UFSM

CIÊNCIA e NATURA

ISSN 2179-460X
OPEN ACCESS

Ci. e Nat., Santa Maria, v. 48, e91810, 2026 • https://doi.org/10.5902/2179460X91810
Submitted: 30/04/2025 • Approved: 18/11/2025 • Published: 02/07/2026

Statistics

Bayesian modeling of Tuberculosis notifications using the generalized Poisson distribution

Modelagem Bayesiana das notificações de Tuberculose utilizando a distribuição Poisson generalizada

Luiz Otávio de Oliveira Pala [I], **Ryan Rodrigo Oliveira de Paula** [I], **Luciano José Pereira** [I], **Thelma Sáfadi** [I]

[I]Universidade Federal de Lavras, Lavras, MG, Brasil.

ABSTRACT

Time series models are widely applied across various scientific areas, enabling forecasting and trend identification. Traditional approaches, such as those based on the Autoregressive Moving Average class, have been expanded in the literature, with the generalized Autoregressive Moving Average (GARMA) models being an example of this expansion, allowing the analysis of discrete, rate, or proportion time series. However, when dealing with count time series, applied studies commonly assume normality for the response or adopt distributions such as the Poisson or negative Binomial, which, in some cases, may not accommodate features such as overdispersion. In this context, this study proposes the use of the generalized Poisson and zero-adjusted generalized Poisson distributions as alternatives to these models. The models were defined using a temporal dependence structure similar to that of the GARMA model, and inference was performed through a Bayesian approach. The models were evaluated through a simulation study, and functions were developed in `R`, via the `shiny` interface, for sample generation from the zero-adjusted generalized Poisson, ensuring the reproducibility of the study. Finally, we modeled tuberculosis notifications in Minas Gerais, Brazil, providing forecasts for public health use.

Keywords: Bayesian inference; Hamiltonian Monte Carlo; Time series analysis; Tuberculosis

RESUMO

Modelos de séries temporais são amplamente utilizados em diversas áreas da ciência, permitindo a realização de previsões e a identificação de tendências. Abordagens tradicionais, como aquelas baseadas na classe de modelos Autorregressivos e de Médias Móveis, têm sido expandidas na literatura, sendo os modelos Autorregressivos e de Médias Móveis Generalizados (GARMA) um exemplo dessa expansão, permitindo a análise de séries temporais discretas, de taxas ou de proporções. Entretanto, ao lidar com séries de contagem, estudos aplicados comumente assumem normalidade para a variável

resposta ou adotam distribuições como a Poisson ou a Binomial negativa, que, em alguns casos, podem não acomodar características como a alta dispersão. Nesse contexto, este estudo propõe a utilização das distribuições Poisson generalizada e Poisson generalizada zero-ajustada como alternativas à essas distribuições. Para a definição dos modelos, foi empregada uma estrutura de dependência temporal semelhante à do modelo GARMA, com a inferência realizada por meio da abordagem Bayesiana. Os modelos foram avaliados por meio de um estudo de simulação e também foram desenvolvidas funções em `R`, por meio da interface `shiny`, para a geração de amostras da distribuição Poisson generalizada zero-ajustada, possibilitando a reprodutibilidade do estudo. Por fim, modelamos as notificações de tuberculose em Minas Gerais, Brasil, fornecendo previsões para uso em saúde pública.

Palavras-chave: Inferência Bayesiana; Hamiltoniano Monte Carlo; Análise de séries temporais; Tuberculose

1 INTRODUCTION

Recent advances in modeling count time series have been extensively discussed in the statistical literature. An important methodological review on the topic is presented by Davis et al. (2021), which addresses models based on the binomial thinning operator, such as the integer-valued autoregressive model, as well as models with observation-driven or parameter-driven structures. The characteristics of these structures have also been explored by Franco et al. (2019).

Regarding observation-driven, the generalized Autoregressive Moving Average models, also known as the GARMA family, were proposed by Benjamin et al. (2003) and can be seen as an extension of the generalized linear models, incorporating a temporal dependence structure in the form of the Autoregressive Moving Average (ARMA) process in the linear predictor. According to Franco et al. (2019), they arise from a general formulation of the studies by Zeger and Qaqish (1988) and Li (1994).

In addition to GARMA, the Generalized Linear Autoregressive Moving Average family was proposed by Davis et al. (2003). This approach is considered flexible and easy to fit when dealing with counts. Besides, it is related to the generalized AR score and dynamic conditional score models (Davis et al., 2021). Regarding parameter-driven structures, in which the state equation follows a stochastic mechanism, Franco et al. (2019) adopted an approach that assumes a latent autoregressive process in the mean.

The literature on modeling count time series using observation-driven models offers various alternatives. In `R` (R Core Team, 2024), for example, the packages `garma` (Hunt, 2022), `glarma` (Dunsmuir & Scott, 2015), and `gamlss.util` (Stasinopoulos et al., 2016) allow fitting models via classical approaches for distributions such as Poisson

and negative Binomial. Generally, observation-driven structures can be attractive when compared to parameter-driven ones, especially in terms of parameter estimation and the complexity of the likelihood function (Davis et al., 2021).

Concerning the different observation-driven perspectives, several extensions have been proposed, including the use of distributions such as the Poisson inverse Gaussian, Bernoulli Geometric, and Conway–Maxwell–Poisson (Melo & Alencar, 2020; Sales et al., 2022); the adoption of zero-adjusted and zero-inflated distributions (Pala & Sáfadi, 2024; Sathish et al., 2021); the expansion to modified model structures (Albarracin et al., 2019); and the incorporation of seasonal components in the form of SARIMA process (Briët et al., 2013).

In high dispersion scenarios, which can result from an excess of zeros, the presence of two or more mixtures (Yadav et al., 2021), or outliers (Payne et al., 2015), models with a negative Binomial response are often adopted. In cases of zero inflation or deflation, the negative Binomial may still not perform well, making it necessary to consider other distributions that accommodate these characteristics (Melo & Alencar, 2020).

Given the diversity of distributions for count data, the generalized Poisson distribution is an interesting option in situations of high dispersion triggered by an excess of zeros and includes the Poisson distribution as a special case (Yadav et al., 2024). This distribution can be easily extended to a hurdle specification to simultaneously address zero inflation/deflation and differing mean-variance relationships (Zuo et al., 2021).

Despite the availability of numerous distributions in the statistical literature for modeling count time series, many applied studies still assume normality for the response variable or adopt distributions such as Poisson or negative Binomial, which, in some cases, may not fully accommodate features like high dispersion and excess zeros, characteristics that are commonly observed in epidemiological series. This underscores the practical relevance of flexible alternatives, such as the generalized Poisson.

Our goal in this paper is to adopt the generalized Poisson distribution and the zero-adjusted version for modeling count series in a Bayesian perspective. Although this distribution has been used to model count data, as in the Generalized Additive Models for Location, Scale and Shape (GAMLSS), we conducted a simulation to evaluate its performance when dealing with time series. Our approach incorporates the predictor structure proposed by Benjamin et al. (2003) and also includes seasonal components, as detailed in Briët et al. (2013) and Pala et al. (2022). For the zero-adjusted generalized Poisson case, we evaluate models that allow for modeling both the mean and the parameter that estimates the exact probability of zero occurrences, similar to the approach used by Pala and Sáfadi (2024) when considering the zero-adjusted Poisson. We also provide functions in `R`, via the `shiny` interface, that allow for generating samples and calculating quantiles of the zero-adjusted generalized Poisson distribution.

2 MODEL DEFINITION

Suppose $\boldsymbol{Y} = \{y_1, y_2, \ldots, y_n\}$ represents a time series and denote $\boldsymbol{X}_t = (\boldsymbol{x}_{1,t}, \ldots, \boldsymbol{x}_{r,t})^\top$, for $t = \{1, \ldots, n\}$, as a vector containing r explanatory variables, which can be viewed as a design matrix, where $r < n$. These variables are associated with $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_r)^\top$, being a set of unknown parameters.

Adhering to the conventional notation of time series models, we define the autoregressive and moving average polynomials of orders p and q, respectively, as $\boldsymbol{\Phi}_p(B) = (1 - \phi_1 B^1 - \ldots - \phi_p B^p)$ and $\boldsymbol{\Theta}_q(B) = (1 - \theta_1 B^1 - \ldots - \theta_q B^q)$. We also consider $\boldsymbol{\Phi}_P(B^s) = (1 - \Phi_1 B^{s1} - \ldots - \Phi_P B^{sP})$ and $\boldsymbol{\Theta}_Q(B^s) = (1 - \Theta_1 B^{s1} - \ldots - \Theta_Q B^{sQ})$ as the seasonal autoregressive and seasonal moving average polynomials of orders P and Q, respectively, where B is the backshift operator and s is the length of the period.

In order to define the model structure, $\mathcal{F}_{t-1}$ denotes the set of previous information up to time $t-1$. The parameter sets $\boldsymbol{\Phi} = (\phi_1, \ldots, \phi_p; \Phi_1, \ldots, \Phi_P)^\top$ and $\boldsymbol{\Theta} = (\theta_1, \ldots, \theta_q; \Theta_1, \ldots, \Theta_Q)^\top$ are analogous to those in the SARMA(p, q)(P, Q)$_s$ model, representing the seasonal autoregressive moving average process. Based on these definitions, in sections 2.1 and 2.2, the approaches considering the generalized Poisson and zero-adjusted generalized Poisson are presented, respectively.

2.1 Generalized Poisson

Assuming that $y_t \mid \mathcal{F}_{t-1}$ follows a generalized Poisson, GPO(μ_t, σ), distribution. The probability mass function of each y_t can be written as shown in equation (1):

$$P(y_t \mid \mu_t, \sigma) = \left(\frac{\mu_t}{1+\sigma\mu_t}\right)^{y_t} \frac{(1+\sigma y_t)^{y_t-1}}{y_t!} \times \\ \times \quad \exp\left[\frac{-\mu_t(1+\sigma y_t)}{1+\sigma\mu_t}\right] \tag{1}$$

where $\mu_t > 0$ and $\sigma > 0$ denote the mean and dispersion parameters. It is noteworthy that we can define the model to accommodate temporal variations in both parameters.

Given the domain of the parameter space satisfying $\mathbf{\Omega} = \{\mu_t, \sigma \mid \mu_t > 0, \sigma > 0\}$, we consider the observation-driven structure displayed in equation (2):

$$\begin{aligned} \log(\mu_t) &= \mathbf{\Phi}_p(B)\mathbf{\Phi}_P(B^s)\left[\boldsymbol{X}_t^\top\boldsymbol{\beta} - \log(y_t)\right] \\ &\quad - \mathbf{\Theta}_q(B)\mathbf{\Theta}_Q(B^s)\left[\log(y_t) - \log(\mu_t)\right] \\ &\quad + \log\left(\frac{y_t^2}{\mu_t}\right) \\ \sigma &= \sigma_0. \end{aligned} \tag{2}$$

As we can see in Equation (2), μ_t follows a SARMAX(p, q)(P, Q)$_s$ process, which can be expanded to SARIMAX through the inclusion of the integration operator or simplified to a basic AR(p). In some cases, it may be necessary to include a threshold, c, to ensure the existence of the function $\log(\cdot)$; in our case, we considered $y_t^* = \max(y_t, c)$, with $c = 0.10$.

The GPO(μ_t, σ) distribution, as outlined in equation (1), can be extended to its zero-adjusted version, referred to as ZAGPO(μ_t, σ, ν), where ν represents the probability of the series being equal to zero. For comprehensive details regarding the construction of the ZAGPO(μ_t, σ, ν), see the section 6.

2.2 Zero-adjusted generalized Poisson

Considering that $y_t \mid \mathcal{F}_{t-1}$ follows a ZAGPO(μ_t, σ, ν) distribution, we have:

$$P(y_t \mid \mu_t, \sigma, \nu) = \nu + (1-\nu)\times \frac{\left(\frac{\mu_t}{1+\sigma\mu_t}\right)^{y_t} \frac{(1+\sigma y_t)^{y_t-1}}{y_t!} \exp\left[\frac{-\mu_t(1+\sigma y_t)}{1+\sigma\mu_t}\right]}{1-\exp\left(\frac{-\mu_t}{1+\sigma\mu_t}\right)},$$

where $0 < \nu < 1$. The μ_t and σ structures are similar to those described in the previous subsection.

When establishing the GPO(μ_t, σ) model in subsection (2.1), an analogous framework can be employed for both μ_t and σ, and we can also incorporate a $logit$ link function for ν when using the ZAGPO(μ_t, σ, ν) distribution.

An interesting approach for modeling the ν parameter is to introduce explanatory variables or lags of the response variable, allowing it to vary over time. For example, one might consider specifying $logit(\nu_t) = \omega_0 + \omega_1 y_{t-1}$. For further details, see Bertoli et al. (2021) and Pala and Sáfadi (2024).

To create and evaluate the ZAGPO(μ_t, σ, ν) distribution in `R`, we adapted functions from the package `gamlss.dist` (Stasinopoulos & Rigby, 2022) to generate pseudo-random samples (`rZAGPO`) and calculate quantiles (`qZAGPO`) of the distribution. These functions and documentation are available in a `shiny` interface at *https://luizpala1.github.io/website/apps/zagpo.html*.

3 BAYESIAN ANALYSIS

For inference, we adopt a full Bayesian approach for the models described in subsections (2.1) and (2.2). Following Bayes' Theorem, our joint posterior distribution can be written as $P(\boldsymbol{\theta} \mid \cdot) \propto L(\boldsymbol{\theta} \mid \boldsymbol{Y})P(\boldsymbol{\theta})$, where $P(\boldsymbol{\theta})$ is the joint distribution of the parameter vector $\boldsymbol{\theta}$, which is composed of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\Theta}, \sigma_0)^\top$ in the GPO(μ_t, σ) model and $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\Theta}, \sigma_0, \nu)^\top$ in the ZAGPO(μ_t, σ, ν) model. In this analysis, non-informative and independent priors were assumed for all components of $\boldsymbol{\theta}$.

To illustrate the prior structure, we considered a multivariate Gaussian distribution for $\boldsymbol{\beta}$; hence:

$$p(\boldsymbol{\beta} \mid \boldsymbol{\mu_0}, \tau_0^2 \boldsymbol{I_0}) \propto \exp\left[\frac{-1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu_0})^\top \tau_0^2 \boldsymbol{I_0} (\boldsymbol{\beta} - \boldsymbol{\mu_0})\right].$$

Here, $\boldsymbol{\mu_0}$ is a vector of length r, $\boldsymbol{I_0}$ is a $r \times r$ identity matrix, and τ_0^2 is the variance. To ensure proper and less informative priors, we fixed the hyperparameters as $\boldsymbol{\mu_0} = \mathbf{0}$ and τ_0^2 equal to 100. The same design used for $\boldsymbol{\beta}$ was applied to $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$; that is, normal priors with mean zero and high variability.

For the dispersion parameter (σ_0), a Gamma distribution with hyperparameters ($\alpha = 1$, $\gamma = 0.01$) was considered. Its probability density function is given by:

$$p(\sigma_0 \mid \alpha, \gamma) = \frac{\sigma_0^{\alpha-1}\gamma^\alpha}{\Gamma(\alpha)} \exp\left(-\gamma\sigma_0\right), \quad \sigma_0 \in (0, \infty).$$

Our strategy for modeling ν in the ZAGPO(μ_t, σ, ν) model was to consider logit$(\nu) = \omega_0$, where ω_0 is normally distributed with a mean of zero and a variance of one hundred. Thus, we assume that the probability of zero occurrences remains constant over time.

Since we have no closed forms for $P(\boldsymbol{\theta} \mid \cdot)$, our Bayesian approach relies on MCMC for inference. Therefore, our strategy was to sample from the joint posterior distribution using the No-U-Turn Sampler (NUTS) algorithm, proposed by Hoffman and Gelman (2014), which extends the Hamiltonian Monte Carlo (HMC) algorithm (Duane et al., 1987). These methods are available in the `RStan` package (Stan Development Team, 2022).

The HMC algorithm uses Hamiltonian dynamics for sampling, effectively addressing the local random walk behavior seen in other algorithms (Gelman et al., 2014), and using gradient information to better describe the posterior distribution (Conceição et al., 2021). This approach produces efficient samples and performs well in high-dimensional settings, with low correlation between samples (Burda & Maheu, 2013; McElreath, 2020). For details on the HMC procedure, we suggest Conceição et al. (2021) or Gelman et al. (2014).

4 SIMULATION ANALYSIS AND FINDINGS

In this section, we present a simulation study of the models outlined in section 2. Subsection 4.1 details the simulation methodology, while subsection 4.2 reports the results obtained.

4.1 Simulation methodology and procedures

We established scenarios for the GPO(μ_t, σ) and ZAGPO(μ_t, σ, ν) models, with sample sizes n = $\{150, 300, 600, 1200\}$ and some autoregressive parameter configurations. The dispersion parameter (σ_0) was fixed at 0.500 in all cases. For the ZAGPO(μ_t, σ, ν) model, the parameter ω_0 was evaluated at {1.500, -0.900, -0.300, 0.300} to assess the precision of estimates in scenarios with a high proportion of zeros.

Each model was replicated $w = 1,000$ times. To simulate each time series, we employed the inverse transform method available in the `gamlss.dist` package (Stasinopoulos & Rigby, 2022) for the GPO(μ_t, σ) model and the `rZAGPO` function for the ZAGPO(μ_t, σ, ν) model.

For the MCMC procedure, the configurations for warmup, thinning, and total samples were established based on a prior pilot study. Accordingly, we set the warmup to 8,000 iterations, thinning to 2, and collected a total of 8,000 samples across two Markov chains. During the simulation process, we stored statistics such as the posterior mean, standard deviation, and mode.

In each of the w iterations, we evaluated the convergence to the stationary distribution using the $\widehat{R}$-split statistic, proposed by Gelman et al. (2014), and assumed convergence if $\widehat{R}$-split was less than 1.050. Additionally, we verified the bulk-ESS and tail-ESS estimates in accordance with the recommendations of the `RStan` package. If convergence was not achieved in any iteration, the simulation was discarded and restarted. Traceplots and other graphical diagnostics are not presented due to the large number of simulations, but readers may request the simulation results from the authors.

The simulation was conducted on a device equipped with an 11th Gen Intel(R) Core(TM) i7-11390H @ 3.40GHz processor (2.92 GHz) and 16.0 GB of RAM (15.7 GB usable), running a 64-bit operating system on an x64-based processor, using R version 4.2.3 (2023-03-15 ucrt) – “Shortstop Beagle”.

4.2 Simulation results

The results are presented in Tables 1, 2, 3, and 4, considering sample sizes ranging from 150 to 1200. Tables 1 and 2 display processes based on the GPO(μ_t, σ) model with AR(2) and SAR(2) orders in the μ_t equation. Table 3 and 4 present processes generated from the ZAGPO(μ_t, σ, ν) distribution, also with AR(2) and SAR(2) orders in the μ_t equation.

In the GPO(μ_t, σ) model, which results are shown in Tables 1 and 2, the estimates for the parameters β_1, ϕ_1, ϕ_2, and σ_0 tend to approach the true values as the sample size increases. Additionally, we observed a reduction in the standard deviations as the number of observations increases. For instance, when estimating β_1, the average standard deviation decreased from 0.286 (n = 150) to 0.075 (n = 1200), a reduction of approximately 73,70%.

We also found that the posterior mode estimates provided better results compared to the posterior mean estimates when estimating the dispersion parameter (σ_0). Across the scenarios presented, the mode estimates were generally closer to the true values, suggesting that they can be a useful option for representing a point estimate of the parameter, particularly in scenarios with fewer observations.

In the analysis of scenario IV with n = 150 and ϕ_1 = -0.700, which indicates a stronger first-order correlation, we observed that the posterior mean led to an underestimation of the β_1 parameter. However, this issue was corrected with the increase in the sample size, as observed in the other scenarios.

Table 1 – Simulation results for the GPO(μ_t, σ) model.

Scenario	n	θ	Real	Mean	Mode	SD
I	150	β_1	1.000	1.050	0.964	0.286
		ϕ_1	0.500	0.518	0.508	0.093
		ϕ_2	-0.300	-0.316	-0.310	0.088
		σ_0	0.500	0.519	0.503	0.081
	300	β_1	1.000	1.015	0.981	0.162
		ϕ_1	0.500	0.503	0.500	0.061
		ϕ_2	-0.300	-0.304	-0.302	0.059
		σ_0	0.500	0.507	0.500	0.055
	600	β_1	1.000	1.003	0.987	0.109
		ϕ_1	0.500	0.503	0.502	0.042
		ϕ_2	-0.300	-0.300	-0.299	0.041
		σ_0	0.500	0.504	0.500	0.038
	1200	β_1	1.000	0.999	0.992	0.075
		ϕ_1	0.500	0.500	0.499	0.029
		ϕ_2	-0.300	-0.299	-0.299	0.028
		σ_0	0.500	0.501	0.499	0.027
II	150	β_1	1.000	1.071	0.941	0.389
		ϕ_1	0.700	0.723	0.707	0.110
		ϕ_2	-0.400	-0.414	-0.404	0.101
		σ_0	0.500	0.516	0.499	0.083
	300	β_1	1.000	1.032	0.944	0.300
		ϕ_1	0.700	0.712	0.701	0.091
		ϕ_2	-0.400	-0.406	-0.399	0.084
		σ_0	0.500	0.511	0.498	0.070
	600	β_1	1.000	0.995	0.974	0.133
		ϕ_1	0.700	0.696	0.694	0.049
		ϕ_2	-0.400	-0.396	-0.395	0.046
		σ_0	0.500	0.502	0.498	0.039
	1200	β_1	1.000	1.000	0.990	0.092
		ϕ_1	0.700	0.703	0.701	0.034
		ϕ_2	-0.400	-0.398	-0.397	0.032
		σ_0	0.500	0.498	0.496	0.027

Table 2 – Simulation results for the GPO(μ_t, σ) model (Scenarios III and IV).

Scenario	n	θ	Real	Mean	Mode	SD
III	150	β_1	1.000	1.084	0.948	0.369
		ϕ_1	-0.500	-0.506	-0.498	0.080
		Φ_1	0.400	0.403	0.396	0.092
		σ_0	0.500	0.519	0.503	0.093
	300	β_1	1.000	1.038	0.989	0.183
		ϕ_1	-0.500	0.504	-0.499	0.059
		Φ_1	0.400	0.407	0.404	0.060
		σ_0	0.500	0.504	0.497	0.052
	600	β_1	1.000	1.019	0.998	0.117
		ϕ_1	-0.500	-0.502	-0.501	0.041
		Φ_1	0.400	0.403	0.402	0.040
		σ_0	0.500	0.504	0.500	0.036
	1200	β_1	1.000	1.006	0.996	0.079
		ϕ_1	-0.500	-0.502	-0.502	0.028
		Φ_1	0.400	0.401	0.401	0.028
		σ_0	0.500	0.502	0.500	0.025
IV	150	β_1	1.000	0.767	0.742	0.240
		ϕ_1	-0.700	-0.658	-0.640	0.107
		Φ_1	0.200	0.192	0.190	0.090
		σ_0	0.500	0.505	0.492	0.069
	300	β_1	1.000	1.014	0.992	0.110
		ϕ_1	-0.700	-0.711	-0.703	0.068
		Φ_1	0.200	0.202	0.202	0.060
		σ_0	0.500	0.508	0.502	0.046
	600	β_1	1.000	1.007	0.997	0.072
		ϕ_1	-0.700	-0.704	-0.701	0.045
		Φ_1	0.200	0.201	0.201	0.041
		σ_0	0.500	0.503	0.500	0.032
	1200	β_1	1.000	1.003	0.998	0.050
		ϕ_1	-0.700	-0.703	-0.702	0.031
		Φ_1	0.200	0.200	0.200	0.028
		σ_0	0.500	0.502	0.500	0.022

When analyzing the simulation results of the ZAGPO(μ_t, σ, ν) model, presented in Tables 3 and 4, we observed a reduction in the standard deviations as the sample size increased. We also found that the average standard deviations of the parameters related to μ_t were higher compared to the GPO(μ_t, σ) model, which can be explained by the presence of an additional parameter.

In the first scenario, the estimate for β_1 improves from 1.011 with a standard deviation of 0.392 when $n = 150$ to 0.991, with a lower standard deviation, when $n = 600$. Similarly, the estimate for the dispersion parameter σ_0 is 0.622 when $n = 150$, but as the sample size increases to 1200, the estimate converges to 0.515, much closer to the true value of 0.500.

Table 3 – Simulation results for the ZAGPO(μ_t, σ, ν) model (Scenarios I and II).

Scenario	n	θ	Real	Mean	Mode	SD
I	150	β_1	1.000	1.011	0.968	0.392
		ϕ_1	0.500	0.537	0.525	0.115
		ϕ_2	-0.300	-0.304	-0.293	0.106
		σ_0	0.500	0.622	0.523	0.218
		ω_0	-1.500	-1.540	-1.527	0.218
	300	β_1	1.000	0.973	0.977	0.186
		ϕ_1	0.500	0.515	0.510	0.074
		ϕ_2	-0.300	-0.306	-0.301	0.069
		σ_0	0.500	0.558	0.511	0.129
		ω_0	-1.500	-1.512	-1.506	0.151
	600	β_1	1.000	0.988	0.990	0.122
		ϕ_1	0.500	0.509	0.507	0.050
		ϕ_2	-0.300	-0.303	-0.301	0.047
		σ_0	0.500	0.526	0.504	0.079
		ω_0	-1.500	-1.508	-1.505	0.106
	1200	β_1	1.000	0.991	0.992	0.084
		ϕ_1	0.500	0.504	0.503	0.035
		ϕ_2	-0.300	-0.299	-0.298	0.032
		σ_0	0.500	0.515	0.505	0.053
		ω_0	-1.500	-1.504	-1.502	0.075
II	150	β_1	1.000	1.171	0.949	0.775
		ϕ_1	0.700	0.765	0.742	0.137
		ϕ_2	-0.400	-0.412	-0.399	0.120

Continued

Table 3 – (Continued)

Scenario	n	θ	Real	Mean	Mode	SD
		σ_0	0.500	0.643	0.530	0.243
		ω_0	-0.900	-0.912	-0.905	0.183
	300	β_1	1.000	0.994	0.966	0.284
		ϕ_1	0.700	0.727	0.719	0.085
		ϕ_2	-0.400	-0.407	-0.402	0.075
		σ_0	0.500	0.574	0.519	0.142
		ω_0	-0.900	-0.906	-0.904	0.128
	600	β_1	1.000	0.991	0.983	0.171
		ϕ_1	0.700	0.713	0.709	0.057
		ϕ_2	-0.400	-0.404	-0.402	0.050
		σ_0	0.500	0.531	0.506	0.085
		ω_0	-0.900	-0.901	-0.899	0.090
	1200	β_1	1.000	0.998	0.995	0.115
		ϕ_1	0.700	0.705	0.703	0.039
		ϕ_2	-0.400	-0.400	-0.399	0.035
		σ_0	0.500	0.513	0.502	0.055
		ω_0	-0.900	-0.899	-0.899	0.064

Table 4 – Simulation results for the ZAGPO(μ_t, σ, ν) model (Scenarios III and IV).

Scenario	n	θ	Real	Mean	Mode	SD
III	150	β_1	1.000	1.334	0.972	1.012
		ϕ_1	-0.500	-0.527	-0.506	0.133
		Φ_1	0.400	0.455	0.440	0.122
		σ_0	0.500	0.714	0.545	0.337
		ω_0	-0.300	-0.301	-0.299	0.175
	300	β_1	1.000	1.029	0.974	0.307
		ϕ_1	-0.500	-0.509	-0.502	0.076
		Φ_1	0.400	0.415	0.410	0.072
		σ_0	0.500	0.592	0.519	0.172
		ω_0	-0.300	-0.307	-0.306	0.120
	600	β_1	1.000	0.993	0.974	0.180
		ϕ_1	-0.500	-0.506	-0.503	0.051
		ϕ_2	0.400	0.405	0.403	0.048
		σ_0	0.500	0.548	0.515	0.100
		ω_0	-0.300	-0.302	-0.302	0.084
	1200	β_1	1.000	1.005	0.996	0.120
		ϕ_1	-0.500	-0.503	-0.502	0.034
		ϕ_2	0.400	0.405	0.404	0.033
		σ_0	0.500	0.523	0.509	0.062
		ω_0	-0.300	-0.300	-0.299	0.059
IV	150	β_1	1.000	1.600	0.996	1.427
		ϕ_1	-0.700	-0.992	-0.761	0.667
		Φ_1	0.200	0.224	0.237	0.328
		σ	0.500	0.749	0.556	0.382
		ω_0	0.300	0.309	0.307	0.175
	300	β_1	1.000	1.078	0.980	0.341
		ϕ_1	-0.700	-0.727	-0.710	0.108
		Φ_1	0.200	0.216	0.208	0.085
		σ_0	0.500	0.580	0.514	0.158
		ω_0	0.300	0.304	0.303	0.120
	600	β_1	1.000	1.021	0.990	0.171
		ϕ_1	-0.700	-0.709	-0.703	0.066
		ϕ_2	0.200	0.207	0.204	0.053
		σ_0	0.500	0.535	0.508	0.086
		ω_0	0.300	0.301	0.300	0.084
	1200	β_1	1.000	1.007	0.993	0.112
		ϕ_1	-0.700	-0.708	-0.705	0.044
		ϕ_2	0.200	0.203	0.202	0.036
		σ_0	0.500	0.515	0.503	0.054
		ω_0	0.300	0.300	0.301	0.059

It is important to note that in the scenario with the highest number of zeros, approximately $\frac{\exp(0.300)}{1+\exp(0.300)} \approx 57.44\%$ zeros, the estimates for β_1 and σ_0 were significantly impaired when $n = 150$. These estimates improved considerably when the sample size was increased to 1200, again showing convergence toward the true value as the sample size grows.

The findings from the simulation process suggest that, in the evaluated scenarios, the parameter estimates converge toward the true values in both the GPO(μ_t, σ) model and the ZAGPO(μ_t, σ, ν) model. However, we also observed the need for larger sample sizes, especially in situations involving seasonal components and a high number of zeros.

5 TUBERCULOSIS NOTIFICATIONS: A TIME-SERIES ANALYSIS

Tuberculosis (TB) is an infectious disease caused by the *Mycobacterium tuberculosis* that commonly affects the lungs but can also affect other organs, such as the kidneys, and other systems of the human body (Brasil, 2024; World Health Organization, 2023). The pulmonary form is the most frequent and it contributes significantly to the transmission chain of the disease (Brasil, 2024). Its transmission occurs via the respiratory route when susceptible individuals inhale aerosols containing the bacillus emitted by people with active tuberculosis, whether pulmonary or laryngeal, who are not undergoing treatment (Brasil, 2024; World Health Organization, 2023).

In Brazil, Ministério da Saúde (2024) highlights the urgent need to implement policies to achieve the goal of eliminating TB by 2030, as set by the government in accordance with the recommendations of the United Nations. In 2022, TB was the second leading cause of mortality from a single infectious agent, underscoring the urgency of these measures to achieve the set goals (Ministério da Saúde, 2024).

In 2023, 11 of Brazil's 27 federative units (UFs) reported TB incidence rates higher than the national average of 37 cases per 100,000 inhabitants. Among these 11 UFs, the state of *Roraima*, located at 1.5958°N, 60.5821°W, had the highest incidence (87.50 cases per 100,000 inhabitants). It was followed by the state of *Amazonas*, with 81.60 cases per 100,000 inhabitants, and *Rio de Janeiro*, with an incidence of 70.70 (Ministério da Saúde, 2024).

Although the incidence rate was lower than the national average in 2023, the state of *Minas Gerais*, located at 17.9302°S, 43.7908°W, has reported at least one case of TB in approximately 65% of its 853 municipalities. With respect to the health regions of this state, which are groups of neighboring municipalities designed to integrate health actions (Ministério da Saúde, 2022), the regions of *Belo Horizonte*, *Juiz de Fora*, and *Uberlândia* accounted for the highest number of cases (Secretaria de Estado de Saúde de Minas Gerais, 2023).

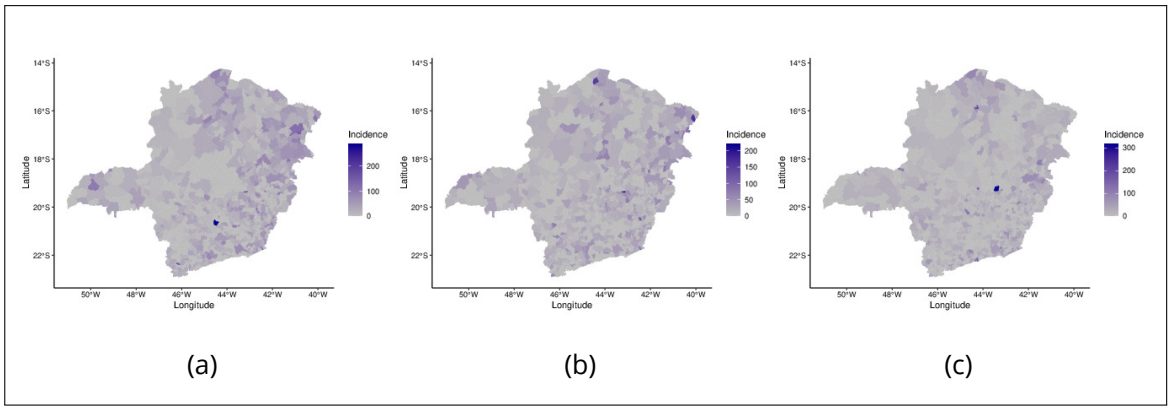
This application aims to analyze the number of tuberculosis notifications in *Minas Gerais* from January 2002 to December 2023, by sex (female and male) and age groups (0-14; 15-59; and 60+ years) using the GPO(μ_t, σ) model. Data on confirmed TB cases were extracted from the *Sistema de Informação de Agravos de Notificação* (Departamento de Informática do Sistema Único de Saúde, 2024b), and resident population data were obtained from Departamento de Informática do Sistema Único de Saúde (2024a) on June 27, 2024. The results of the analysis are presented in subsection 5.1.

5.1 Background of Tuberculosis in Minas Gerais, Brazil

In Figure 1, we present the spatial distribution of TB incidence per 100,000 inhabitants for the years 2002, 2010, and 2022, shown in (a), (b), and (c), respectively[1]. According to the data, the incidence in MG was 31.32 in 2002, 22.20 in 2010, and 23.00 in 2022. In 2023, the incidence was 19.50 cases per 100,000 inhabitants (Secretaria de Estado de Saúde de Minas Gerais, 2024).

[1]The year 2023 is not displayed in this figure because data on the resident population was unavailable when the data was obtained.

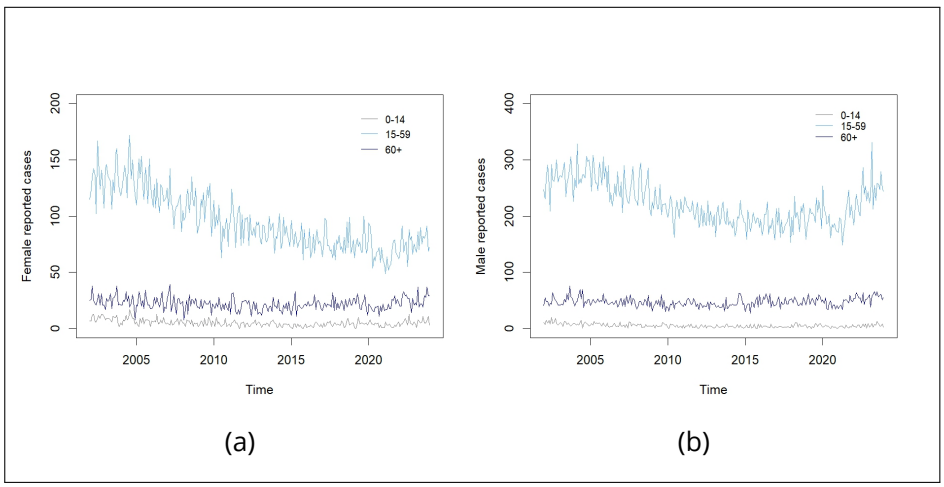
Figure 1 – TB incidence per 100,000 inhabitants in the state of Minas Gerais, Brazil, calculated for the years 2002, 2010, and 2022, shown in panels (a), (b), and (c), respectively. .



For the years shown in Figure 1, the municipalities of *Passa Tempo*, located at 20.6524ºS, 44.4955ºW (with 290 cases per 100,000 inhabitants in 2002), *Passabém*, located at 19.3524ºS, 43.1387ºW (with 222.47 cases per 100,000 inhabitants in 2010), and *Morro do Pilar*, located at 19.2170ºS, 43.3769ºW (with 319.18 cases per 100,000 inhabitants in 2022), had the highest TB incidences compared to other municipalities; however, these are considered small municipalities.

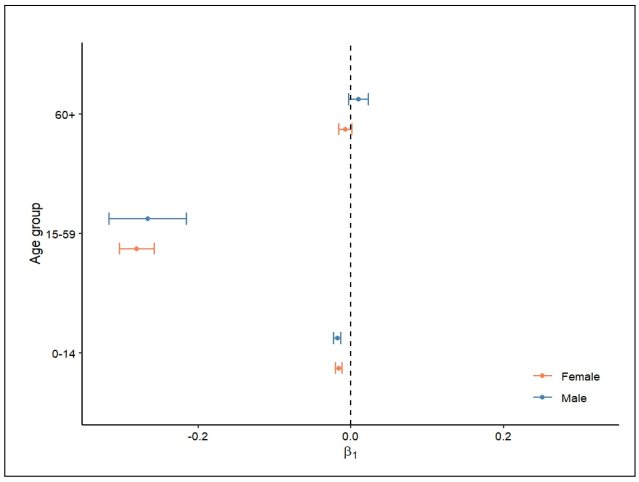
Regarding the number of reported cases for the age groups 0-14, 15-59, and 60 years or older, the distribution is presented in Figure 2. It can be observed that the 15-59 age group had the highest number of TB notifications, with a predominance of male cases (Figure 2b).

Figure 2 – Number of Tuberculosis (TB) cases in Minas Gerais, Brazil... stratified by sex and age group.



Through an initial study of the trend in the number of cases using a simple linear model, we found evidence of a reduction in the number of cases for the age groups 0-14 and 15-59, for both females and males. However, for the population aged 60 or older, there was no evidence of a reduction in cases. These results are presented in Figure 3, which shows the estimates of the parameter β_1, associated with the trend, along with the corresponding 95% confidence intervals.

Figure 3 – Coefficients β_1 of the simple linear regression models, given by the model $Y_t = \beta_0 + \beta_1 t + e_t$, fitted to TB case data for the age groups 0 to 14, 15 to 59, and 60 or older.



Although the time series of the number of cases are not directly comparable due to differences in population composition, the results presented in Figure 3 can be used to monitor trends in cases over time, as well as to help evaluate and quantify the effects of TB-targeted policies in the state.

For the modeling of TB time series, we used the models presented in Table 5. The orders were chosen based on the behavior of the autocorrelation and partial autocorrelation functions of the series. Regarding the seasonal component, the test by Canova and Hansen (1995) at the $\alpha = 0.05$ significance level did not suggest rejection of the null hypothesis of seasonal stationarity for any of the series.

The prior structure described in section 3 was adopted for all models, using two chains with 15,000 final samples, a warm-up period of 5,000 iterations, and a thinning interval of 2. The models were trained using data from January 2002 to August 2023, with the last four observations of 2023 used to analyze the forecasting error through the mean absolute percentage error (MAPE) and the root mean squared error (RMSE).

The posterior mean estimates of the parameters are presented in Table 5, along with the standard deviation (SD), the $\widehat{R}$-split, and the p-value of the Q statistic from Box and Pierce (1970), evaluated up to lag 20. For residual analysis, we used randomized and normalized quantile residuals (Dunn & Smyth, 1996) and calculated the mean p-value of the $Q(\cdot)$ statistic for one hundred realizations.

According to the results described in Table 5, we observed a seasonal component associated with lag 12 in the case series for males aged 15 to 59. For individuals aged 60 or older, the model results suggest a mean of approximately 48 ($\approx \exp(3.864)$) monthly cases for males and 22 ($\approx \exp(3.090)$) monthly cases for females.

Table 5 – Estimates of the parameters of the GPO(μ_t, σ) models used to analyze the time series of reported TB cases, by sex and age group, in *Minas Gerais*, Brazil.

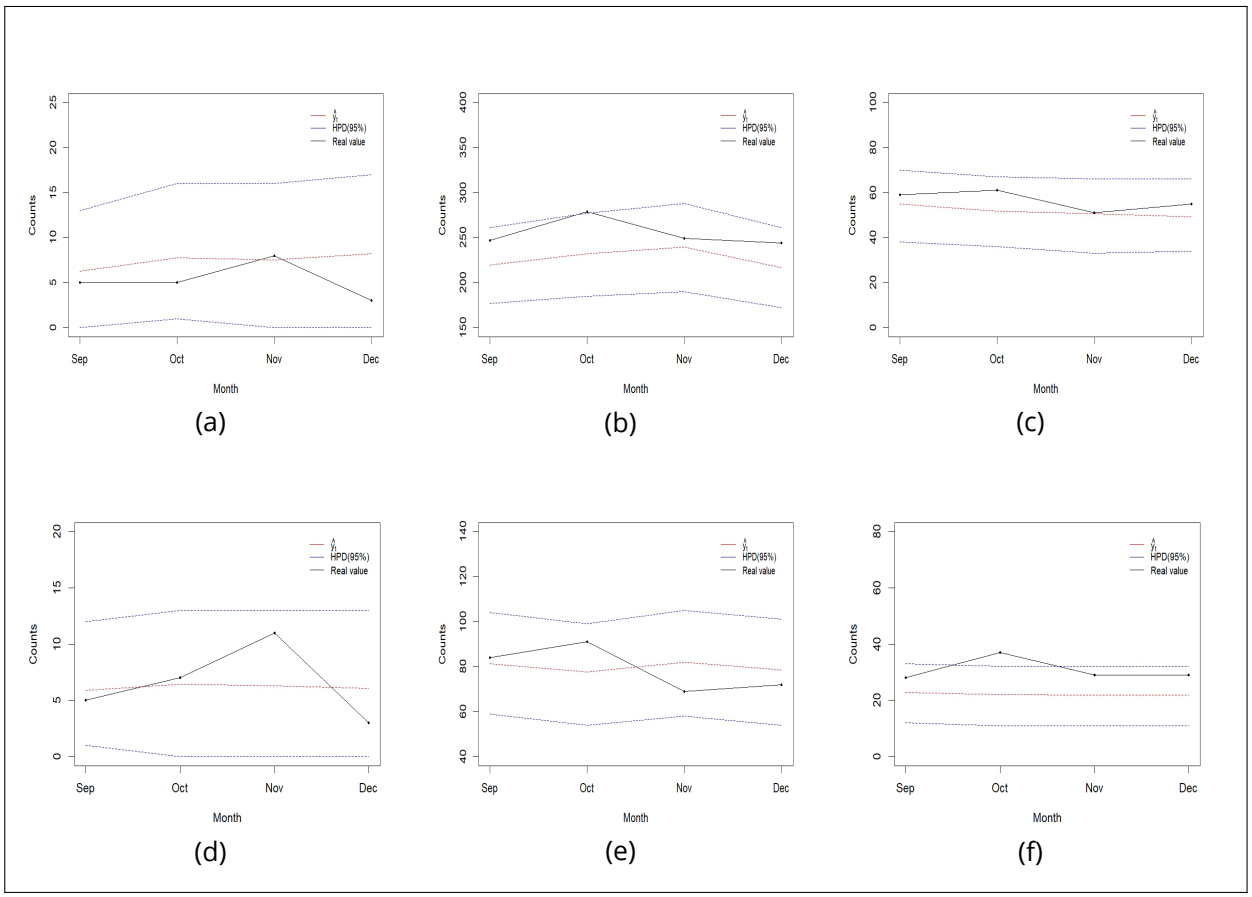
Group	Model structure	Parameters	Mean	SD	$\widehat{R}$-split	Q(20)
		Male				
0-14	ARIMA(4,1,0)	ϕ_1	-0.866	0.047	1.000	0.290
		ϕ_2	-0.670	0.060	1.000	
		ϕ_3	-0.474	0.061	1.000	
		ϕ_4	-0.275	0.050	1.000	
		σ_0	0.068	0.013	1.000	
15-59	SARIMA(4,1,0)(1,0,0)$_{12}$	ϕ_1	-0.721	0.063	1.000	0.853
		ϕ_2	-0.532	0.073	1.000	
		ϕ_3	-0.468	0.074	1.000	
		ϕ_4	-0.340	0.065	1.000	
		Φ_1	0.282	0.065	1.000	
		σ_0	0.002	$<$0.001	1.000	
60+	ARIMA(2,0,0)	β_0	3.864	0.021	1.000	0.173
		ϕ_1	0.175	0.059	1.000	
		ϕ_2	0.332	0.060	1.000	
		σ_0	0.002	0.001	1.000	
		Female				
0-14	ARIMA(4,1,0)	ϕ_1	-0.688	0.055	1.000	0.082
		ϕ_2	-0.614	0.063	1.000	
		ϕ_3	-0.362	0.063	1.000	
		ϕ_4	-0.170	0.053	1.000	
		σ_0	0.051	0.012	1.000	
15-59	ARIMA(5,1,0)	ϕ_1	-0.797	0.060	1.000	0.570
		ϕ_2	-0.654	0.074	1.000	
		ϕ_3	-0.543	0.077	1.000	
		ϕ_4	-0.421	0.074	1.000	
		ϕ_5	-0.298	0.061	1.000	
		σ_0	0.004	0.001	1.000	
60+	ARIMA(1,0,0)	β_1	3.090	0.018	1.000	0.116
		ϕ_1	0.147	0.060	1.000	
		σ_0	0.007	0.002	1.000	

Note: The sensitivity analysis performed for the prior distributions of the parameters ϕ and β, using variances equal to 10 and 100, showed small differences in the Widely Applicable Information Criterion and Expected Log Predictive Density, indicating a certain stability of the model estimates.

The forecasts for the months from September to December of 2023 are presented in Figure 4, along with the 95% credibility intervals and the actual values. The following forecasting errors were obtained for the female and male groups aged 0–14, 15–59, and 60 or older, respectively: MAPE = {42.80, 11.50, 26.87} and {63.87, 10.57, 8.49}, and the

corresponding RMSE for the GPO(μ_t, σ) model were {2.87, 9.88, 9.35} and {3.00, 30.58, 5.78} for females and males groups, respectively. As shown in Figure 4, the number of cases estimated for October in the male group aged 15 to 59 exceeded the upper limit, contributing to a higher forecasting error. A similar pattern occurred for the forecast in the female group aged 60 or older.

Figure 4 – Forecasts of the number of TB cases for the months from September to December of 2023 compared with the actual values, for the state of *Minas Gerais*, covering the following age groups for males: (a) 0-14; (b) 15-59; and (c) 60 or older; and for females: (d) 0-14; (e) 15-59; and (f) 60 or older. The red dashed line represents the predicted values, and the blue dashed bands indicate the 95% HPD intervals.



To perform a comparison, we also considered models from the traditional SARIMA(p, d, q)(P, D, Q)$_s$ class with the same orders presented in Table 5, and parameter estimates obtained via maximum likelihood using the `arima` function from the `stats` package. The following forecasting errors were obtained: MAPE = {50.79, 11.71, 26.83} and {86.72, 4.17, 8.39} for the female and male groups aged 0–14, 15–59, and 60 or older, respectively. The RMSE for this model was {3.01, 10.06, 9.35} and {3.91, 13.32, 5.84} for the female and male groups, respectively.

Thus, we can see that the adoption of the GPO(μ_t, σ) model resulted in a lower average MAPE for forecasting TB cases from September to December of 2023 in *Minas Gerais*, especially for the 0-14 age group. Additionally, including explanatory variables and using five-year age groups are suggestions for future analyses, which could help inform and contribute to policies for the control and prevention of tuberculosis.

6 FINAL REMARKS

In this paper, we examine the utility of the generalized Poisson distribution and its zero-adjusted version for modeling count time series as alternatives for dealing with scenarios of high dispersion and even excess zeros. These approaches can be useful when analyzing epidemiological time series, such as the number of disease cases over time, respecting the characteristics of the response variable. Moreover, the proposed methodology can be extended to other regions and serve as an auxiliary tool for informing public health policies, and can be applied by researchers using the `garmaFit` package.

We adopted a full Bayesian inference and observed an improvement in parameter estimates as the sample size increased, which is a desired characteristic in terms of estimator quality. In certain scenarios, posterior mode estimates perform better than the mean when estimating the dispersion parameter. Particularly, when dealing with excess zeros using the ZAGPO(μ_t, σ, ν) distribution, we observed that the estimates of parameters related to μ_t may be affected as the number of zero counts increases.

As a suggestion for future work, we recommend studying the inclusion of explanatory variables in the equation for ν_t in the ZAGPO(μ_t, σ, ν) model, allowing it to capture and provide more information about the probability of the time series taking a zero value at moments $t + h$, as well as investigating the stability of the estimates and the behavior of the model when dealing with strongly seasonal time series, such as dengue cases.

ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

APPENDIX

Considering the GPO(μ, σ) distribution defined in the parameter space $\mathbf{\Omega} = \{\mu, \sigma \mid \mu > 0, \sigma > 0\}$, we can construct the ZAGPO(μ, σ, ν) distribution as follows:

$$P(Y = y \mid \mu, \sigma, \nu) = \nu + (1 - \nu) \times \times \left[\frac{P(Y = y \mid \mu, \sigma)}{1 - P(Y = 0 \mid \mu, \sigma)} \right], \tag{3}$$

where $P(\cdot \mid \mu, \sigma)$ is the probability function of a random variable with a GPO(μ, σ) distribution, as also shown in Zuo et al. (2021). For example, $P(Y = 0 \mid \mu, \sigma)$ is given by:

$$P(Y = 0 \mid \mu, \sigma) = \exp\left(\frac{-\mu}{1 + \sigma\mu} \right).$$

Note, from Equation (3), that $\nu \in (0, 1)$ is a parameter that measures the exact probability of the variable taking the value zero. Thus, the new parameter space is given by $\mathbf{\Omega} = \{\mu, \sigma, \nu \mid \mu > 0, \sigma > 0, 0 < \nu < 1\}$.

According to the structure presented in expression (3), the ZAGPO(μ, σ, ν) is given as follows:

$$P(Y = y \mid \mu, \sigma, \nu) = \nu + (1 - \nu) \times \times \frac{\left(\frac{\mu}{1+\sigma\mu} \right)^{y} \frac{(1+\sigma y)^{y-1}}{y!} \exp\left[\frac{-\mu(1+\sigma y)}{1+\sigma\mu} \right]}{1 - \exp\left(\frac{-\mu}{1+\sigma\mu} \right)}.$$

Where:

$$E(Y) = \left[\frac{1 - \nu}{1 - \left((1 + \sigma y)^{-1} \exp\left(\frac{-\mu}{1+\sigma\mu} \right) \right)} \right] \mu.$$

REFERENCES

Albarracin, O. Y. E., Alencar, A. P., & Ho, L. L. (2019). Generalized autoregressive and moving average models: Multicollinearity, interpretation and a new modified model. *Journal of Statistical Computation and Simulation*, *89*(10),1819–1840.

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

Bertoli, W., Conceição, K., Andrade, M. G., & Louzada, F. (2021). A new regression model for the analysis of overdispersed and zero-modified count data. *Entropy*, *23*(646),1–25.

Box, G. & Pierce, D. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, *65*(332),1509–1526.

Brasil (2024). *Tuberculose*. *https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/t/tuberculose*.

Briët, O. J. T., Amerasinghe, P. H., & Vounatsou, P. (2013). Generalized seasonal autoregressive integrated moving average models for count data with application to malaria time series with low case numbers. *Plos One*, *8*(6),1–9.

Burda, M. & Maheu, J. (2013). Bayesian adaptive hamiltonian monte carlo with an application to high-dimensional bekk garch models. *Studies in Nonlinear Dynamics and Econometrics*, *17*,345–372.

Canova, F. & Hansen, B. E. (1995). Are seasonal patterns constant over time? a test for seasonal stability. *Journal of Business & Economic Statistics*, *13*(3),237–252.

Conceição, K., Suzuki, A., & Andrade, M. (2021). A bayesian approach for zero-modified skellam model with hamiltonian mcmc. *Statistical Methods & Applications*, *30*(2),747–765.

Davis, R., Dunsmuir, W., & Streett, S. (2003). Observation-driven model for poisson counts. *Biometrika*, *90*,777–790.

Davis, R., Fokianos, K., Holan, S., Joe, H., Livsey, J., Lund, R., Pipiras, V., & Ravishanker, N. (2021). Count time series: A methodological review. *Journal of the American Statistical Association*, *0*(0),1–15.

Departamento de Informática do Sistema Único de Saúde (2024a). *População Residente*. *http://tabnet.datasus.gov.br*.

Departamento de Informática do Sistema Único de Saúde (2024b). *Sistema de Informação de Agravos de Notificação - Tuberculose*. *http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sinannet/cnv/tubercmg.def*.

Duane, S., Kennedy, A., Pendleton, B., & Roweth, D. (1987). Hybrid monte carlo. *Physics Letters B*, *195*(2),216–222.

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

Dunsmuir, W. T. M. & Scott, D. J. (2015). The glarma package for observation-driven time series regression of counts. *Journal of Statistical Software*, *67*(7),1–36.

Franco, G. C., Migon, H. S., & Prates, M. O. (2019). Time series of count data: A review, empirical comparisons and data analysis. *Brazilian Journal of Probability and Statistics*, *33*(4),756–781.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2014). *Bayesian Data Analysis*. Chapman & Hall.

Hoffman, M. D. & Gelman, A. (2014). The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, *15*(47),1593–1623.

Hunt, R. (2022). *Garma: Fitting and Forecasting Gegenbauer ARMA Time Series Models*. *https://CRAN.R-project.org/package=garma*.

Li, W. K. (1994). Time series models based on generalized linear models: Some further results. *Biometrics*, *50*(2),506–511.

McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course With Examples in R and Stan*. Chapman & Hall.

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

Ministério da Saúde (2022). *Regiões de Saúde*. *https://www.gov.br/saude/pt-br/composicao/saps/programa-cuida-mais-brasil/regioes-de-saude*.

Ministério da Saúde (2024). *Boletim Epidemiológico - Tuberculose (2024)*. *https://www.gov.br/aids/pt-br/central-de-conteudo/boletins-epidemiologicos/2024/boletim-epidemiologico-tuberculose-2024/view*.

Pala, L. & Sáfadi, T. (2024). Time-varying zero-adjusted poisson distribution for modeling count time series. *Semina: Ciências Exatas e Tecnológicas*, *45*,1–13.

Pala, L. O. d. O., Carvalho, M. d. M., & Sáfadi, T. (2022). Count time series with excess zeros: A bayesian perspective using zero-adjusted distributions. *Semina: Ciências Exatas e Tecnológicas*, *43*(2),147–160.

Payne, E., Hardin, J., Egede, L., Ramakrishnan, V., Selassie, A., & Gebregziabher, M. (2015). Approaches for dealing with various sources of overdispersion in modeling count data: Scale adjustment versus modeling. *Statistical Methods in Medical Research*, *26*(4),1802–1823.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. *https://www.R-project.org/*.

Sales, L., Alencar, A., & Ho, L. (2022). The berg generalized autoregressive moving average model for count time series. *Computers & Industrial Engineering*, *168*,1–13.

Sathish, V., Mukhopadhyay, S., & Tiwari, R. (2021). Autoregressive and moving average models for zero-inflated count time series. *Statistica Neerlandica*, *76*(2),190–218.

Secretaria de Estado de Saúde de Minas Gerais (2023). *Tuberculose: a Prevenção Começa pela Informação*. *https://www.saude.mg.gov.br/tuberculose*.

Secretaria de Estado de Saúde de Minas Gerais (2024). *Tuberculose*. *https://www.saude.mg.gov.br/termos-de-uso/page/1568-tuberculose*.

Stan Development Team (2022). *RStan: the R interface to Stan*. *https://mc-stan.org/*.

Stasinopoulos, M., Rigby, B., & Eilers, P. (2016). *Gamlss.util: GAMLSS Utilities*. *https://CRAN.R-project.org/package=gamlss.util*.

Stasinopoulos, M. & Rigby, R. (2022). *gamlss.dist: Distributions for Generalized Additive Models for Location Scale and Shape*. *https://CRAN.R-project.org/package=gamlss.dist*.

World Health Organization (2023). *Tuberculosis*. *https://www.who.int/news-room/fact-sheets/detail/tuberculosis*.

Yadav, B., Jeyaseelan, L., Jeyaseelan, V., Durairaj, J., George, S., Selvaraj, K., & Bangdiwala, S. I. (2021). Can generalized poisson model replace any other count data models? an evaluation. *Clinical Epidemiology and Global Health*, *11*,1–6.

Yadav, B., Jeyaseelan, L., Sappani, M., Mani, T., George, S., & Bangdiwala, S. I. (2024). Modeling zero-inflated count data using generalized poisson and ordinal logistic regression models in medical research. *Oman Medical Journal*, *39*,1.

Zeger, S. L. & Qaqish, B. (1988). Markov regression models for time series: A quasi-likelihood approach. *Biometrics*, *44*(4),1019–1031.

Zuo, G., Fu, K., Dai, X., & Zhang, L. (2021). Generalized poisson hurdle model for count data and its application in ear disease. *Entropy*, *23*,9.

Author contributions

1 – Luiz Otávio de Oliveira Pala)
PhD in Statistics and Agricultural Experimentation
https://orcid.org/0000-0002-9941-7951 • luizpala@ufla.br
Contribution: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review editing

2 – Ryan Rodrigo Oliveira de Paula
Medical Degree from the Federal University of Lavras
https://orcid.org/0009-0000-5304-4347 • ryan.paula@estudante.ufla.br
Contribution: Conceptualization, Data curation, Investigation, Resources, Visualization, Writing – original draft, Writing – review editing.

3 – Luciano José Pereira
PhD in Physiology from the State University of Campinas
https://orcid.org/0000-0002-0502-2554 • lucianojosepereira@ufla.br
Contribution: Conceptualization, Investigation, Resources, Supervision, Visualization, Writing – original draft, Writing – review editing.

4 – Thelma Sáfadi
PhD in Statistics from the University of São Paulo
https://orcid.org/0000-0002-5999-4121 • safadi@ufla.br
Contribution: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review editing.

How to cite this article

Pala, L. O. O., Paula, R. R. O., Pereira, L. J., & Sáfadi, T. Bayesian modeling of Tuberculosis notifications using the generalized Poisson distribution. *Ciência e Natura*, Santa Maria, v. 48, e91810, 2026. https://doi.org/10.5902/2179460X91810