initial contribution in this direction.

This paper advances the literature by introducing a new time series model that incorporates the Maxwell distribution within an autoregressive moving average (ARMA)-type framework, referred to as the MxARMA model. The proposed model provides a novel alternative for modeling positive-valued serially dependent data and serves as a competitor to the generalized ARMA (GARMA) model (Benjamin et al., 2003), the Rayleigh ARMA (RARMA) model (Bayer et al., 2020), the Chen ARMA (CHARMA) model (Stone et al., 2023), and the Burr XII ARMA (UBXII-ARMA) model (de Araújo et al., 2024). All these models are designed to handle positive-valued time series but differ in their distributional assumptions and levels of complexity. Most of them, except for the RARMA, are based on more complex distributions indexed by two or more parameters. The RARMA model, on the other hand, relies on the one-parameter Rayleigh distribution, which is closely linked to radar data through its physical motivation. In this sense, the MxARMA model is a direct competitor to the RARMA model: it has the same level of parametric simplicity but is based on a distribution that offers different patterns and wider applicability across different domains.

Besides introducing the MxARMA model, we develop its inferential theory and discuss diagnostic and forecasting tools. A Monte Carlo simulation study is conducted to evaluate the finite-sample performance of the conditional likelihood inference. To illustrate the model's practical relevance, we analyze a wind speed time series, a positive, typically skewed, and autocorrelated process, characteristics well aligned with the MxARMA model's assumptions. Wind speed forecasting plays a crucial role in wind energy generation, a growing renewable energy sector that requires accurate short-term predictions for efficient operation. Finally, we also provide an R package implementing the proposed model, whose functionality is demonstrated in a dedicated Appendix.

The remainder of this paper is organized as follows. In Section 2, we introduce the proposed MxARMA model. Section 3 presents the conditional maximum like-

lihood theory, including closed-form expressions for the conditional score vector, the conditional Fisher information matrix, and related asymptotic results. Section 4 addresses model diagnostics and forecasting procedures. Section 5 contains a Monte Carlo simulation study to evaluate the finite-sample performance of the proposed estimator. Section 6 illustrates the application of MxARMA models to wind speed data, while Section 7 concludes the paper. The Appendix includes supplementary simulation results, mathematical details, and a comprehensive introduction to the `MxARMA` R package.

2. The proposed model

This section introduces a novel time series model based on the Maxwell distribution. A non-negative continuous random variable is said to follow the Maxwell law if its probability density function (pdf) is given by Lemonte (2024)

$$f(y;\lambda) = \sqrt{\frac{2}{\pi}}\frac{y^2}{\lambda^3}\exp\left(-\frac{y^2}{2\lambda^2}\right), \quad y > 0,$$

where $\lambda > 0$ is the scale parameter. Following the reparametrization introduced by Lemonte (2024), we express λ in terms of the mean, where $E(Y) = \mu = 2\lambda\sqrt{2/\pi}$, i.e., $\lambda = (\mu/2)\sqrt{\pi/2}$. We denote this as $Y \sim \text{Mx}(\mu)$ to emphasize the mean-based parameterization. Building upon this reparameterization, we define an ARMA structure by assuming that the conditional distribution of each time series observation follows a Maxwell distribution with a time-varying mean.

Let $\{Y_t\}_{t\in\mathbb{Z}}$ be a stochastic process, where each Y_t, conditioned on the information set $\mathcal{F}_{t-1}$, is distributed according to $\text{Mx}(\mu_t)$. Here, $\mathcal{F}_{t-1}$ denotes the intersection of all $\sigma-$fields with respect to which the random variables $\{Y_1, \ldots, Y_{t-1}\}$ are measurable. This represents the smallest $\sigma-$field containing all the information available up to time $t-1$. Under this conditional structure, the pdf and cumulative distribution function of Y_t are given, respectively, by

$$f(y_t;\mu_t|\mathcal{F}_{t-1}) = \frac{32y_t^2}{\pi^2\mu_t^3}\exp\left(-\frac{4y_t^2}{\pi\mu_t^2}\right), \quad y_t > 0, \tag{1}$$

and

$$F(y_t;\mu_t|\mathcal{F}_{t-1}) = \frac{2\gamma(3/2, 4y_t^2/(\pi\mu_t^2))}{\sqrt{\pi}},$$

where $\mu_t > 0$ and $\gamma(a,x) = \int_0^x t^{a-1}\text{e}^{-t}\text{d}t$ is the lower incomplete gamma function. The conditional variance of Y_t, in terms of μ_t, is $\text{Var}(Y_t|\mathcal{F}_{t-1}) \approx 0.178\mu_t^2$.

Figure 1 compares the pdfs of the reparameterized Maxwell distribution (see Equation (1)) and the mean-parameterized Rayleigh distribution, given by Palm et al. (2019):

$$f_R(y;\mu) = \frac{\pi y}{2\mu^2}\exp\left(-\frac{\pi y^2}{4\mu^2}\right), \quad y > 0, \tag{2}$$

for different values of the mean parameter (0.2, 0.5, and 1.0). It can be observed that the Rayleigh distribution exhibits stronger right skewness and heavier tails than the Maxwell distribution. This behavior is associated with its relatively larger variance, given by $0.273\mu^2$, which is approximately 53% higher than that of the Maxwell distribution with the same mean. As a result, the Rayleigh distribution tends to allocate more probability mass to neighborhood intervals near zero and at higher values. In contrast, the Maxwell distribution concentrates its density more tightly around the mean.

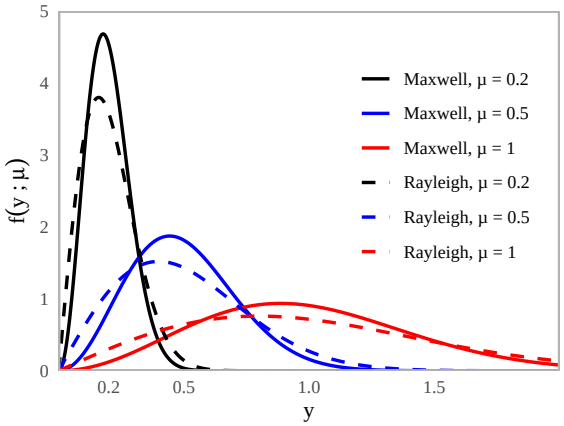


FIGURE 1: Probability density functions of the mean-parameterized Maxwell (solid line) and Rayleigh (dashed line) distributions for different mean values $\mu \in \{0.2, 0.5, 1.0\}$.

To define the dynamic structure of the model, we consider a linear predictor represented as $\eta_t = g(\mu_t)$, where $g(\cdot)$ is an invertible, monotonic and twice differentiable link function, such that $g : \mathbb{R}^+ \to \mathbb{R}$. Following Benjamin et al. (2003), we adopt the following functional relationship:

$$\eta_t = g(\mu_t) = \alpha + \boldsymbol{x}_t^\top \boldsymbol{\beta} + \tau_t, \tag{3}$$

where α is an intercept term, $\boldsymbol{x}_t = (x_{t1}, \ldots, x_{tk})^\top$ is a k-dimensional vector of known covariates at time t, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^\top$ is the corresponding k-dimensional vector of unknown parameters, and τ_t is a dynamic component that captures ARMA-type dependence, included additively in the linear predictor. This component can be expressed as

$$\tau_t = \sum_{i=1}^{p} \phi_i \left(g(y_{t-i}) - \boldsymbol{x}_{t-i}^\top \boldsymbol{\beta} \right) + \sum_{j=1}^{q} \theta_j r_{t-j},$$

where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^\top$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^\top$ are the vectors of autoregressive (AR) and moving average (MA) parameters, respectively. The MA error term is defined recursively on the prediction scale as $r_t = g(y_t) - g(\mu_t)$.

We define the Maxwell ARMA model, denoted as MxARMA(p, q), using Equations (1) and (3), which describe the stochastic specification and systematic struc-

ture of the model, respectively. In practice, the model parameters are unknown and require estimation. The following section addresses the inference theory.

One important property in time series modeling concerns the conditions for stationarity. Using similar arguments to those in Benjamin et al. (2003) for the canonical exponential family, it is possible to establish conditions for the stationarity of the mean and variance only when the identity link function is used. However, for other link functions, such as the log link, the first two moments of the marginal distribution appear to be analytically intractable. As a result, deriving stationarity conditions for this class of dynamic models remains a theoretical problem still open in the literature and lies beyond the scope of the present work. Nevertheless, as is common in the related literature (Rocha & Cribari-Neto, 2009; Bayer et al., 2017; Melo & Alencar, 2020; Stone et al., 2023), inference can be reliably conducted via conditional maximum likelihood without explicitly imposing or verifying marginal stationarity conditions, since the conditional score vector and information matrix remain well-behaved under mild regularity conditions.

3. Conditional Likelihood Inference

In this section, we present the parameter estimation procedure for the MxARMA model based on the conditional maximum likelihood method. The conditional maximum likelihood estimators (CMLEs) are obtained by maximizing the logarithm of the conditional likelihood function under the MxARMA dynamics. Given an observed sample $y_1, \ldots, y_T$ from a sequence of random variables following an MxARMA(p, q) process, that is, according to Equations (1) and (3), with parameter vector $\boldsymbol{\gamma} = \left(\alpha, \boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \boldsymbol{\theta}^\top\right)^\top$ of dimension $(r + p + q + 1)$, the conditional log-likelihood function, excluding the first $m = \max(p, q)$ observations, is expressed as

$$\ell(\boldsymbol{\gamma}) = \sum_{t=m+1}^{T} \log f(y_t; \mu_t \mid \mathcal{F}_{t-1}) = \sum_{t=m+1}^{T} \ell_t(\mu_t), \tag{4}$$

where

$$\ell_t(\mu_t) = \log(32) - 2\log(\pi) - 3\log(\mu_t) + 2\log(y_t) - \frac{4y_t^2}{\pi\mu_t^2}.$$

It is worth noting that the function $\ell_t(\mu_t)$ has a particularly simple form, allowing closed-form derivations of both the conditional score vector and the conditional Fisher information matrix. This analytical simplicity is one of the main advantages of the proposed model, facilitating parameter estimation and enabling straightforward large-sample inference for the model parameters.

3.1. Conditional Score Vector

In a similar fashion to Lemonte (2024) and Bayer et al. (2020), the conditional score vector is obtained by differentiating the conditional log-likelihood Function (4) with respect to each j-th element of the parameter vector $\boldsymbol{\gamma}$, for $j = 1, \ldots, (r + p + q + 1)$.

It is given by the gradient of the log-likelihood, obtained by the chain rule as

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

where

$$\frac{\partial \ell_t(\mu_t)}{\partial \mu_t} = \frac{8y_t^2}{\mu_t^3 \pi} - \frac{3}{\mu_t} = c_t, \quad \frac{d\mu_t}{d\eta_t} = \frac{1}{g'(\mu_t)}.$$

In this case, the logarithmic function is the most appropriate choice for the link function $g(\cdot)$, whose derivative is $g'(\mu_t) = 1/\mu_t$. The derivatives of η_t with respect to each element of $\boldsymbol{\gamma}$ are given by

$$\begin{aligned}
\frac{\partial \eta_t}{\partial \alpha} &= 1 - \sum_{j=1}^{q} \theta_j \frac{\partial \eta_{t-j}}{\partial \alpha}, \\
\frac{\partial \eta_t}{\partial \beta_i} &= x_{ti} - \sum_{i=1}^{p} \phi_i x_{(t-j)i} - \sum_{j=1}^{q} \theta_j \frac{\partial \eta_{t-j}}{\partial \beta_i}, i = 1, \ldots, k, \\
\frac{\partial \eta_t}{\partial \phi_i} &= g(y_{t-i}) - x_{t-i}\boldsymbol{\beta} - \sum_{j=1}^{q} \theta_j \frac{\partial \eta_{t-j}}{\partial \phi_i}, i = 1, \ldots, p, \\
\frac{\partial \eta_t}{\partial \theta_i} &= r_{t-i} - \sum_{j=1}^{q} \theta_j \frac{\partial \eta_{t-i}}{\partial \theta_i}, i = 1, \ldots, q.
\end{aligned}$$

A matrix representation of the score vector provides a compact formulation that unifies the theoretical framework while facilitating computational implementation via vectorized operations. Therefore, the score vector assumes the matrix form

$$\boldsymbol{U}(\boldsymbol{\gamma}) = \left(\boldsymbol{U}_{\alpha}(\boldsymbol{\gamma}), \boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\gamma})^{\top}, \boldsymbol{U}_{\boldsymbol{\phi}}(\boldsymbol{\gamma})^{\top}, \boldsymbol{U}_{\boldsymbol{\theta}}(\boldsymbol{\gamma})^{\top}\right)^{\top},$$

where

$$\boldsymbol{U}_{\boldsymbol{\alpha}}(\boldsymbol{\gamma}) = \boldsymbol{v}^{\top}\boldsymbol{T}\boldsymbol{c}, \quad \boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}) = \boldsymbol{M}^{\top}\boldsymbol{T}\boldsymbol{c}, \quad \boldsymbol{U}_{\boldsymbol{\phi}}(\boldsymbol{\gamma}) = \boldsymbol{P}^{\top}\boldsymbol{T}\boldsymbol{c}, \quad \boldsymbol{U}_{\boldsymbol{\theta}}(\boldsymbol{\gamma}) = \boldsymbol{R}^{\top}\boldsymbol{T}\boldsymbol{c},$$

with

$$\mathbf{T} = \operatorname{diag}\left(\frac{1}{g'(\mu_{m+1})}, \cdots, \frac{1}{g'(\mu_T)}\right), \quad \boldsymbol{v} = \left(\frac{\partial \eta_{m+1}}{\partial \alpha}, \ldots, \frac{\partial \eta_T}{\partial \alpha}\right)^{\top}, \quad \boldsymbol{c} = (c_{m+1}, \ldots, c_T)^{\top}.$$

Moreover, the matrices $\boldsymbol{M}, \boldsymbol{P}$, and $\boldsymbol{R}$ have dimension $(T-m) \times k$, $(T-m) \times p$, and $(T-m) \times q$, respectively, whose (i,j)-th elements are given by

$$M_{i,j} = \frac{\partial \eta_{i+m}}{\partial \beta_j}, \quad P_{i,j} = \frac{\partial \eta_{i+m}}{\partial \phi_j}, \quad R_{i,j} = \frac{\partial \eta_{i+m}}{\partial \theta_j}.$$

The CMLEs, $\hat{\boldsymbol{\gamma}}$, of the parameter vector $\boldsymbol{\gamma}$ are obtained by solving the nonlinear system $\boldsymbol{U}(\boldsymbol{\gamma}) = \mathbf{0}$, where $\mathbf{0}$ is the null vector of dimension $(r+p+q+1)$. Since this system does not have a closed-form solution, numerical optimization techniques are required. In this study, we employ the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm (Vetterling & Press, 1992), a quasi-Newton method that uses analytical first-order derivatives. This iterative algorithm requires initial values for all parameters. The starting values for the intercept (α), the AR coefficients ($\boldsymbol{\phi}$), and the regression parameters ($\boldsymbol{\beta}$) are obtained via ordinary least squares estimation from a linear regression, where the response vector is $\mathbf{Y} = (g(y_{m+1}), g(y_{m+2}), \ldots, g(y_T))^\top$. For the MA parameters ($\boldsymbol{\theta}$), the initial values are set to zero, as commonly adopted in related works (Bayer et al., 2017, 2020).

3.2. Confidence Intervals and Hypothesis Testing Inference

The asymptotic properties of the CMLEs (Andersen, 1970; Fahrmeir & Kaufmann, 1985) imply that the estimator has the following asymptotic distribution:

$$\hat{\boldsymbol{\gamma}} \xrightarrow{d} \mathcal{N}\big(\boldsymbol{\gamma}, \boldsymbol{K}^{-1}(\boldsymbol{\gamma})\big),$$

where $\boldsymbol{K}(\boldsymbol{\gamma})$ denotes the conditional Fisher information matrix, as defined in Equation (A1) in Appendix A. That is, the estimator $\hat{\boldsymbol{\gamma}}$ converges in distribution to a multivariate normal random vector with mean $\boldsymbol{\gamma}$ and covariance matrix $\boldsymbol{K}^{-1}(\boldsymbol{\gamma})$. Based on this large-sample result, interval estimation and hypothesis testing for the model parameters can be straightforwardly performed.

Let us consider the following set of hypotheses for an individual parameter of the model:

$$\begin{cases} \mathcal{H}_0 : \gamma_j = \gamma_j^0, \\ \mathcal{H}_1 : \gamma_j \neq \gamma_j^0, \end{cases}$$

where γ_j denotes the j-th element of $\boldsymbol{\gamma}$, and γ_j^0 is the hypothesized value of the parameter under the null hypothesis. The corresponding test statistic is obtained by standardizing the estimator:

$$Z = \frac{\hat{\gamma}_j - \gamma_j^0}{\mathrm{se}(\hat{\gamma}_j)},$$

which follows approximately a standard normal distribution, $\mathcal{N}(0,1)$, under $\mathcal{H}_0$. Here, $\mathrm{se}(\hat{\gamma}_j)$ denotes the standard error of $\hat{\gamma}_j$, computed as the square root of the (j,j)-th diagonal element of $\boldsymbol{K}^{-1}(\boldsymbol{\gamma})$ evaluated at the estimated parameter values.

Let ω denote the significance level of the test. Thus, the test can be performed by comparing the calculated Z statistic with the usual quantiles of the standard normal distribution. The null hypothesis is rejected when the absolute value of Z exceeds the upper (1-ω/2) quantile of the standard normal distribution.

Based on the asymptotic normal distribution of the CMLEs, we can construct a $100(1-\omega)\%$ confidence interval for each γ_j, which is given by

$$\left[\hat{\gamma}_j - z_{1-\omega/2} \cdot \mathrm{se}(\hat{\gamma}_j) \quad ; \quad \hat{\gamma}_j + z_{1-\omega/2} \cdot \mathrm{se}(\hat{\gamma}_j)\right],$$

where $z_{1-\omega/2}$ is the $1-\omega/2$ quantile of the standard normal distribution.

4. Diagnostic Analysis and Forecasting

In this section, we present diagnostic measures and model selection criteria that are essential for evaluating how well a fitted model captures the underlying data dynamics. Quantile residuals (Dunn & Smyth, 1996) are particularly recommended because of their advantages over other residual types in non-Gaussian settings, as they approximately follow a standard normal distribution when the model is correctly specified. Accordingly, the quantile residual is defined as

$$e_t = \Phi^{-1}\left(F(y_t; \hat{\mu}_t | \mathcal{F}_{t-1})\right),$$

where $\Phi^{-1}(\cdot)$ denotes the standard normal quantile function, and $\hat{\mu}_t$ is given in Equation (5). A well-fitted model should produce residuals that behave as white noise, exhibiting zero mean and constant variance. The Ljung-Box test (Ljung & Box, 1978) can be employed to test the null hypothesis of no autocorrelation in the residual time series.

For selecting the order (p, q) and comparing different fitted models, model selection criteria such as the Akaike information criterion (AIC) (Akaike, 1974) and the Bayesian information criterion (BIC) (Konishi & Kitagawa, 2008) can be employed. These criteria are computed from the maximized conditional log-likelihood function, with an added penalty term for model complexity. Accordingly, the model with the lowest AIC or BIC value is preferred, as it provides the best trade-off between goodness of fit and parsimony.

After order selection, parameter estimation, and diagnostic analysis, the model can be used to generate predictions. Time series forecasting with ARMA-type models is a well-established approach for predicting future values based on historical observations (Brockwell & Davis, 1991). In-sample predictions within the MxARMA(p, q) framework are obtained by estimating $\hat{\mu}_{m+1}, \ldots, \hat{\mu}_T$ using the CMLEs $\hat{\boldsymbol{\gamma}}$ in Equation (3). The recursive prediction structure, starting at $t = m + 1$, is given by

$$\hat{\mu}_t = g^{-1}\left(\hat{\alpha} + \boldsymbol{x}_t^\top \hat{\boldsymbol{\beta}} + \sum_{i=1}^{p} \hat{\phi}_i \left[g(y_{t-i}) - \boldsymbol{x}_{t-i}^\top \hat{\boldsymbol{\beta}}\right] + \sum_{j=1}^{q} \hat{\theta}_j \hat{r}_{t-j}\right), \tag{5}$$

where $\hat{r}_t = g(y_t) - g(\hat{\mu}_t)$ for $m < t \leq T$, and $\hat{r}_t = 0$ for $t \leq m$. Similarly, out-of-sample forecast at a future time point $T+h$, where h denotes the forecast horizon, with $h = 1, \ldots, H$, can be computed as

$$\hat{\mu}_{T+h} = g^{-1}\left(\hat{\alpha} + \boldsymbol{x}_{T+h}^\top \hat{\boldsymbol{\beta}} + \sum_{i=1}^{p} \hat{\phi}_i \left[g^*(y_{T+h-i}) - \boldsymbol{x}_{T+h-i}^\top \hat{\boldsymbol{\beta}}\right] + \sum_{j=1}^{q} \hat{\theta}_j \hat{r}_{T+h-j}\right).$$

For out-of-sample forecasts, the covariates $\boldsymbol{x}_t$ must be also available, which is not a problem if they represent seasonal or trend patterns, for example. For $t > T$, the transformed responses $g(y_t)$ are unobserved and therefore $g^*(y_t)$ is replaced by their predicted values $g(\hat{\mu}_t)$. Regarding the moving-average component, since

$g(y_t)$ is replaced by $g(\hat{\mu}_t)$ for $t > T$, the corresponding error terms in the out-of-sample forecasting stage become $\hat{r}_t = g(\hat{\mu}_t) - g(\hat{\mu}_t) = 0$. Thus, the MA component has no further contribution beyond the observed sample. This procedure enables the iterative computation of $\hat{\mu}_{T+h}$ for all $h > 0$.

To assess the empirical forecasting performance of the MxARMA model and benchmark it against other fitted models, we use the mean squared error (MSE) and the mean absolute percentage error (MAPE). These metrics are respectively defined as

$$\text{MSE} = \frac{1}{T-m}\sum_{t=m+1}^{T}(y_t - \hat{\mu}_t)^2, \qquad \text{and} \qquad \text{MAPE} = \frac{1}{T-m}\sum_{t=m+1}^{T}\left|\frac{y_t - \hat{\mu}_t}{y_t}\right|,$$

where y_t denotes the observed value and $\hat{\mu}_t$ the predicted value at time t, for $t = 1, \ldots, T$. These measures are widely used in statistical modeling to quantify the discrepancy between observed and predicted values, allowing the evaluation and comparison of model performance (Hyndman & Koehler, 2006).

In a similar way as in Benjamin et al. (2003) for GARMA models, simulation techniques can be used to construct prediction intervals (PI). Fixing $\hat{\boldsymbol{\gamma}}$ and the known information set up to time T, B simulated sample paths of y_{T+h}, for $h \geq 1$, yield an empirical predictive distribution of $Y_{T+h} \mid \mathcal{F}_T$, from which prediction intervals can be obtained. Algorithmically, the procedure can be described as follows:

1. Given the observed time series $\mathbf{y} = (y_1, \ldots, y_T)^\top$, obtain the estimate of the parameter vector $\hat{\boldsymbol{\gamma}}$ and compute the one-step-ahead prediction $\hat{\mu}_{T+1}$ using (5);
2. Generate y_{T+1} as a pseudo-random draw from the distribution $\text{Mx}(\hat{\mu}_{T+1})$ using the inversion method;
3. Update the time series to $\mathbf{y}^* = (y_1, \ldots, y_T, y_{T+1})^\top$;
4. Using $\mathbf{y}^*$ and $\hat{\boldsymbol{\gamma}}$ from the Step (1), compute the next prediction $\hat{\mu}_{T+2}$ according to (5);
5. Generate $y_{T+2} \sim \text{Mx}(\hat{\mu}_{T+2})$ and update the time series to $\mathbf{y}^* = (y_1, \ldots, y_T, y_{T+1}, y_{T+2})^\top$;
6. Repeat the previous steps iteratively until reaching
$$\mathbf{y}^* = (y_1, \ldots, y_T, y_{T+1}, y_{T+2}, \ldots, y_{T+H})^\top;$$
7. Repeat Steps 2–6 a total of B times, producing B simulated paths;
8. For each forecast horizon $h = 1, \ldots, H$, compute the empirical quantiles $\omega/2$ and $1 - \omega/2$ from the B simulated values of y_{T+h};
9. Construct the prediction interval for horizon h using the empirical $\omega/2$ quantile as the lower limit and the $1 - \omega/2$ quantile as the upper limit, yielding a $(1 - \omega)$ confidence level prediction interval.

5. Numerical Evaluation

This section presents a Monte Carlo simulation study conducted to evaluate the performance of the CMLEs for the MxARMA model parameters. The simulations were based on 5000 Monte Carlo replications, considering sample sizes $T \in \{50, 100, 300, 500, 1000\}$ and setting the log link function. All implementations were performed in the R programming language (R Core Team, 2024), and the corresponding code is available at `https://github.com/arthurhintz/MxARMA`. Additional implementation details and the developed R package are discussed in Appendix C.

To assess the CMLEs performance, we expect that, as the sample size increases, the percentage relative bias (RB%) and mean square error (MSE) decrease toward zero, being numerical evidence of the consistency of the estimators. The RB% is defined as the ratio between the bias and the true parameter value, multiplied by 100, and the MSE is calculated as the variance of the estimator plus the squared bias. To evaluate the confidence intervals, we calculated the coverage rates (CRs) as the proportion of intervals obtained from Monte Carlo replications that contain the true parameter, using a nominal confidence level of 95%. Ideally, this proportion should be as close as possible to the nominal confidence level.

As defined in Equation (3), the model can be specified either with or without covariates. In this section, and throughout the remainder of the paper, we use the notation MxARMAX to denote MxARMA models that include regressors, while MxARMA refers to models without regressors. In the simulation study, we consider four scenarios: MxARMAX(1,1), MxARMA(1,1), MxARMAX(0,1), and MxARMAX(1,0). Our discussion here focuses on the MxARMAX(1,1) scenario, which is the most complex one, while simulation results for the other three configurations are provided in Appendix B.

In the first scenario, the series y_t was generated according to the dynamic structure of the MxARMAX(1,1) model, which includes a sinusoidal covariate defined as $x_t = \sin(2\pi t/12)$. The parameter values were set to $\alpha = 0.5$, $\phi_1 = 0.45$, $\theta_1 = -0.3$, and $\beta_1 = 0.6$, following the configuration adopted by Bayer et al. (2020) in the context of RARMA model. The results for this scenario are summarized in Table 1. The mean CMLEs obtained from the Monte Carlo replications converge toward the true parameter values as the sample size increases. Similarly, the CR approaches the nominal confidence level with larger samples, confirming the adequacy of inferences based on the conditional Fisher information matrix. Overall, the estimator of θ_1 exhibits higher bias compared to the others, a well-documented behavior in the literature for models involving moving average components (Ansley & Newbold, 1980; Palm & Bayer, 2018). In contrast, the covariate parameter estimator exhibits very low bias, even for relatively small sample sizes.

Figure 2 summarizes the results of all scenarios in terms of RB, consistent with the findings reported in Appendix B. As expected, the RB of all estimators across all scenarios approach zero as the sample size increases. In particular, the bias of the estimator of the parameter β remains small even for small samples, and the inclusion of the regressor further reduces the bias of the other parameters. Overall,

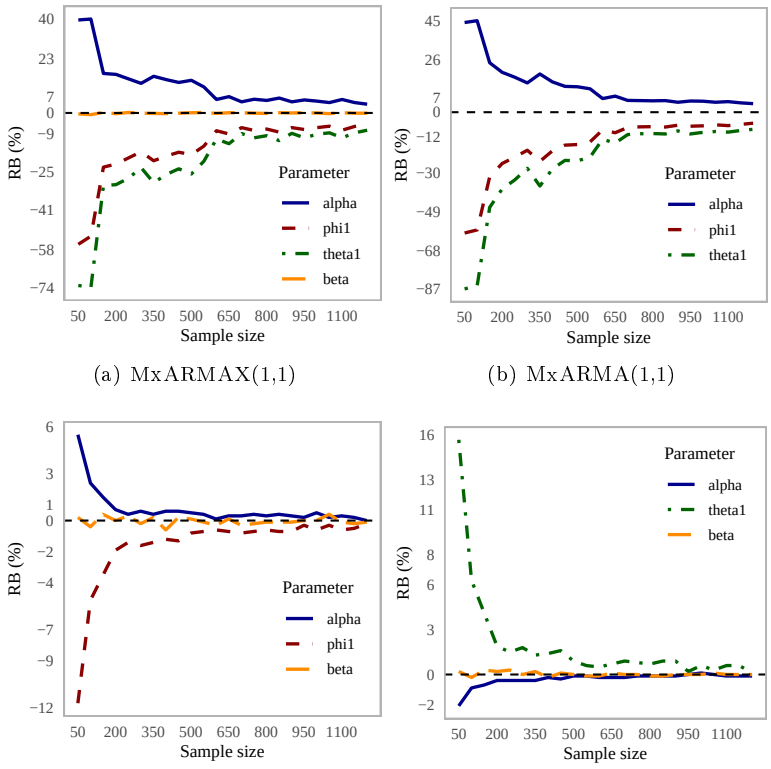
the numerical results validate the proposed methodology and its implementation, confirming the practical feasibility of applying the MxARMA model in real-world problems.

Table 1: Monte Carlo simulation results for the MxARMAX(1,1) model.

Parameter	T	Mean	RB(%)	MSE	CR
$\alpha = 0.5$	50	0.6923	38.4665	0.1771	0.8076
	100	0.6943	38.8604	0.1891	0.7742
	300	0.5704	14.0778	0.0690	0.8588
	500	0.5581	11.6165	0.0497	0.8838
	1000	0.5206	4.1125	0.0175	0.9140
	2000	0.5116	2.3109	0.0082	0.9330
$\beta = 0.6$	50	0.5960	−0.6711	0.0098	0.9264
	100	0.6009	0.1512	0.0049	0.9312
	300	0.5991	−0.1522	0.0015	0.9398
	500	0.5999	−0.0150	0.0009	0.9462
	1000	0.6001	0.0224	0.0005	0.9386
	2000	0.6004	0.0681	0.0002	0.9524
$\phi_1 = 0.45$	50	0.2091	−53.5437	0.2496	0.8034
	100	0.2182	−51.5175	0.2568	0.7756
	300	0.3672	−18.4068	0.0897	0.8600
	500	0.3827	−14.9603	0.0644	0.8806
	1000	0.4262	−5.2892	0.0222	0.9138
	2000	0.4365	−3.0001	0.0105	0.9348
$\theta_1 = -0.3$	50	−0.0874	−70.8718	0.2660	0.7838
	100	−0.0788	−73.7343	0.2534	0.7638
	300	−0.2215	−26.1831	0.0945	0.8538
	500	−0.2343	−21.9067	0.0655	0.8786
	1000	−0.2776	−7.4645	0.0247	0.9146
	2000	−0.2870	−4.3289	0.0117	0.9298

6. Real Data Application

To illustrate the practical applicability of the proposed model, we analyze a wind speed time series from Brasília, Brazil. Wind speed data are positive, typically skewed, and exhibit temporal dependence, characteristics that align well with the assumptions of the proposed model. In recent years, wind energy has demonstrated the fastest growth among renewable energy sources worldwide, with gigawatt-scale power generation now feasible through modern conversion technologies (International Renewable Energy Agency (IRENA), 2024). Within this global context, Brazil stands out as a leading producer of wind power in South America, with an estimated technical potential of 3.1 terawatts, approximately three times the total installed electricity generation capacity of all European Union countries combined (Energy Sector Management Assistance Program, 2019). Wind turbine energy production depends directly on wind speed, which remains one of the most difficult meteorological variables to forecast due to its stochastic and highly intermittent nature (Douak et al., 2013; Khosravi et al., 2018).



(a) MxARMAX(1,1) (b) MxARMA(1,1)

(c) MxARMAX(1,0) (d) MxARMAX(0,1)

FIGURE 2: Relative biases of all estimators across all simulated scenarios.

In this context, time series models emerge as valuable tools for addressing this challenge. For instance, the RARMA model, proposed by Bayer et al. (2020), has been successfully applied to wind speed data. Building on this idea, the proposed MxARMA model provides a suitable alternative for modeling data of this nature, as it is defined on the positive real line and, like the RARMA model, is based on a one-parameter distribution. In this study, we analyze the monthly average wind speed time series recorded at meteorological station 83377 in Brasília, Brazil, covering the period from January 2001 to December 2018. Brasília, located on the Central Plateau of Brazil, exhibits distinct climatic characteristics shaped by its high altitude and well-defined seasonal cycle, with alternating rainy and dry periods. Wind speed plays a key role in the local climate, influencing not only perceived temperature but also the dispersion of pollutants and other atmospheric processes.

The dataset contains 216 monthly observations, with an overall mean of 2.1012 m/s and a standard deviation of 0.6261 m/s, indicating moderate variability in wind speed over time. The black line in Figure 3 shows the observed time series, while descriptive statistics are summarized in Table 2. The average wind speed reaches its highest values in August and September, exceeding 2.8 m/s, coinciding

with the peak of the dry season, when stronger winds prevail. Conversely, during March and November, a marked reduction is observed, with mean speeds around 1.8 m/s. Seasonality is confirmed by the Kruskal–Wallis test, which rejects the null hypothesis of no seasonal effect (p-value < 0.0001). Finally, the Augmented Dickey–Fuller test rejects the null hypothesis of a unit root (p-value $= 0.01$), indicating that the series is stationary.

Table 2: Summary statistics for wind speed data.

T	Mean	SD	Median	Minimum	Maximum	Range	Kurtosis
216	2.1012	0.6261	2.0367	0.6921	4.5054	3.8133	1.3033

For model selection, we fitted MxARMA(p, q) models with different AR and MA orders, identifying those with the lowest AIC and BIC values. Seasonality was incorporated into the model through a covariate. The series was decomposed into seasonal, trend, and irregular components using moving averages, following the additive formulation $Y_t = T_t + S_t + e_t$, commonly employed in time series analysis (e.g., Kendall, 1948). The decomposed seasonal component S_t was then included in the mean structure of the model described in Equation (3) as an explanatory covariate. The last 12 observations are reserved for forecast evaluation only.

For comparison purposes, the ETS, HW, and RARMA models were also fitted. The ETS model was estimated under the usual (A, N, A) configuration — additive errors, no trend, and additive seasonality — and the HW model employed an additive seasonal formulation. The RARMA model is defined using the pdf in Equation (2) as its stochastic component and Equation (3) as its systematic structure. Accordingly, the same procedure was applied to identify the best-fitting specification for this model, and the notation RARMAX was adopted to denote RARMA models that include regressors.

Table 3 presents the selected fitted MxARMAX(1,1) and RARMAX(3,2) models, both including a seasonal covariate. The Ljung–Box test applied to the quantile residuals of the fitted MxARMAX model indicates no significant autocorrelation (p-value $= 0.6413$), confirming the adequacy of the model. Figure 4 displays the sample autocorrelation function (ACF) and partial ACF (PACF) for both the original time series and the quantile residuals, further supporting the model validity. Similarly, for the RARMAX(3,2) model, the Ljung–Box test (p-value $= 0.2061$) also suggests that residuals are uncorrelated. In the MxARMAX model, the intercept term was not statistically significant, and in the RARMAX model, the θ_1 component was also nonsignificant at usual levels. Nevertheless, both parameters were retained, since the primary goal of this analysis is predictive performance rather than inference and interpretations.

Figure 3 displays the observed time series along with the fitted values from the MxARMAX(1,1) model, represented by the dashed line. A 12-step-ahead out-of-sample forecast, together with its prediction interval with $B = 1000$ and $\omega = 5\%$, is also presented, appearing after the dotted vertical line. Overall, the MxARMAX model effectively captures the main dynamics of the data and provides accurate out-of-sample forecasts. Table 4 summarizes the in-sample and out-of-sample prediction performance metrics for all competing models.

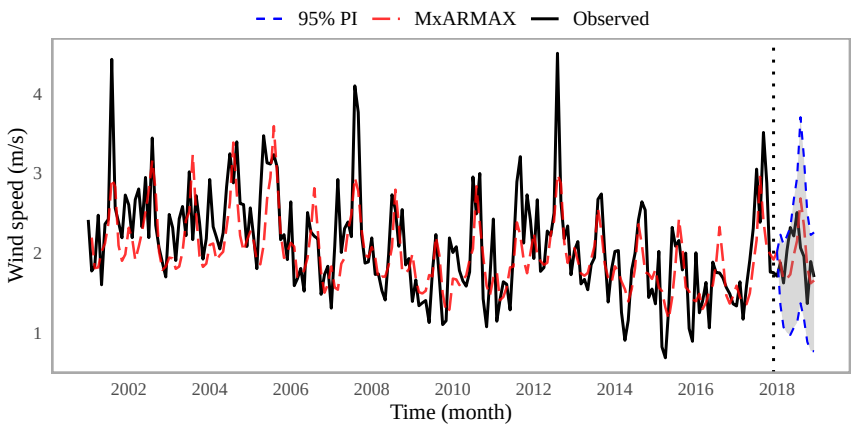
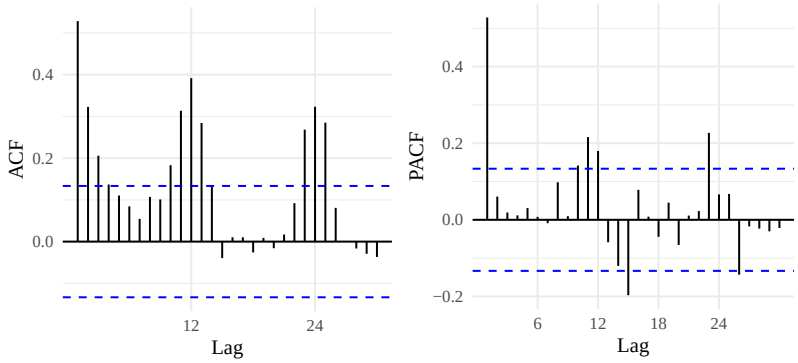
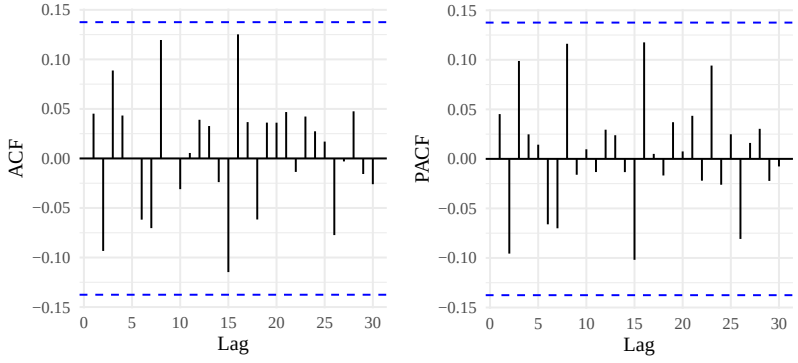


FIGURE 3: Observed (solid line) and fitted (dashed line) monthly wind speed series of Brasília.



(a) ACF of y (b) PACF of y



(c) ACF of residuals (d) PACF of residuals

FIGURE 4: Correlograms of wind speed time series and residuals of the fitted MxARMA model.

According to Table 4, the MxARMAX(1,1) model achieved the lowest MAPE values both in-sample and out-of-sample, indicating superior predictive accuracy compared to the alternative models. Moreover, its out-of-sample MSE was the smallest among all methods, confirming that the proposed model produced forecasts closest to the observed wind speed measurements over the twelve-month forecast horizon. Therefore, we conclude that the MxARMAX(1,1) model is the most appropriate option for these data among the four competing models evaluated, being a good alternative for modeling positive time series.

Table 3: Fitted MxARMAX and RARMAX models for the wind speed data in Brasília.

MxARMAX(1,1)				
	Estimate	Standard error	Z stat	p-value
α	0.0666	0.0607	1.0977	0.2724
ϕ_1	0.8841	0.0922	9.5904	≤ 0.0001
θ_1	-0.6143	0.1702	3.6099	≤ 0.0003
β_1	0.4560	0.0856	5.3261	≤ 0.0001
AIC = 383.21		BIC = 399.80		
Ljung-Box (p-value) = 0.6413 (with 12 lags)				
RARMAX(3,2)				
	Estimate	Standard error	Z stat	p-value
α	0.2855	0.1223	2.3335	0.0196
ϕ_1	0.6882	0.2513	2.7382	0.0062
ϕ_2	-0.6323	0.2876	2.1987	0.0279
ϕ_3	0.4046	0.1528	2.6486	0.0081
θ_1	-0.3071	0.2621	1.1715	0.2414
θ_2	0.5309	0.2587	2.0521	0.0402
β_1	0.3346	0.1080	3.0973	0.0020
AIC = 464.89		BIC = 491.44		
Ljung-Box (p-value) = 0.2061 (with 12 lags)				

Table 4: In-sample and out-of-sample prediction accuracy measures. Best results are highlighted in bold.

	In-Sample		Out-of-Sample	
Model	MSE	MAPE	MSE	MAPE
MxARMAX	0.2160	**0.1738**	**0.1294**	**0.1560**
RARMAX	0.2537	0.1815	0.1963	0.1964
ETS	0.2033	0.2183	0.2400	0.2065
HW	**0.2030**	0.1802	0.2326	0.2043

7. Conclusion

This paper introduced a new time series model for positive data, named MxARMA. The proposed model is based on the one-parameter Maxwell distribution, which offers a suitable balance between simplicity and flexibility. The dynamic specification of the model was presented, along with the corresponding parameter estimation theory, diagnostic tools, model selection criteria, and forecasting procedures.

In particular, closed-form matrix expressions for the conditional score vector and the conditional Fisher information matrix were derived. All methodological developments were implemented in a comprehensive `R` package, freely available on GitHub. Monte Carlo simulations and an empirical application evidenced the adequacy of the MxARMA model and its inferential framework. Notably, in the wind speed application for Brasília, Brazil, the proposed model outperformed the main competing models available in the literature.

Acknowledgments

This work was partially funded by the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Swedish-Brazilian Research and Innovation Centre (CISB), and Saab AB.

[Received: September 2025 — Accepted: November 2025]

References

Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**(6), 716–723.

Andersen, B. A. (1970), 'Asymptotic properties of conditional maximum-likelihood estimators', *Journal of the Royal Statistical Society. Series B* **32**(1), 283–301.

Ansley, C. F. & Newbold, P. (1980), 'Finite sample properties of estimators for autoregressive moving average models', *Journal of Econometrics* **13**(2), 159–183.

Bayer, F. M., Bayer, D. M., Marinoni, A. & Gamba, P. (2020), 'A novel Rayleigh dynamical model for remote sensing data interpretation', *IEEE Transactions on Geoscience and Remote Sensing* **58**(7), 4989–4999.

Bayer, F. M., Bayer, D. M. & Pumi, G. (2017), 'Kumaraswamy autoregressive moving average models for double bounded environmental data', *Journal of Hydrology* **555**, 385–396.

Bekker, A. J. J. J. & Roux, J. J. J. (2005), 'Reliability characteristics of the Maxwell distribution: a Bayes estimation study', *Communications in Statistics – Theory and Methods* **34**(11), 2169–2178.

Benjamin, M. A., Rigby, R. A. & Stasinopoulos, D. M. (2003), 'Generalized autoregressive moving average models', *Journal of the American Statistical Association* **98**(461), 214–223.

Boltzmann, L. (1868), 'Wiener Ber.', *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften in Wien* **58**, 517–560.

Box, G. E., Jenkins, G. M., Reinsel, G. C. & Ljung, G. M. (2015), *Time series analysis: forecasting and control*, John Wiley & Sons.

Brockwell, P. J. & Davis, R. A. (1991), *Time series: theory and methods*, Springer, New York.

Chaturvedi, A. & Rani, U. (1998), 'Classical and Bayesian reliability estimation of the generalized Maxwell failure distribution', *Journal of Statistical Research* **32**(1), 113–120.

de Araújo, F. J. M., Guerra, R. R. & Peña-Ramírez, F. A. (2024), 'Quantile-based dynamic modeling of asymmetric data: a novel Burr XII approach for positive continuous random variables', *International Journal of Data Science and Analytics* pp. 1–20.

Dey, S. & Maiti, S. S. (2010), 'Bayesian estimation of the parameter of Maxwell distribution under different loss functions', *Journal of Statistical Theory and Practice* **4**, 279–287.

Douak, F., Melgani, F. & Benoudjit, N. (2013), 'Kernel ridge regression with active learning for wind speed prediction', *Applied Energy* **103**, 328–340.

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

Energy Sector Management Assistance Program (2019), *Going global: expanding offshore wind to emerging markets*, World Bank.

Fahrmeir, L. & Kaufmann, H. (1985), 'Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models', *The Annals of Statistics* **13**(1), 342–368.

Hyndman, R. J. & Khandakar, Y. (2008), 'Automatic time series forecasting: the forecast package for R', *Journal of Statistical Software* **27**, 1–22.

Hyndman, R. J. & Koehler, A. B. (2006), 'Another look at measures of forecast accuracy', *International Journal of Forecasting* **22**(4), 679–688.

International Renewable Energy Agency (IRENA) (2024), 'Wind energy'. Accessed: October 2025. https://www.irena.org/Energy-Transition/Technology/Wind-energy

Kazmi, S. M. A., Aslam, M. & Ali, S. (2011), 'A note on the maximum likelihood estimators for the mixture of Maxwell distributions using type-I censored scheme', *The Open Statistics and Probability Journal* **3**(1).

Kazmi, S. M. A., Aslam, M. & Ali, S. (2012), 'Preference of prior for the class of lifetime distributions under different loss functions', *Pakistan Journal of Statistics* **28**(4).

Kendall, M. G. (1948), *The Advanced Theory of Statistics: Distribution Theory*, Vol. 1, Charles Griffin and Company, London.

Khosravi, A., Machado, L. & Nunes, R. O. (2018), 'Time-series prediction of wind speed using machine learning algorithms: a case study of Osorio wind farm, Brazil', *Applied Energy* **224**, 550–566.

Konishi, S. & Kitagawa, G. (2008), Bayesian information criteria, *in* 'Information Criteria and Statistical Modeling', Springer, New York, pp. 211–237.

Lemonte, A. J. (2024), 'A simple mean-parameterized Maxwell regression model for positive response variables', *REVSTAT – Statistical Journal* **22**(4), 503–526.

Ljung, G. M. & Box, G. E. P. (1978), 'On a measure of lack of fit in time series models', *Biometrika* **65**(2), 297–303.

Maxwell, J. C. (1867), 'On the dynamical theory of gases', *Philosophical Transactions of the Royal Society of London* **157**, 49–88.

Melo, M. & Alencar, A. (2020), 'The Conway–Maxwell–Poisson autoregressive moving average model for equidispersed, underdispersed, and overdispersed count data', *Journal of Time Series Analysis* **41**(6), 830–857.

Palm, B. G. & Bayer, F. M. (2018), 'Bootstrap-based inferential improvements in beta autoregressive moving average model', *Communications in Statistics – Simulation and Computation* **47**(4), 977–996.

Palm, B. G., Bayer, F. M., Cintra, R. J., Pettersson, M. I. & Machado, R. (2019), 'Rayleigh regression model for ground type detection in SAR imagery', *IEEE Geoscience and Remote Sensing Letters* **16**(10), 1660–1664.

R Core Team (2024), *R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. `https://www.R-project.org/`

Radha, R. K. & Venkatesan, P. (2013), 'On the double prior selection for the parameter of Maxwell distribution', *International Journal of Scientific & Engineering Research* **4**(5), 1238–1241.

Rocha, A. V. & Cribari-Neto, F. (2009), 'Beta autoregressive moving average models', *TEST* **18**(3), 529–545.

Stone, R. F., Loose, L. H., Melo, M. S. & Bayer, F. M. (2023), 'The Chen autoregressive moving average model for modeling asymmetric positive continuous time series', *Symmetry* **15**(9), 1675.

Tomer, S. K. & Panwar, M. S. (2015), 'Estimation procedures for Maxwell distribution under type-I progressive hybrid censoring scheme', *Journal of Statistical Computation and Simulation* **85**(2), 339–356.

Tyagi, R. K. & Bhattacharya, S. K. (1989), 'A note on the MVU estimation of reliability for the Maxwell failure distribution', *Estadistica* **41**(137), 73–79.

Vetterling, W. T. & Press, W. H. (1992), *Numerical recipes: example book C*, Cambridge University Press.

Winters, P. R. (1960), 'Forecasting sales by exponentially weighted moving averages', *Management Science* **6**(3), 324–342.

Appendix A. Conditional Information Matrix

In this appendix, we derive the conditional Fisher information matrix. To this end, we compute the expected values of the second-order partial derivatives of the conditional log-likelihood function. The general expression for the second derivative is given by:

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

where the quantities $\frac{d\mu_t}{d\eta_t}$ and $\frac{\partial\eta_t}{\partial\gamma_i}$ are already introduced in Section 3.1.

Once $\mathrm{E}\left(d\ell_t(\mu_t)/d\mu_t\middle|\mathcal{F}_{t-1}\right) = 0$, the expected value is given by:

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

where

$$\mathbb{E}\left(\frac{d^2\ell_t(\mu_t)}{d\mu_t^2}\middle|\mathcal{F}_{t-1}\right) = \frac{-24(0.178\mu_t^2+\mu_t^2)}{\mu_t^4\pi} + \frac{3}{\mu_t^2} = -\frac{6}{\mu_t^2}.$$

Considering the log link function, we have

$$\left(\frac{d\mu_t}{d\eta_t}\right)^2 = \mu_t^2,$$

and then

$$-\mathbb{E}\left(\frac{d^2\ell_t(\mu_t)}{d\mu_t^2}\middle|\mathcal{F}_{t-1}\right)\left(\frac{d\mu_t}{d\eta_t}\right)^2 = 6.$$

The conditional Fisher information matrix for the vector $\boldsymbol{\gamma}$ can be written as:

$$\boldsymbol{K} = \boldsymbol{K}(\boldsymbol{\gamma}) = 6\begin{bmatrix} K(\alpha,\alpha) & \boldsymbol{K}(\alpha,\boldsymbol{\beta}) & \boldsymbol{K}(\alpha,\boldsymbol{\phi}) & \boldsymbol{K}(\alpha,\boldsymbol{\theta}) \\ \boldsymbol{K}(\boldsymbol{\beta},\alpha) & \boldsymbol{K}(\boldsymbol{\beta},\boldsymbol{\beta}) & \boldsymbol{K}(\boldsymbol{\beta},\boldsymbol{\phi}) & \boldsymbol{K}(\boldsymbol{\beta},\boldsymbol{\theta}) \\ \boldsymbol{K}(\boldsymbol{\phi},\alpha) & \boldsymbol{K}(\boldsymbol{\phi},\boldsymbol{\beta}) & \boldsymbol{K}(\boldsymbol{\phi},\boldsymbol{\phi}) & \boldsymbol{K}(\boldsymbol{\phi},\boldsymbol{\theta}) \\ \boldsymbol{K}(\boldsymbol{\theta},\alpha) & \boldsymbol{K}(\boldsymbol{\theta},\boldsymbol{\beta}) & \boldsymbol{K}(\boldsymbol{\theta},\boldsymbol{\phi}) & \boldsymbol{K}(\boldsymbol{\theta},\boldsymbol{\theta}) \end{bmatrix}, \tag{A1}$$

where

$$
\begin{aligned}
&K(\alpha,\alpha) = \boldsymbol{v}^\top \boldsymbol{v}, \quad \boldsymbol{K}(\alpha,\boldsymbol{\beta}) = \boldsymbol{v}^\top \boldsymbol{M}, \quad \boldsymbol{K}(\alpha,\boldsymbol{\phi}) = \boldsymbol{v}^\top \boldsymbol{P}, \quad \boldsymbol{K}(\alpha,\boldsymbol{\theta}) = \boldsymbol{v}^\top \boldsymbol{R},\\
&\boldsymbol{K}(\boldsymbol{\beta},\alpha) = \boldsymbol{M}^\top \boldsymbol{v}, \quad \boldsymbol{K}(\boldsymbol{\beta},\boldsymbol{\beta}) = \boldsymbol{M}^\top \boldsymbol{M}, \quad \boldsymbol{K}(\boldsymbol{\beta},\boldsymbol{\phi}) = \boldsymbol{M}^\top \boldsymbol{P}, \quad \boldsymbol{K}(\boldsymbol{\beta},\boldsymbol{\theta}) = \boldsymbol{M}^\top \boldsymbol{R},\\
&\boldsymbol{K}(\boldsymbol{\phi},\alpha) = \boldsymbol{P}^\top \boldsymbol{v}, \quad \boldsymbol{K}(\boldsymbol{\phi},\boldsymbol{\beta}) = \boldsymbol{P}^\top \boldsymbol{M}, \quad \boldsymbol{K}(\boldsymbol{\phi},\boldsymbol{\phi}) = \boldsymbol{P}^\top \boldsymbol{P}, \quad \boldsymbol{K}(\boldsymbol{\phi},\boldsymbol{\theta}) = \boldsymbol{P}^\top \boldsymbol{R},\\
&\boldsymbol{K}(\boldsymbol{\theta},\alpha) = \boldsymbol{R}^\top \boldsymbol{v}, \quad \boldsymbol{K}(\boldsymbol{\theta},\boldsymbol{\beta}) = \boldsymbol{R}^\top \boldsymbol{M}, \quad \boldsymbol{K}(\boldsymbol{\theta},\boldsymbol{\phi}) = \boldsymbol{R}^\top \boldsymbol{P}, \quad \boldsymbol{K}(\boldsymbol{\theta},\boldsymbol{\theta}) = \boldsymbol{R}^\top \boldsymbol{R}.
\end{aligned}
$$

Appendix B. Simulation Results

In this appendix, we present additional Monte Carlo simulation results for alternative model configurations. Specifically, the outcomes for the MxARMA(1,1), MxARMAX(0,1), and MxARMAX(1,0) models are reported in Tables A1, A2, and A3, respectively. We focus on low-order ARMA structures for illustration, but extensions to higher orders are straightforward and show similar numerical results.

The same simulation design and evaluation criteria described in Section 5 are employed, including the computation of the RB, MSE, and CR for each coefficient. These additional scenarios provide complementary evidence on the finite-sample performance and accuracy of the proposed estimation procedure under different model specifications.

TABLE A1: Monte Carlo simulation results for the MxARMA(1,1) model.

Parameter	T	Mean	RB(%)	MSE	CR
$\alpha = 0.5$	50	0.7292	45.8309	0.2086	0.8088
	100	0.7281	45.6258	0.2122	0.7784
	300	0.5824	16.4762	0.0725	0.8704
	500	0.5646	12.9208	0.0544	0.8844
	1000	0.5220	4.3976	0.0176	0.9188
	2000	0.5122	2.4325	0.0081	0.9352
$\phi_1 = 0.45$	50	0.1755	−61.0094	0.2809	0.8054
	100	0.1832	−59.2852	0.2836	0.7770
	300	0.3544	−21.2447	0.0942	0.8716
	500	0.3758	−16.4884	0.0702	0.8834
	1000	0.4248	−5.6012	0.0225	0.9208
	2000	0.4359	−3.1250	0.0104	0.9352
$\theta_1 = -0.3$	50	−0.0312	−89.6067	0.2955	0.7826
	100	−0.0344	−88.5220	0.2779	0.7642
	300	−0.2058	−31.3946	0.0977	0.8636
	500	−0.2259	−24.6956	0.0712	0.8816
	1000	−0.2754	−8.1991	0.0249	0.9182
	2000	−0.2860	−4.6555	0.0116	0.9336

Table A2: Monte Carlo simulation results for the MxARMAX(0,1) model.

Parameter	T	Mean	RB(%)	MSE	CR
$\alpha = 0.5$	50	0.4893	−2.1393	0.0020	0.8856
	100	0.4952	−0.9603	0.0009	0.9294
	300	0.4982	−0.3659	0.0003	0.9426
	500	0.4992	−0.1560	0.0002	0.9462
	1000	0.4997	−0.0658	0.0001	0.9484
	2000	0.4997	−0.0577	0.0000	0.9508
$\beta = 0.6$	50	0.6007	0.1235	0.0042	0.9988
	100	0.6020	0.3252	0.0020	0.9996
	300	0.6001	0.0190	0.0007	0.9996
	500	0.6002	0.0279	0.0004	0.9996
	1000	0.5997	−0.0428	0.0002	0.9998
	2000	0.6001	0.0188	0.0001	0.9996
$\theta_1 = -0.3$	50	−0.3470	15.6689	0.0240	0.5958
	100	−0.3151	5.0255	0.0086	0.6508
	300	−0.3059	1.9741	0.0023	0.6960
	500	−0.3025	0.8181	0.0014	0.7016
	1000	−0.3019	0.6499	0.0007	0.7110
	2000	−0.3009	0.3064	0.0003	0.6986

Table A3: Monte Carlo simulation results for the MxARMAX(1,0) model.

Parameter	T	Mean	RB(%)	MSE	CR
$\alpha = 0.5$	50	0.5267	5.3442	0.0120	0.9516
	100	0.5091	1.8224	0.0051	0.9470
	300	0.5036	0.7193	0.0016	0.9518
	500	0.5018	0.3602	0.0009	0.9510
	1000	0.5015	0.2983	0.0005	0.9470
	2000	0.5006	0.1136	0.0002	0.9486
$\beta = 0.6$	50	0.5991	−0.1428	0.0168	0.9170
	100	0.6021	0.3570	0.0085	0.9016
	300	0.5997	−0.0463	0.0027	0.8984
	500	0.6003	0.0475	0.0016	0.8996
	1000	0.6007	0.1098	0.0008	0.9062
	2000	0.6001	0.0198	0.0004	0.9148
$\phi_1 = 0.45$	50	0.3982	−11.5065	0.0169	0.9016
	100	0.4298	−4.4791	0.0069	0.9430
	300	0.4423	−1.7112	0.0020	0.9614
	500	0.4463	−0.8300	0.0012	0.9702
	1000	0.4476	−0.5334	0.0006	0.9758
	2000	0.4488	−0.2650	0.0003	0.9762

Appendix C. MxARMA Package

We developed an R package, named `MxARMA`, to facilitate the use of the proposed model by practitioners and researchers, which is available on GitHub at `https://github.com/arthurhintz/MxARMA`. The package implements the main

functions of the mean-parameterized Maxwell distribution, including the probability density, cumulative distribution, and quantile functions, as well as random number generation. Additionally, it provides routines for simulating data from the MxARMA model, with or without regressors, and for fitting the model to empirical data. The package also includes the dataset used in the empirical application presented in this paper. The `MxARMA` package only depends on the `stats` package, which is included in base R (R Core Team, 2024). A summary of the available functions is given in Table A4.

TABLE A4: Summary of main functions available in the `MxARMA` package.

Basic Functions	
`dmax()`	Probability density function of the parametrized Maxwell distribution
`pmax()`	Cumulative distribution function of the parametrized Maxwell distribution
`qmax()`	Quantile function of the parametrized Maxwell distribution
`rmax()`	Random generation from the parametrized Maxwell distribution
Simulation	
`mxarma.sim()`	Simulation from the MxARMA model with and without regressors
Model Fitting	
`mxarma.fit()`	Estimation procedure for the MxARMA model with and without regressors
Dataset	
`wind_speed_brasilia`	Time series of average wind speed recorded in Brasília, Brazil

The functions in the `MxARMA` package are designed to support model fitting for various orders, both with and without covariates. The structures of the available model specifications are summarized in Table A5.

TABLE A5: MxARMA model structures with and without covariates.

Dynamic	Without covariates	With covariates
AR(p)	$\eta_t = \alpha + \sum_{i=1}^{p} \phi_i g(y_{t-i})$	$\eta_t = \alpha + \sum_{i=1}^{p} \phi_i \left(g(y_{t-i}) - \boldsymbol{x}_{t-i}^{\top}\boldsymbol{\beta} \right)$
MA(q)	$\eta_t = \alpha + \sum_{j=1}^{q} \theta_j r_{t-j}$	$\eta_t = \alpha + \sum_{j=1}^{q} \theta_j r_{t-j} + \boldsymbol{x}_{t-i}^{\top}\boldsymbol{\beta}$
ARMA(p, q)	$\eta_t = \alpha + \sum_{i=1}^{p} \phi_i g(y_{t-i}) + \sum_{j=1}^{q} \theta_j r_{t-j}$	$\eta_t = \alpha + \sum_{i=1}^{p} \phi_i \left(g(y_{t-i}) - \boldsymbol{x}_{t-i}^{\top}\boldsymbol{\beta} \right) + \sum_{j=1}^{q} \theta_j r_{t-j} + \boldsymbol{x}_{t-i}^{\top}\boldsymbol{\beta}$

The general simulation function, `mxarma.sim()`, has the following arguments:

```
R> mxarma.sim(n, phi = NULL, theta = NULL, alpha = 0.0,
+                 beta = NULL, xreg = NULL)
```

This function generates a sample $\{y_t\}_{t=1}^{T}$ from an MxARMA process according to the dynamic structure defined in Equation (3) and the stochastic component in Equation (1). The argument `n` specifies the sample size T, `alpha` represents the intercept α, while `phi` and `theta` correspond to the vectors of autoregressive

$\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^\top$ and moving average coefficients $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^\top$, respectively. The argument `beta` defines the vector of covariate parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^\top$, and `xreg` specifies the matrix of covariates. Setting `phi = NULL` and `theta = NULL` implies $p = 0$ and $q = 0$, respectively, meaning that the model includes no autoregressive or moving average components. Likewise, when both `beta = NULL` and `xreg = NULL`, the model is simulated without covariates.

The function `mxarma.fit()` is used to fit the MxARMA, with the following syntax:

```
R> fit <- mxarma.fit(y, xreg, ar = NA, ma = NA, resid = 1,
+                    h1 = 0, X_hat = 0)
```

Here, `y` denotes the response vector representing the observed time series, and `xreg` is a matrix of covariates with dimensions $T \times k$, where k is the number of covariates. The arguments `ar` and `ma` specify the autoregressive (p) and moving average (q) orders, respectively. The argument `h1` defines the number of out-of-sample forecasts to be generated, while `X_hat` corresponds to the matrix of covariates for prediction, with dimensions $(h_1 \times k)$. Finally, the argument `resid` specifies the type of residuals to be computed, where the default `resid = 1` corresponds to the quantile residuals, `resid = 2` to the standard residuals, and `resid = 3` to the deviance residuals.

When the function `mxarma.fit()` is executed, it returns the following outputs:

```
names(fit)
[1] "conv" "errorhat" "K" "fitted" "forecast"
[6] "residc" "loglik" "aic" "bic" "model"
```

The object `conv` indicates whether the parameter estimates converged during the optimization procedure. The component `errorhat` corresponds to the error term r_t defined in Equation (3), while `K` denotes the conditional Fisher information matrix, defined in Equation (A1), evaluated at the CMLE of parameters. The elements `fitted` and `forecast` provide, respectively, the in-sample fitted values and the out-of-sample forecasts. The component `residc` contains the model residuals. The `loglik` component stores the maximized conditional log-likelihood value of the fitted model, as defined in Equation (4), whereas `aic` and `bic` correspond to the Akaike and Bayesian information criteria, respectively. Finally, the `model` component returns a summary including the parameter estimates, standard errors, Z-statistics, and p-values for each coefficient.

Unlike the classical `ARIMA()` function from the `forecast` package (Hyndman & Khandakar, 2008), where the initial specification of autoregressive and moving average orders is fixed and all parameters are estimated regardless of significance, the `MxARMA` package allows the removal of non-significant parameters after fitting. The following code illustrates the usage of these functions.

```
R> devtools::install_github("arthurhintz/MxARMA")
R> library(MxARMA)
```

```
R> library(forecast)
R> library(lmtest)
R> set.seed(1248)

R> y <- arima.sim(n = 1000, list(ar = c(0.04, 0.2, 0.5),
+                                 ma = c(0.002, 0.3)))
R> fit <- Arima(y, order = c(3,0,2))
R> coeftest(fit)

##              Estimate Std. Error z value  Pr(>|z|)
## ar1        -0.042582   0.055783 -0.7634  0.44525
## ar2         0.259971   0.041637  6.2438  4.271e-10 ***
## ar3         0.579865   0.033312 17.4070  < 2.2e-16 ***
## ma1         0.088212   0.063998  1.3783  0.16810
## ma2         0.209929   0.050037  4.1954  2.723e-05 ***
## intercept   0.358098   0.193833  1.8475  0.06468 .

R> y <- mxarma.sim(n = 1000, alpha = 0.08, phi = c(0.04, 0.2, 0.5),
+                theta = c(0.002, 0.3))
R> fit <- mxarma.fit(y, ar = c(2,3), ma = c(2))
R> fit\$model

##          Estimate Std. Error z value Pr(>|z|)
## alpha     0.0961     0.0168   5.7233   0.0000 ***
## phi2      0.2605     0.0317   8.2116   0.0000 ***
## phi3      0.5085     0.0239  21.2877   0.0000 ***
## theta2    0.2788     0.0371   7.5084   0.0000 ***
```