



Research Article

Copyright © Norou Diawara

Poisson Regression Model with Application to Wastewater Surveillance Under a Threshold Linear Mixed Model for COVID-19 Sensitivity Rates

Norou Diawara^{1*}, Hueiwang Anna Jeng², Kyle Curtis⁴, Raul Gonzalez⁴, Nancy Welch⁵, Cynthia Jackson⁵, Rekha Singh³, David Jurgens⁶, Sasanka Adikari² and Omotomilola Jegede²

¹Department of Mathematics & Statistics, Old Dominion University Norfolk VA, United States of America

²School of Community and Environmental Health, Old Dominion University, Norfolk VA, United States of America

³Virginia Department of Health, Richmond, United States of America

⁴Virginia Department of Health, Richmond, United States of America

Technical Services Division, Hampton Road Sanitation District, United States of America

⁵Chesapeake Health Department, Chesapeake, United States of America

⁶Public Utilities, City of Chesapeake, United States of America

***Corresponding author:** Norou Diawara, Department of Mathematics & Statistics, College of Sciences, Old Dominion University Norfolk VA, United States of America

To Cite This Article: Norou Diawara*, Hueiwang Anna Jeng, Kyle Curtis, Raul Gonzalez, Nancy Welch, et al. Poisson Regression Model with Application to Wastewater Surveillance Under a Threshold Linear Mixed Model for COVID-19 Sensitivity Rates. *Am J Biomed Sci & Res.* 2023 20(2) AJBSR.MS.ID.002683, DOI: [10.34297/AJBSR.2023.20.002683](https://doi.org/10.34297/AJBSR.2023.20.002683)

Received: 📅 September 15, 2023; **Published:** 📅 September 25, 2023

Abstract

A Threshold Linear Mixed Model (TLMM) has been developed to identify specific thresholds based on wastewater SARS-CoV-2 viral concentrations, which reflect COVID-19 cases. The thresholds can guide decisions regarding public health responses and prevention measures. To assess the practical application of TLMM, a simple simulation was conducted using a sample size of 100 and 500 replications. The simulation allowed for comparing parameter estimators by assessing bias and standard deviation and the root of the mean square error. The model and estimation procedures were applied to reported wastewater and clinic data to test its application for real-world scenarios. Our results demonstrated the efficacy of TLMM in selecting threshold values corresponding to specific levels of wastewater SARS CoV-2 vial concentrations. In particular, TLMM successfully determined threshold values of 0.2 and 0.3, corresponding wastewater SARS-CoV-2 viral concentrations of 2530.1 gene copies/μL and 7,432.6 gene copies/μL, respectively. These values were indicative of a concerning level of COVID-19 cases. However, threshold values at or above 0.5 were associated with a need for warranted public health responses. TLMM presents a valuable modeling approach for effectively determining critical thresholds for wastewater SARS-CoV-2 viral concentrations, guiding targeted public health actions to address the ongoing pandemic.

Keywords: Linear mixed model, Threshold, COVID-19, Wastewater levels

Introduction

On February 11, 2020, the WHO officially identified SARS-CoV-2, commonly referred to as coronavirus [1]. As of July 28, 2022,

the global tally at over 571,198,904 confirmed cases of COVID-19 and over 6,387,863 deaths [2,3]. As the pandemic has evolved, the



efforts to manage the COVID-19 pandemic in the communities have shifted from focusing solely on eliminating SARS-CoV-2 transmission. Instead, there is now an emphasis on identifying and utilizing more relevant metrics and tools to prevent and manage the virus (US CDC, 2022).

Wastewater-based surveillance has emerged as a valuable tool for monitoring the presence of COVID-19 cases within communities across many countries (Agrawal, et al., 2021) [4,5]. This approach is crucial in determining testing resource allocation and preventive measures. The rationale for the surveillance lies in its ability to detect SARS-CoV-2 RNA in symptomatic and asymptomatic individuals [6-8]. Also, wastewater-based analysis is more efficient as it reduces the number of tests required to assess a large population. Furthermore, well-established molecular techniques are available for SARS-CoV-2 detection in wastewater, but the research for successful detection of SARS-COV2 in wastewater is ongoing [5,9,10].

The US Centers for Disease Control and Prevention (CDC) launched the National Wastewater Surveillance System in September 2020. This system holds the potential for early detection of COVID-19 outbreaks, often days or weeks ahead of clinical testing methods [11,12]. However, to fully harness its potential, the national coordination and standardized testing, analysis, and practice shall be tailored to the specific needs of individual communities [11], enabling local public health officials to craft community- or city-specific management strategies.

Recent modelling effort has leveraged wastewater surveillance data to forecast COVID-19 cases and trends using various techniques, e.g., the Autoregressive Integrated Moving Average (ARIMA) [13], Neural Network Graphs [14], and Random Forest Method [5]. Our newly developed model, the Gaussian copula marginal regression model, has demonstrated a significant correlation between wastewater SARS-CoV-2 RNA concentrations and COVID-19 cases, allowing for predicting COVID-19 trends [15]. However, this modelling effort may not be directly applied to establish specific thresholds of wastewater SARS-CoV-2 viral concentration for identifying communities or regions where sizeable COVID-19 cases warrant public health response and actions.

In our current study, we have introduced Threshold Linear Mixed Models (TLMMs), an extension of simple linear models, to assess thresholds of wastewater SARS-CoV RNA concentrations, which reflect certain levels of COVID-19 cases calling for public health response. The TLMMs are particularly well-suited for fixed and random effects, and non-independence data as generalized linear models are less appropriate. The temporal dependence of COVID-19 case counts is a notable example. The TLMMs included a simple threshold marginal which can pinpoint the sensitivity level at which longitudinal outcomes, counts, or rates become detected. This innovative approach can potentially select risk tolerance thresholds while considering covariates consistently used to predict COVID-19 cases for the monitoring and management purpose.

Methodology

Wastewater Viral Load Data

The wastewater samples were collected from five pumping stations in Chesapeake VA, from June 2021 to June 2022 [16]. The selection of pumping stations was based on the need for consistency with clinical data and a sufficient sample size for merging clinical cases and wastewater SARS-CoV-2 RNA concentration data. A 1 L grab sample was collected weekly between 8:00 a.m. and 11:00 a.m. at each pumping station. All the samples were then returned to the Hampton Road Sanitation District laboratories, and wastewater SARS CoV-2 RNA concentrations recorded as gene copies/100mL were quantified using the Reverse transcription droplet digital PCR (RT-ddPCR) method [10]. If a SARS-CoV-2 RNA concentration was below the Limit of Detection (LOD) the viral load was calculated based on half of the LOD.

Clinic Data

Clinical data related to the COVID-19 pandemic included day-to-day reported cases of COVID-19. The cases were registered in the statewide database maintained by the Virginia Chesapeake Health Department, which collects information on clinical cases related to COVID-19 infection. The cases occurred in the period of June 2021 to June 2022, consistent with the timeframe of the wastewater sampling and data collection.

Statistical Model and Notation

The model proposed is an extension from the Ge, et al., [16] paper. A column vector is denoted as $Y_i = (y_{i1}, y_{i2}, \dots, y_{in_i})'$ for the longitudinal measurements of the i th pump station with n_i observations.

The j^{th} element in y_{ij} is the j^{th} observation of pump station i measured at time t_{ij} , where n_i is the number of observations from the i^{th} pump station, $i=1,2,\dots,N$ and N is the number of pump stations.

For each pump station, $X_i = (x_{i1}, x_{i2}, \dots, x_{in_i})'$ is a $(n_i \times P)$ designed matrix of covariates for fixed effect β .

$Z_i = (z_{i1}, z_{i2}, \dots, z_{in_i})'$ is a $(n_i \times Q)$ designed matrix of covariates for random effect α_i .

Substituting in b_i , an indicator of the treatment received by pump station i with either $b_i = 1$ if the pump station is detecting a new threshold or $b_i = 0$ if not.

Denote W_i as a continuous covariate, wastewater SARS-CoV-2 RNA concentration, at the baseline for pump station i and assume two subgroups of pump stations can be defined based on whether the wastewater reading N2Concentration or their transformed values $w_i = \frac{N2conc - \min(N2conc)}{\max(N2conc) - \min(N2conc)}$ exceeds an unknown threshold denoted as c . This threshold, often referred to as the cut point or cut-off point, is a value that we aim to estimate. The threshold c gov-

erns the two-part decision steps, which mirrors the hurdle model cross-sectional: “below-the-threshold” group for no immediate public health response action vs. “above-the-threshold” group for warranted public health response action. Within the “above-the-threshold” group, we further differentiate two subgroups: the lower and higher risk levels associated with the wastewater SARS-CoV-2 RNA concentrations. This differentiation helps better understand the degree of risk in this category. The rationale for employing these thresholds lies in the established link between wastewater SARS CoV-2 RNA concentrations and the occurrence of COVID-19 cases [17]. That means wastewater SARS CoV-2 RNA concentrations can reflect COVID-19 cases within the community. By measuring wastewater SARS CoV-2 RNA concentration with these thresholds, public health officials can gain valuable guidance on when and where to implement targeted public health responses to control the spread of the disease effectively.

The following threshold linear mixed model to assess the potential differential cutoff levels and effects between these two subgroups proposed is as follows:

$$y_i = X_i\beta + Z_i\alpha_i + \eta_i I(W_i > c) + \eta_2 b_i I(w_i > c) + \varepsilon_i \quad (1)$$

where $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{im})'$ is a vector of random errors and 1 is a n_i – dimensional vector with its all elements as 1.

From Equation (1), the response y_{ij} of pump station i measured at the time t_{ij} is modelled by three components:

1. the fixed effects of all covariates $x'_{ij}\beta + \eta_i I(w_i > c) + \eta_2 b_i I(w_i > c)$,
2. the pump station random effect $z'_{ij}\alpha_i$ and
3. the random error ε_{ij} independent and identically distributed (iid) $N(0, \sigma^2)$.

The columns of X_i may include intercept, time or its function, vaccination information, societal behaviours, treatment, and other confounding variables, and we assume that the columns of Z_i are a subset of the columns of X_i .

In order to simplify the presentation, the model equation (1) is re-written in the matrix form defined as

$$Y = X\beta + W\eta + Z\alpha + \varepsilon \quad (2)$$

$$Y = [Y'_1, Y'_2, \dots, Y'_N]'$$

$$X = [X'_1, X'_2, \dots, X'_N]'$$

$$\alpha = (\alpha'_1, \alpha'_2, \dots, \alpha'_N)'$$

$$\varepsilon = (\varepsilon'_1, \varepsilon'_2, \dots, \varepsilon'_N)'$$

$$W = (W'_1, W'_2, \dots, W'_N)'$$

and

$$Z = \begin{pmatrix} Z_1 & 0 & 0 & \dots & 0 \\ 0 & Z_2 & 0 & \dots & 0 \\ & & \vdots & & \\ & & 0 & \dots & Z_N \end{pmatrix};$$

$$W_i = \begin{pmatrix} I(w_i > c) & b_i & \times I(w_i > c) \\ I(w_i > c) & b_i & \times I(w_i > c) \\ & \vdots & \\ I(w_i > c) & b_i & \times I(w_i > c) \end{pmatrix}$$

$$i = 1, \dots, N.$$

For the vector of random effects α and vector of random errors ε in the model, we assume that $E(\alpha) = 0$ and $E(\varepsilon) = 0$.

It is also assumed that α and ε are independent and distributed as multivariate normal.

Inferences Via Kernel Smoothing

From the assumptions and notations, Y follows a multivariate normal distribution. Then, using $n = \sum_{i=1}^N n_i$ as the total number of observations, and the log-likelihood for the unknown parameters $\theta = (\hat{\alpha}, \sigma^2, \rho^2, \sigma^2)$ in equation 2 based on longitudinal outcomes Y can be written as:

$$l(\theta | Y, X, Z) = -\frac{1}{2} \left\{ \log(2\pi) + n \log \sigma^2 + \log |\mathbf{H}| + \frac{(Y - X\beta - W\eta)' \mathbf{H}^{-1} (Y - X\beta - W\eta)}{\sigma^2} \right\} \quad (3)$$

where \mathbf{H} is the variance of Y .

As mentioned in *Ge, et al., (2020)[16]* and *Brown and Wang, et al., (2007) [17]*, the presence of an indicator function $I(W_i > c)$ in the log-likelihood function is not continuous with respect to the threshold or cut-off point c ; instead of the conventional maximum likelihood theory and algorithm, the approach is to take a kernel smooth function as approximation to the indicator variable $I(W_i > c)$, defined as:

$$\Phi\left(\frac{w_i - c}{h}\right), \text{ where } (4)$$

is the distribution function of the standard normal distribution and h is a bandwidth.

Using the approximation above, we can define a smoothed log-likelihood function by replacing in the definition of in equation 3 with the following:

$$\tilde{W}_i = \begin{pmatrix} \Phi\left(\frac{w_i - c}{h}\right) & b_i & \times \Phi\left(\frac{w_i - c}{h}\right) \\ \Phi\left(\frac{w_i - c}{h}\right) & b_i & \times \Phi\left(\frac{w_i - c}{h}\right) \\ & \vdots & \\ \Phi\left(\frac{w_i - c}{h}\right) & b_i & \times \Phi\left(\frac{w_i - c}{h}\right) \end{pmatrix}$$

Therefore, the smoothed log-likelihood function of θ is given by

$$sl(\theta|Y, X, Z) = -\frac{1}{2} \left\{ \log(2\pi) + n \log \sigma^2 + \log |\mathbf{H}| + \frac{(\mathbf{Y} - \mathbf{X}\beta - \mathbf{W}\eta)' \mathbf{H}^{-1} (\mathbf{Y} - \mathbf{X}\beta - \mathbf{W}\eta)}{\sigma^2} \right\} \quad (5)$$

$$\tilde{\mathbf{W}} = (\tilde{\mathbf{W}}'_1, \tilde{\mathbf{W}}'_2, \dots, \tilde{\mathbf{W}}'_N).$$

The Maximum Smoothed Likelihood Estimates (MSLE) of θ can be obtained by maximizing the smoothed log-likelihood function from equation (5) or, equivalently, solving the following equations:

$$sl(\theta|Y, X, Z) = -\frac{1}{2} \left\{ \log(2\pi) + n \log \sigma^2 + \log |\mathbf{H}| + \frac{(\mathbf{Y} - \mathbf{X}\beta - \mathbf{W}\eta)' \mathbf{H}^{-1} (\mathbf{Y} - \mathbf{X}\beta - \mathbf{W}\eta)}{\sigma^2} \right\}$$

Or

$$sl(\theta|Y, X, Z) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} n \log \sigma^2 - \frac{1}{2} \log |\mathbf{H}| - \frac{1}{2} \frac{(\mathbf{Y} - \mathbf{X}\beta - \mathbf{W}\eta)' \mathbf{H}^{-1} (\mathbf{Y} - \mathbf{X}\beta - \mathbf{W}\eta)}{\sigma^2}$$

The standard errors can be calculated based on the inverse matrix of the Hessian matrix $J(\hat{\mathbf{a}}, \eta, c, \rho^2, \sigma^2)$ after the estimates of θ are obtained. For the selection of the bandwidth, based on the theoretical and empirical evidence presented by Lin, et al., (2011) [18] and He, et al., (2018) [19] we propose to select $h = \hat{d}n^{-1/3}$ for the proposed estimator, where \hat{d} is the sample standard deviation of covariate w_i .

Simulation Example

To assess the applicability of the TLMM, a simple simulation

based on a samples of size $n=100$ was conducted. The performance of the parameters in the TLMM proposed are being assessed.

The model is written as:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 b_i + \eta_1 I(w_i > c) + \eta_2 b_i I(w_i > c) + \alpha_i + \varepsilon_{ij},$$

where β_0 is the intercept, α_i is a random variable for the pump station effect, β_1 is the effects of measurement time β_2 , is the treatment indicator η_1 is the subgroup indicator, η_2 is the interaction between the treatment and subgroup indicators, and ε_{ij} is the random error, as described in equation (1).

In the simulation study done for this paper, we assume that pump stations are randomly assigned to either treatment group $b_i=1$ or control group $b_i=0$, each with probability 1/2 and each pump station is measured at the same four equal spaced timepoints 1, 2, 3, and 4. For each pump station i , the baseline covariate w_i and random effect α_i are generated, respectively, from a uniform distribution [0,1]; and a normal distribution with mean zero and covariance σ^2 , while the random error ε_{ij} is generated from a normal distribution with mean zero and covariance $\sigma^2 = 2.25 = 1.5^2$. The coefficients of fixed effects in the model are taken as, $\beta_0=1$, $\beta_1=5$, $\beta_2=2$, $\eta_1=4.9$, $\eta_2=3$ while the true cut-point of w_i for the definition of subgroup is $c = 0.3$. In each scenario, 500 replications are used to obtain the bias, Standard Deviation (SD), and the Root of the Mean Square Error (RMSE) of the parameter estimators under bootstrap. Table 1 describes the parameters derived from the simulated LMM example.

Table 1: shows the parameters, the initial values chosen for the simulation studies and the generated values, the Bias and RMSE.

Parameters	True Values	Initial Values	New Values	Bias	RMSE
β_0	1	0	0.2380772	-0.761922	0.068565
β_1	5	4	4.691046	-0.308954	0.461408
β_2	2	6	1.975093	-0.024907	0.131972
η_1	4.9	4	5.120305	0.220305	0.343441
η_2	3	3	3.260354	0.2603535	0.562743
C	0.3	0.7	0.557878	0.2578781	0.003123
σ	1.5	1.2	1.490339	-0.00966	0.052181
α	0.8	0.5	0.172179	-0.62782	0.018352

From Table 1, we can see that the bias and RMSE of the estimates for all parameters, except the biases of estimates for σ^2 and q^2 , decrease as the total number of observations increases when h is fixed at the optimal value and RMSE becomes very small for most of parameters. This value of h is very close to that calculated from the optimal bandwidth formula.

We also compared the empirical standard deviations of the parameter estimates with the averages of the estimated standard errors based on the inverse of the Hessian matrix (Table 1).

From the Table 1, the generated values for β_0 , β_1 , β_2 , η_1 , η_2 and c do converge even if they share different characteristics and spreads. To gain more insight, we plot the histograms. The histo-

grams show the convergent value(s) of the parameter. The target solution is consequently obtained and close to the optimal values. The change in the c value, from the initial 0.7 to 0.56, indicates that there is a reliable estimate for the threshold. Out of the 100 observations, when 43 data points are found under the threshold data limit of $c=0.3$, the new estimation method finds 85 of them under the threshold 0.56 (Figure 1).

The plot of the simulated transformed Y values contrasted with the threshold c is described in Figure 1. The gap in the threshold (true vs estimated) values are quite large, but the simulation could gain a lot by considering starting parameter for c closer to 0.3.

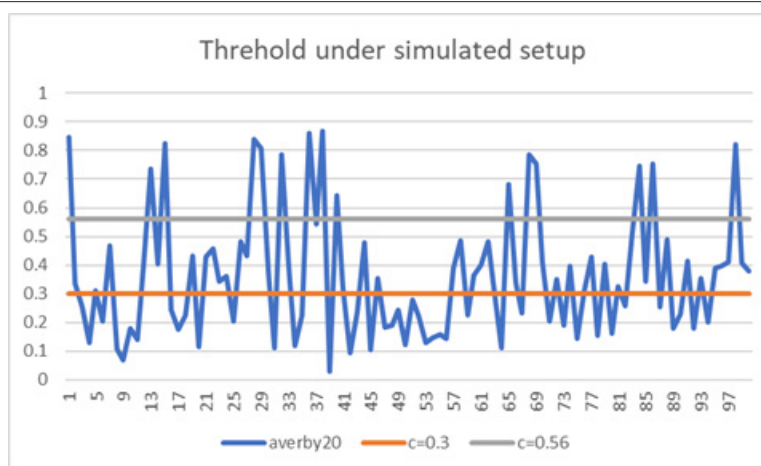


Figure 1: Plot of the simulated data example contrasted with threshold c.

Application to Wastewater and Clinic Data

We applied the proposed model and estimation procedures to the wastewater and clinic data. The model assesses the performance of the proposed cutoff and threshold estimators. The response variable y follows a Poisson and is described as: $y_i \sim P(\mu_i, \theta)$, for $i=1, \dots, 43$, using data from the 43 time points. The various threshold values considered are: $c=0.1, 0.3, 0.5, 0.7$ and 0.9 , corresponding to the values of wastewater viral load concentrations of 2530.1, 7432.6, 12335.0, 17237.4, 22139.9 gene copies/ μL , respectively. In the clinic data, y represents the number of COVID-19 cases, x_1 represents the hospitalization cases, x_2 represents the death cases, x_3 represents the wastewater SARS-CoV concentration, and x_4 represents the interaction between hospitalization cases and the wastewater SARS-CoV concentration. The methodology from Section 2 is applied on each threshold value, and the ranges of observations captured across the COVID-19 case plots were compared.

We display the cutoff structures in Figure 2. The accuracy and

measure of uncertainty of the parameters are plotted via their kernel density estimators in Figure 3. By providing multiple ranges, different tailored decisions may be called for, e.g., no public health response action or warranted public health response action. Values of wastewater readings were deemed to be acceptable, or they could be found to be above tolerance level, requiring further exploration.

The coefficients of fixed effects in the model were estimated, with the different cut points of c . Replications were done for 10,000 bootstrap processes to obtain the bias, SD, RMSE, of the parameter estimators.

The fitted lines resulting from the TLMM at different cutoff values are shown in Figure 2. For any cutoff value between 0.2 (2530.1) and 0.3 (7432.6), the scale of the COVID-19 cases is not alarming. Wastewater SARS-CoV-2 RNA concentrations at the thresholds greater than 0.5 (12335.0) indicate increased COVID-19 cases warranted public health response.

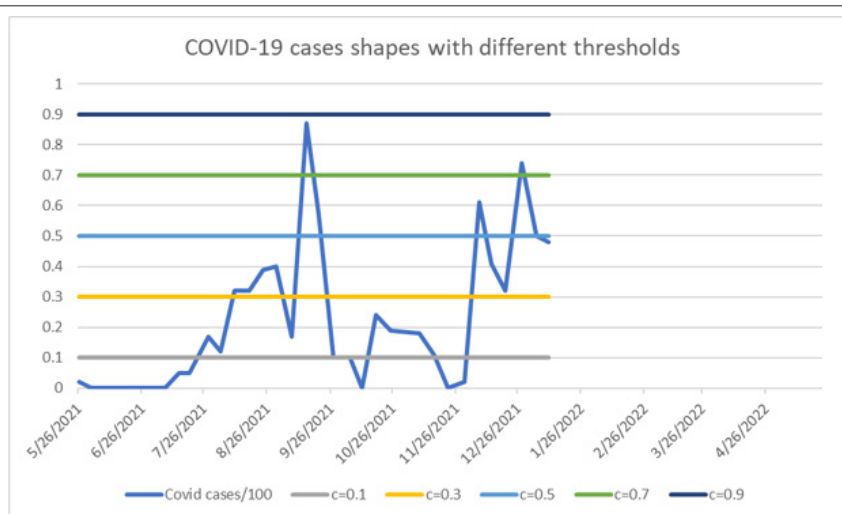


Figure 2: Fitted lines resulting from the projected threshold values with respect to the 100th COVID-19 cases.

Discussion and Conclusions

A threshold linear mixed model based on Poisson regression has been introduced and designed for the identification of subsets of public health responses based on longitudinal COVID-19 cases and wastewater SARS CoV-2 RNA concentrations. This model can include time-dependent effects and both fixed and random effects from covariates in the analysis. To estimate the unknown parameters within this model, statistical procedures are constructed based on a smooth approximation of an indicator function. Efficient computational algorithms are developed to compute the estimators and associated variances. The effectiveness of this proposed approach has been assessed through simulation studies, where its performance was evaluated using SD and RMSE. The results from these simulations indicate that the proposed procedures perform adequately. Furthermore, when applied to the analysis of clinical and environmental data, our modeling procedures have demonstrated practicality and stability. This suggests that the modeling can be effectively employed in real-world scenarios.

The main objective of conducting the threshold linear mixed model was to simultaneously determine the cut point of the wastewater SARS CoV-2 viral concentrations that reflect the levels of the COVID-19 cases calling for public health responses. We therefore propose a smoothing likelihood function to approximate the original likelihood function from the wastewater and clinic data and to make inferences on the model parameters based on the new likelihood function given. The proposed model is quite robust to determine the thresholds based on wastewater SARS-CoV-2 viral concentrations. To truly reflect current scenarios of the levels of COVID-19 cases, the model requires constant updates by ongoing monitoring of wastewater SARS CoV-2 viral concentrations.

The use of wastewater SARS CoV-2 viral concentration to establish thresholds is an innovative approach. Beyond the inherent benefits of wastewater surveillance, the analysis of SRAS-CoV-2 viral concentration in wastewater, coupled with the advanced modelling effort, could provide invaluable data for timely decisions and implementing practical actions and prevention measures at the community level. The study confirmed that there is a relationship between wastewater SARS-CoV-2 viral concentrations and COVID-19 cases. In essence, the use of wastewater SARS-CoV-2 viral concentration as a tool for threshold determination creates new avenues for effective public health responses, providing a balance readiness and early warning settings [20-24].

This research has several limitations, including the privacy of data, political decisions, and economic effects associated with coordinated public health responses based on wastewater surveillance data in general. Additionally, while the proposed procedure has undergone evaluation through simulation studies and an application to both wastewater and clinic data, the goodness of fit of threshold estimates may vary from one community to another. Consequently, the results cannot be extrapolated over a large geographic area, such as an entire or state. This makes the strategy of disease surveillance locally focused. To address this limitation, it is recommended

to consider implementing statewide wastewater surveillance. Such an approach would help address the inherent variability in threshold estimates across different communities [25-28].

Acknowledgment

We acknowledge the staff and laboratory technicians from the Hampton Roads Sanitation District for wastewater sampling and SARS CoV-2 analysis. Also, we thank the staff from the Chesapeake Health Department for participating in data interpretation.

Conflict of Interest

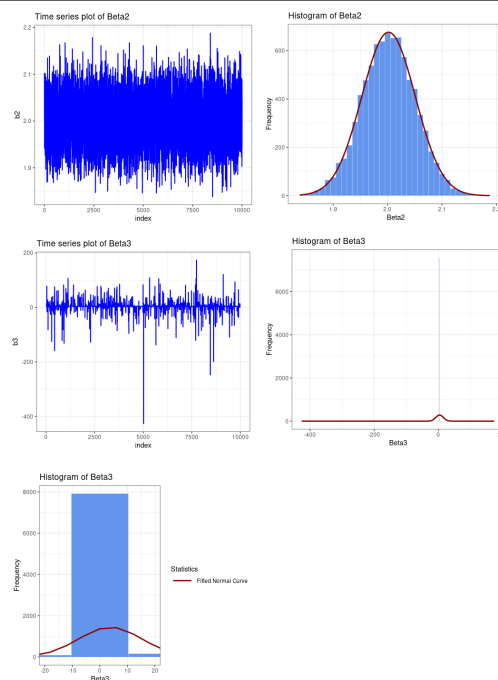
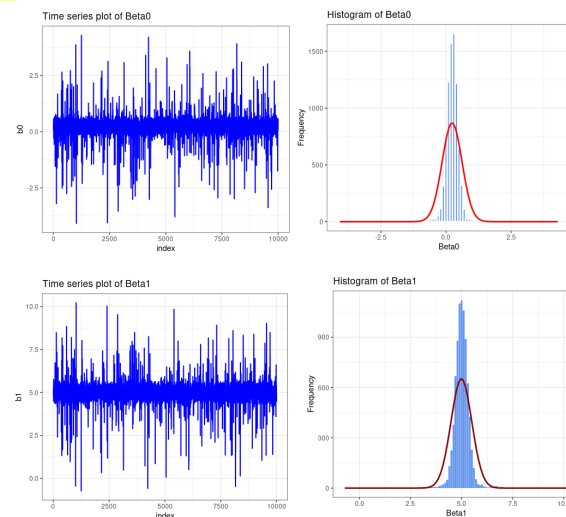
None.

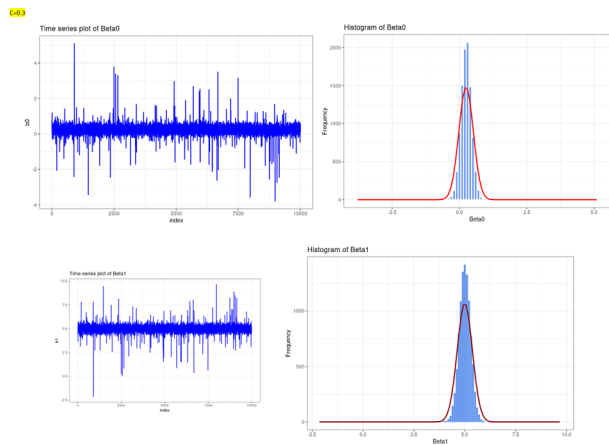
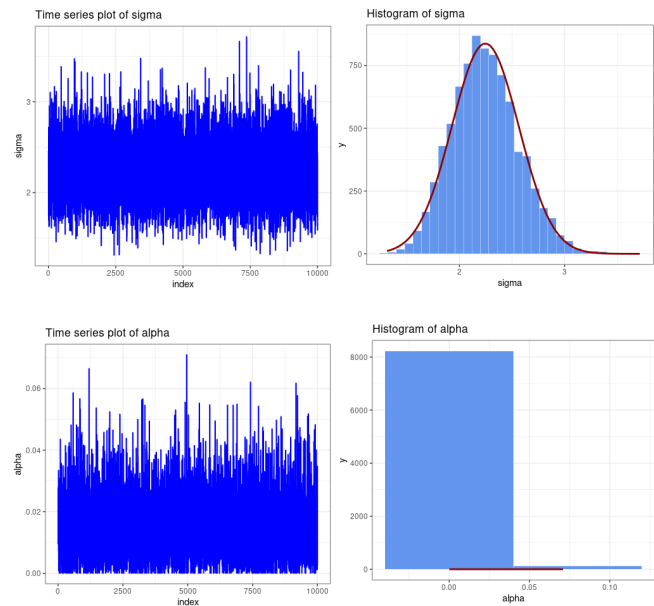
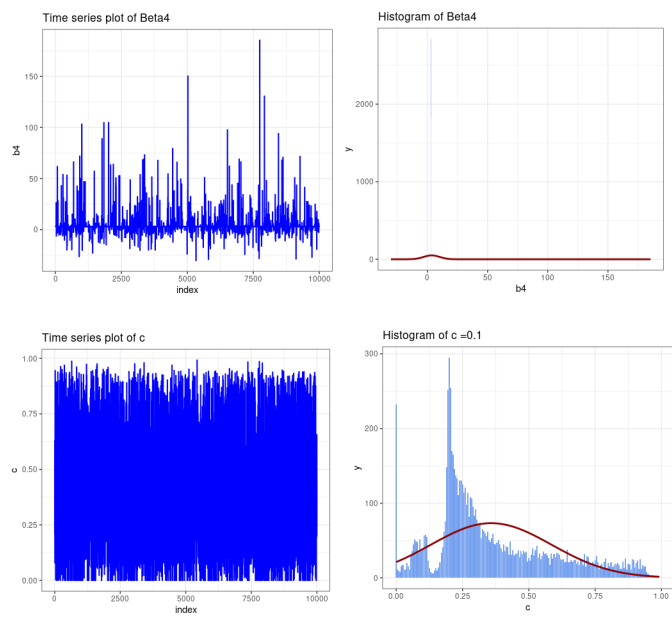
References

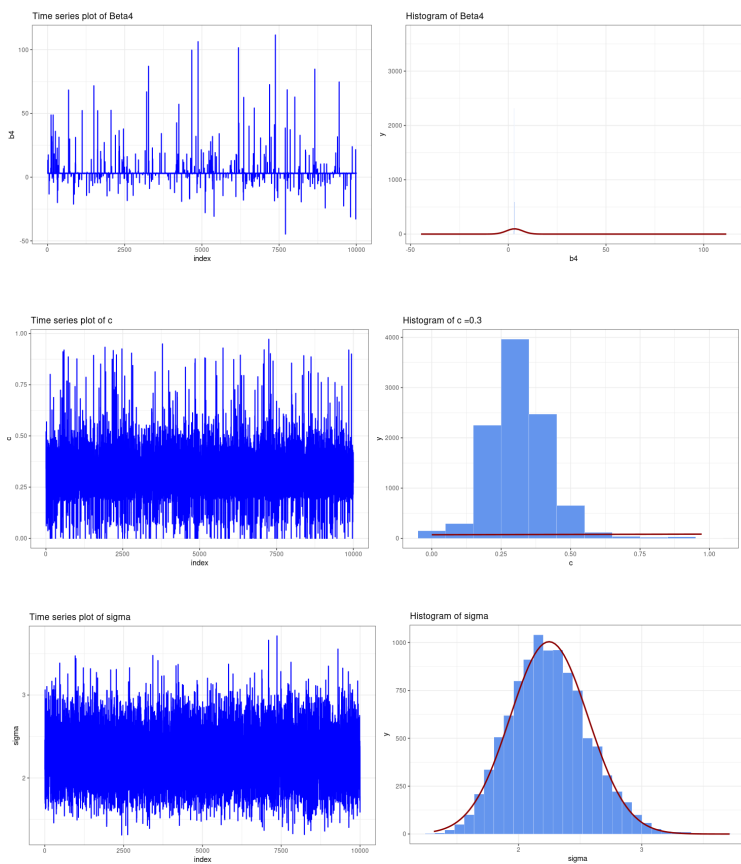
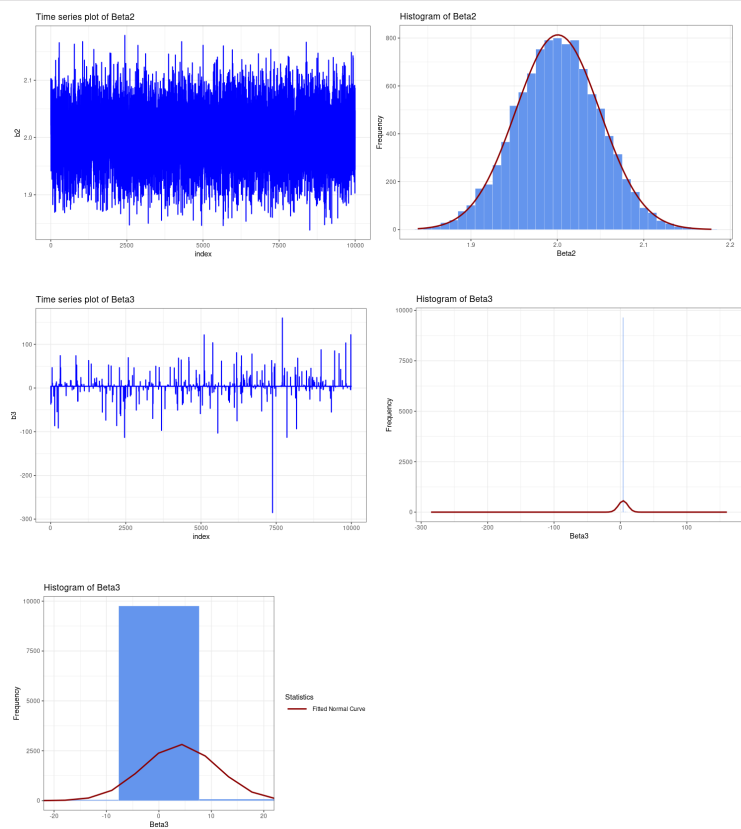
1. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019?>
2. <https://www.vdh.virginia.gov/coronavirus/see-the-numbers/covid-19-in-virginia/>
3. <https://covid19.who.int/>
4. <https://covid.cdc.gov/covid-data-tracker/#datatracker-home>
5. Gonzalez R, Curtis K, Bivins A, Bibby K, Weir MH, et al. (2020) COVID-19 surveillance in Southeastern Virginia using wastewater-based epidemiology. *Water Res* 186: 116296.
6. Parasa Sravanthi, Nageshwar Reddy, Douglas O Faigel, Alessandro Repici, Fabian Emura, et al. (2020) Global impact of the COVID-19 pandemic on endoscopy: an international survey of 252 centers from 55 countries. *Gastroenterology* 159(4): 1579-1581.
7. Peccia Jordan, Alessandro Zulli, Doug E Brackney, Nathan D Grubaugh, Edward H Kaplan, et al. (2020) SARS-CoV-2 RNA concentrations in primary municipal sewage sludge as a leading indicator of COVID-19 outbreak dynamics. *MedRxiv* 2020-05.
8. Tang Yi Wei, Jonathan E Schmitz, David H Persing, Charles W Stratton (2020) Laboratory diagnosis of COVID-19: current issues and challenges. *J Clin Microbiol* 58(6): e00512-e00520.
9. Shah S, Gwee SXW, Ng JQX, Lau N, Koh J, et al. (2022) Wastewater surveillance to infer COVID-19 transmission: A systematic review. *Sci Total Environ* 804: 150060.
10. Bogler A, Packman A, Furman A, Amit Gross, Ariel Kushmaro, et al. (2020) Rethinking wastewater risks and monitoring in light of the COVID-19 pandemic. *Nat Sustain* 3: 981-990.
11. CDC (2022) Overview and Scientific Rationale. Indicators for Monitoring COVID-19 Community Levels and COVID-19 and Implementing COVID-19 Prevention Strategies.
12. (2022) WHO estimates nearly 15 million excess deaths during pandemic, Significance 3.
13. Dansana D, Kumar R, Das Adhikari J, Mohapatra M, Sharma R, et al. (2020) Global forecasting confirmed and fatal cases of COVID-19 outbreak using autoregressive integrated moving average model. *Front Public Health* 8: 580327.
14. Kapoor Amol, Xue Ben, Luyang Liu, Bryan Perozzi, Matt Barnes, et al. (2020) Examining covid-19 forecasting using spatio-temporal graph neural networks. *arXiv preprint arXiv:2007.03113*.
15. Yeşilkanat Cafer Mert (2020) Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm. *Chaos Solitons Fractals* 140: 110210.
16. Jeng HA, R Singh, N Diawara, K Curtis, R Gonzalez, et al. (2023) Application of wastewater-based surveillance and copula time-series model for COVID-19 forecasts. *Sci Total Environ* 885: 163655.

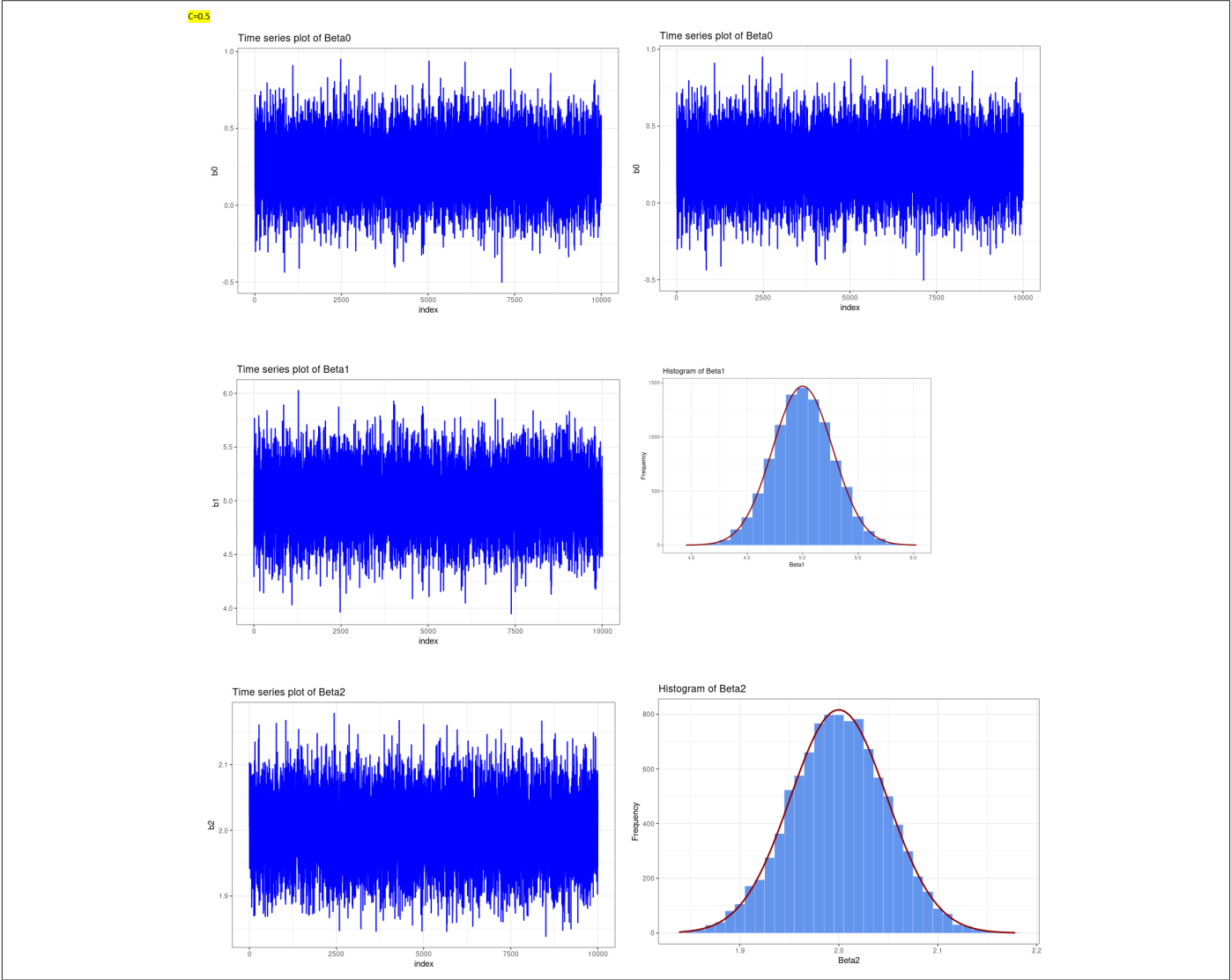
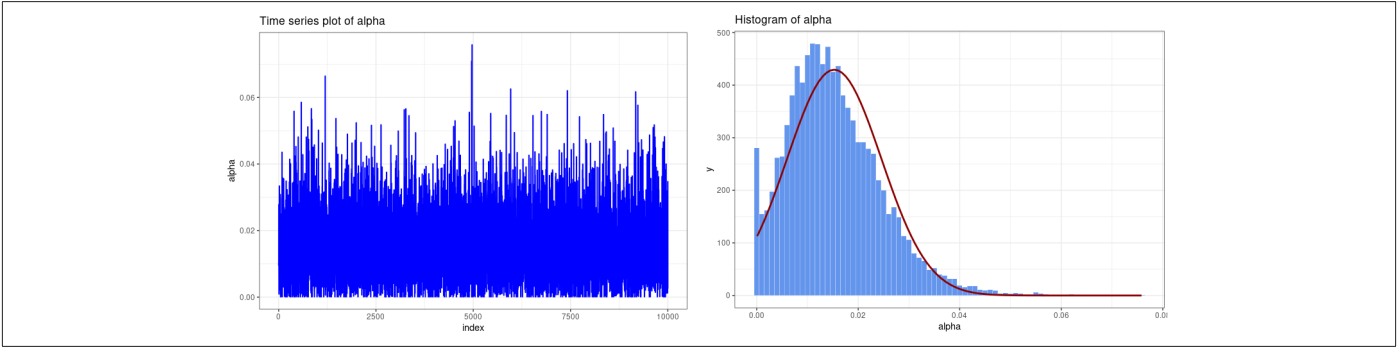
17. Ge X, Peng Y, Tu D (2020) A threshold linear mixed model for identification of treatment-sensitive subsets in a clinical trial based on longitudinal outcomes and a continuous covariate. *Stat Methods Med Res* 29(10): 2919-2931.
18. Brown B, Wang YG 2007 Induced smoothing for rank regression with censored survival times. *Stat Med* 26(4): 828-836.
19. Lin H, Zhou L, Peng H, Xiao Hua (2011) selection and combination of biomarkers using ROC method for disease classification and prediction. *Can J Stat* 39: 324-343.
20. He Y, Lin H and Tu D (2018) A single-index threshold Cox proportional hazard model for identifying a treatment-sensitive subset based on multiple biomarkers. *Stat Med* 37(23): 3267-3279.
21. Hasan SW, Y Ibrahima, M Daou, H Kannout, N Jan, et al. (2021) Detection and quantification of SARS-CoV-2 RNA in wastewater and treated effluents: Surveillance of COVID-19 epidemic in the United Arab Emirates. *Sci Total Environ* 764: 142929.
22. Broyden CG 1970 The convergence of a class of double-rank minimization algorithms 1. General considerations. *IMA J Appl Math* 6(1): 76-90.
23. (2022) WHO estimates nearly 15 million excess deaths during pandemic Significance. 3.
24. <https://stats.oarc.ucla.edu/other/mult-pkg/introduction-to-linear-mixed-models/>
25. <https://www.vdh.virginia.gov/environmental-health/wastewater-surveillance-for-covid-19/>
26. Fletcher R (1970) A new approach to variable metric algorithms. *Comput J* 13: 317-322.
27. Goldfarb D (1970) A family of variable-metric methods derived by variational means. *Math Comput* 24(109): 23-26.
28. Shanno DF (1970) Conditioning of quasi-newton methods for function minimization. *Math Comput* 24: 647-656.

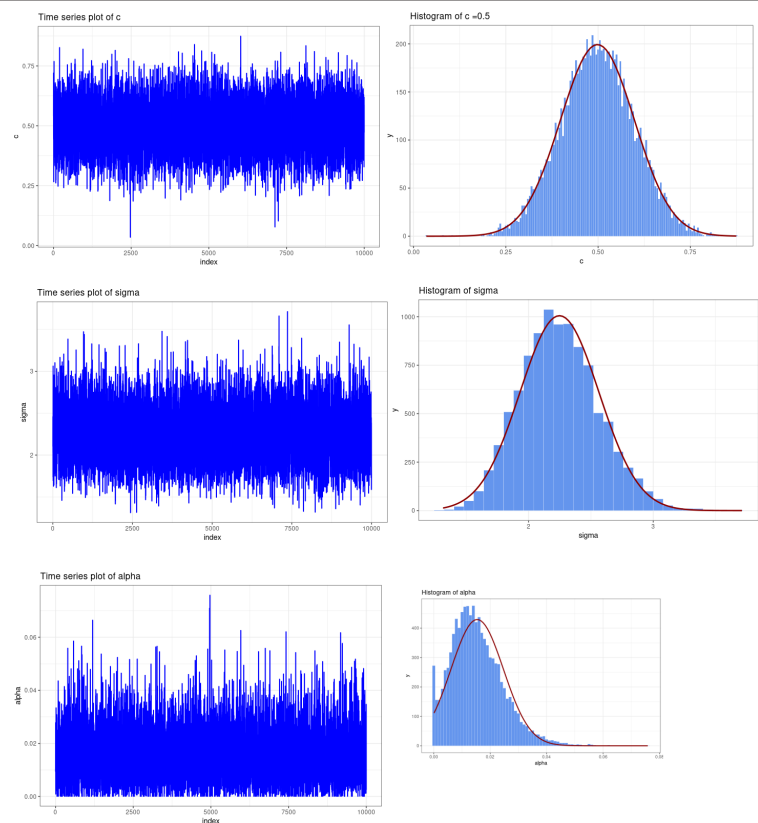
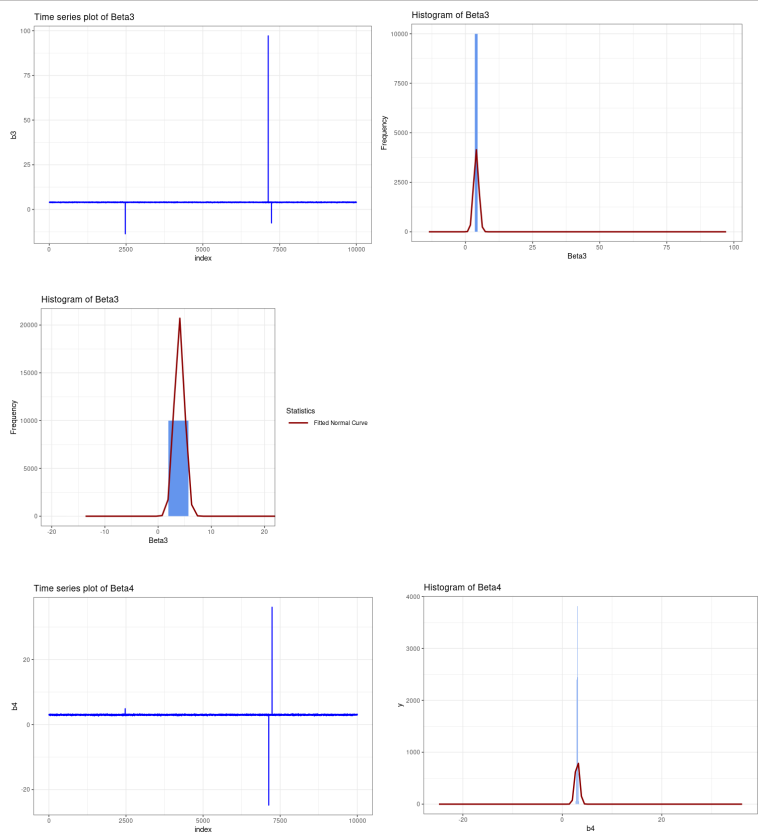
c=0.1

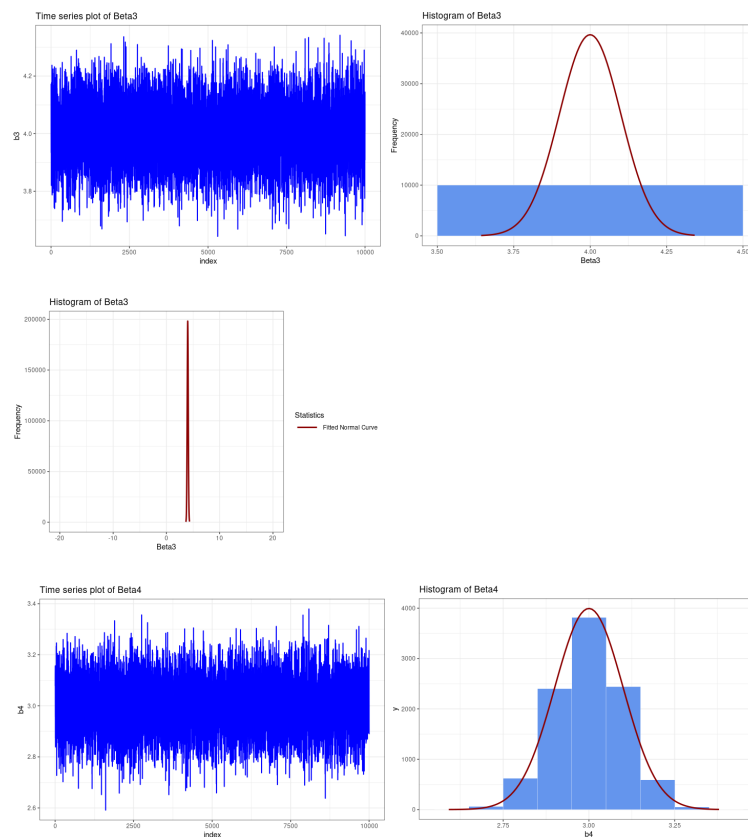
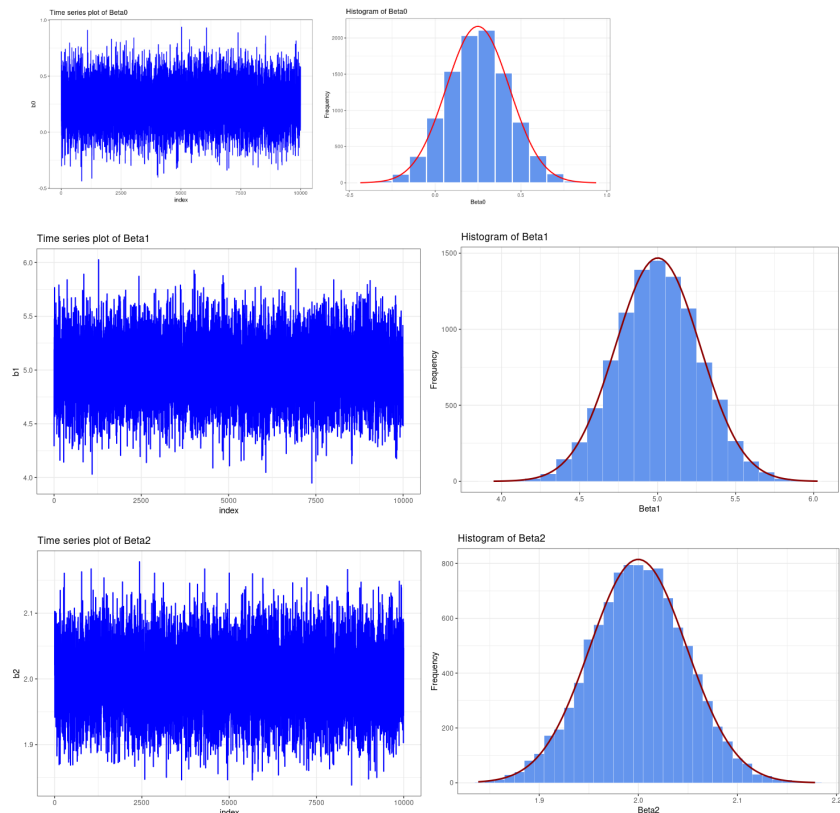


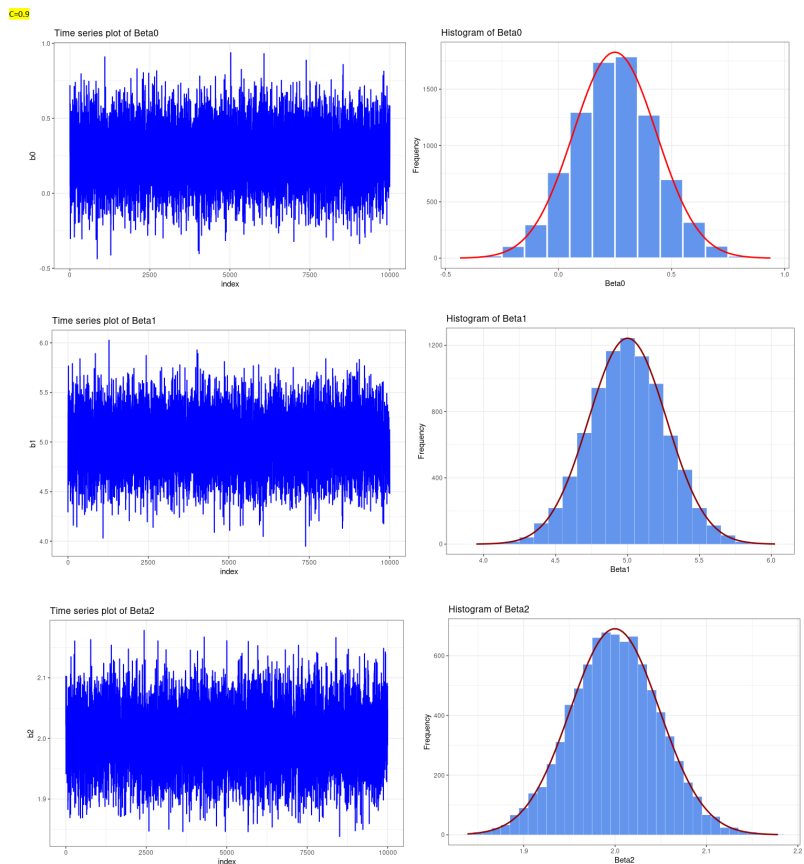
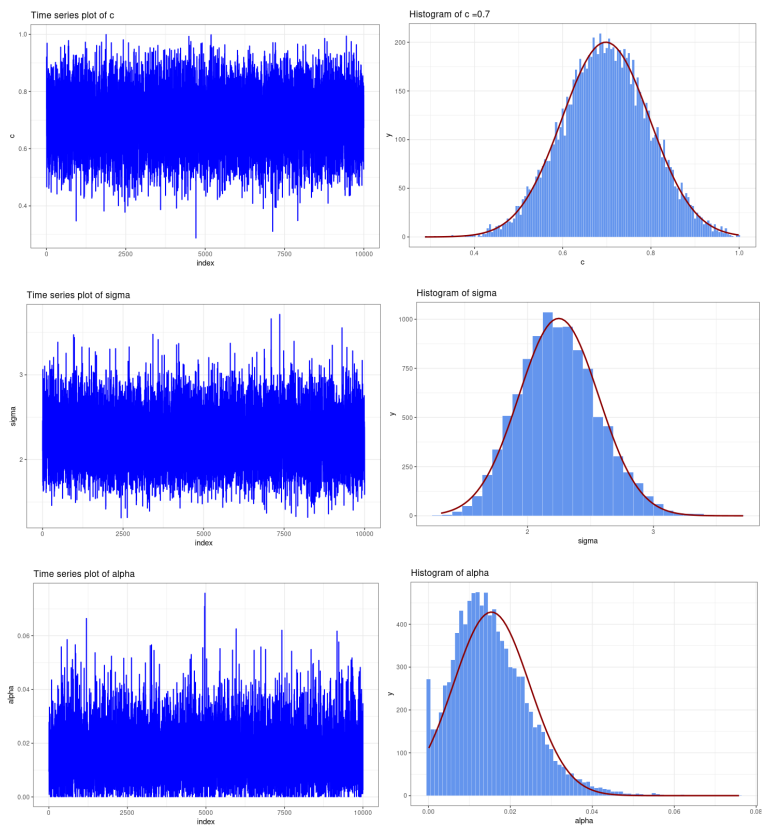








$\phi=0.7$ 



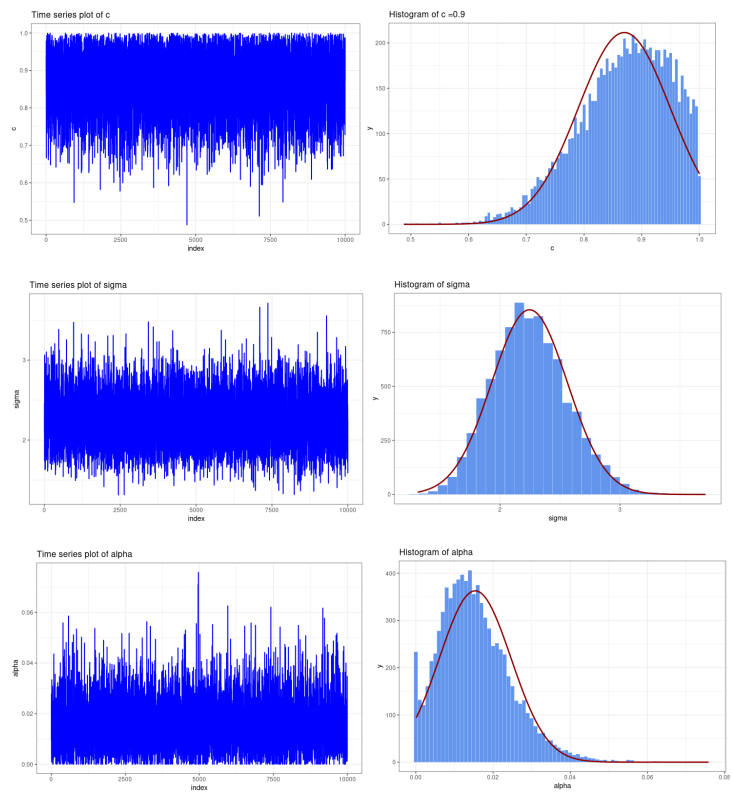
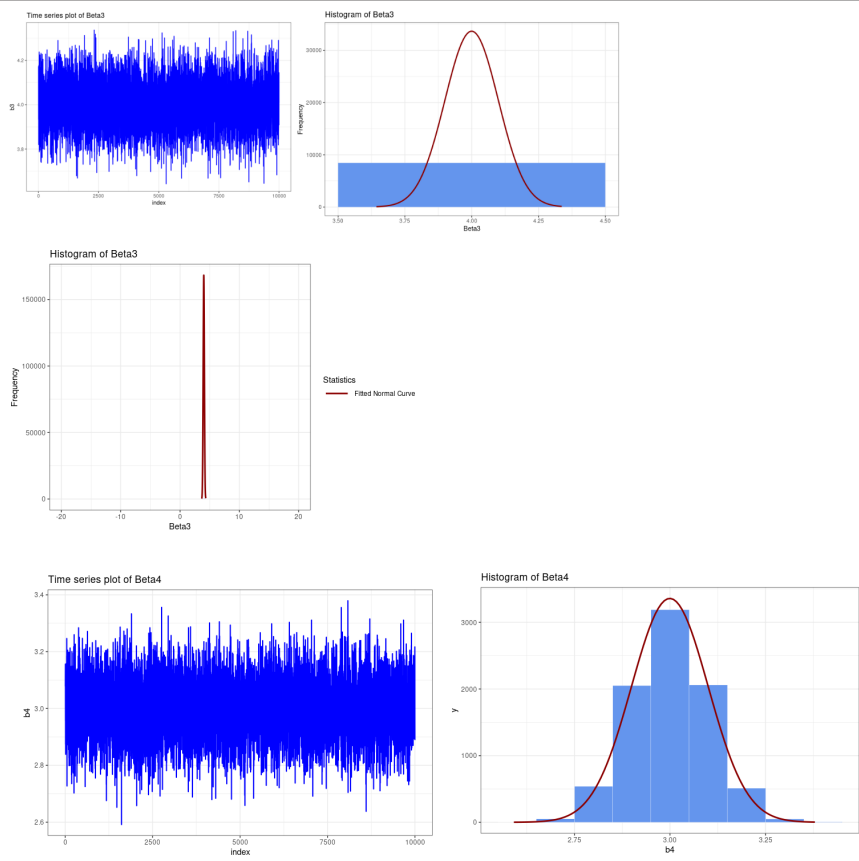


Figure 3: Trace plots of the parameters associated with the different threshold c values