Agrociencia

TIME SERIES WITH MULTIPLE CHANGE POINTS AND CENSORED OBSERVATIONS

René **Castro-Montoya**[1*], Gabriel Arcángel **Rodríguez-Yam**[2], Felipe de Jesús **Peraza-Garay**[1], José Vidal **Jiménez-Ramírez**[1]

[1] Universidad Autónoma de Sinaloa. Facultad de Ciencias Físico-Matemáticas. Avenida de las Américas y Blvd. Universitarios s/n, Ciudad Universitaria, Culiacán, Sinaloa, Mexico. C. P. 80010.
[2] Universidad Autónoma Chapingo. División de Ciencias Forestales. Carretera México-Texcoco km 38.5, Chapingo, Texcoco, State of Mexico, Mexico. C. P. 56230.
* Author for correspondence: renec@uas.edu.mx

Citation: Castro-Montoya R, Rodríguez-Yam GA, Peraza-Garay F de J, Jiménez-Ramírez JV. 2024. Time series with multiple change points and censored observations.. **Agrociencia.** doi.org/ 10.47163/agrociencia.v58i1.2856

Editor in Chief:
Dr. Fernando C. Gómez Merino

Received: August 23, 2022.
Approved: November 07, 2023.
Published in Agrociencia: January 22, 2024.

ABSTRACT

This article examines a Bayesian model for a nonstationary time series with an unknown number of change points and censored observations. Each segment is assumed to be an autoregressive process with order one. To estimate the number and locations of change points, we use the reversible jump Markov chain Monte Carlo (RJMCMC) algorithm. The censored problem is solved by imputing the censored values from a multivariate normal distribution based on the observed part. A numerical example shows that the estimates of the number of change points and their localizations have little bias. Additionally, the estimates are robust to the censoring percentage.

Keywords: Parameter estimation, Bayesian inference, prior distributions, Metropolis algorithm, reversible jump Markov chain Monte Carlo algorithm.

INTRODUCTION

A time series can present a change in the model structure and/or in some parameters in response to the effect of external factors on the variables of interest, and due to detection limits, it can also present some type of censoring on the observations. For example, the time series obtained when monitoring air pollutants, such as particles of polycyclic aromatic hydrocarbons (PAHs), carbon monoxide (CO), and sulfur dioxide (SO_2), can have censored observations and changes in the model structure. Another example is the study of cloud ceiling height, where the measurement is restricted to the detection limit of the recording device. This variable is important as it accounts for weather-related accidents and flight delays.

There are several studies and theses on change point analysis and random censorship. However, there are only a few examples in which both topics are combined. Hušková and Neuhaus (2004) proposed a hypothesis test to detect changes in the censoring time distribution using rank-based procedures. Gombay and Liu (2000) proposed a procedure for determining whether a change point occurred in randomly censored

data based on an extension of Wilcoxon's rank statistics. Gijbels and Gürler (2003) considered the problem of estimating a single change point in a piecewise constant hazard function when the observed variables are subject to random censoring.

Several authors have developed methods to analyze time series with censored observations; for example, Robinson (1980) suggested inputting the censored part by its conditional expectation given the observed part. He sub-grouped the data vector so that each subgroup included one censored observation and thus required a single integral. Park *et al*. (2007) used a random vector from the conditional multivariate normal distribution to impute censored observations based on the observed part. The method involved updating parameter estimates by imputing the censored values with the conditional sample.

Some methods for solving the change point problem have been developed under the frequentist approach. For example, Davis *et al*. (2006) modeled a nonstationary time series by dividing the time series into AR process blocks and assuming that the number, locations, and orders of the AR processes are unknown. Lavielle and Teyssière (2006) considered the multiple change-points problem for multivariate time series, which included strongly dependent processes with an unknown number of change points.

Some papers using a Bayesian approach examine change points in time series. Chib (1986) modeled the change point process as a Markov chain with constrained transition probabilities, resulting in a non-reversible sequence. Barbieri and O'hagan (1996) worked on Bayesian analysis for autoregressive time series change points. Giordani and Kohn (2008) addressed the issue of modeling and inference for processes undergoing random parameter shifts at unknown dates. They presented a simple algorithm that further improves sampling efficiency in a class of discrete latent variable models that includes change points, mixture innovation, regimen switching, and outlier detection. The adaptive algorithm uses past draws of the discrete latent variables to design a proposal distribution for a Metropolis-Hasting step, significantly reducing computing time spent on observations where the presence or absence of a break or outlier is rather clear-cut.

This paper proposes a Bayesian model for a time series with an unknown number of change points and censored observations. The RJMCMC algorithm is used to draw a sample from the model. Park *et al*. (2007) describe an algorithm for addressing the censorship problem. A simulated dataset with two change points and 0, 10, and 40 % censoring is analyzed.

MATERIALS AND METHODS

The Bayesian model proposed

Assume the realization $y_1, y_2, ..., y_n$ of a time series y_t, $t = 0, \pm 1, \pm 2, ...$, has k unknown change points at locations $\tau_1, \tau_2, ..., \tau_k$, and that some observations are right censored at c_t, $t = 1, 2, ..., n$. Thus, instead of $y_1, y_2, ..., y_n$, we observe $x_t := min\ (y_t, c_t)$.

The i^{th} segment between $\tau_i s$ is assumed to be an autoregressive (AR) process of order 1,

$$X_t = \mu_i + \phi_i(X_{t-1} - \mu_i) + \varepsilon_t, \tau_{i-1} + 1 \leq t \leq \tau_i, i = 1,2, \ldots, k+1 \quad (1)$$

where $\varepsilon_t \sim N(0, \sigma_i^2)$ are independent and identically distributed. Here, K, τ_1, τ_2, ... τ_k, $(0 < \tau_1 < \tau_2 < \cdots < \tau_k < n)$, $\mu_1, \mu_2, \cdots, \mu_{k+1}, \sigma_1^2, \sigma_2^2, \cdots, \sigma_{k+1}^2, \phi_1, \phi_2, \cdots, \phi_{k+1}$, are the parameters of the model.

The priors for the number of breaks and their locations are as follows:

$$K \sim U(0, 1, 2, \cdots, k_m)$$

$$f(\tau_i | \tau_1, \tau_2, \ldots, \tau_{i-1}, k) \sim U(\tau_{i-1} + 1, \cdots, n-1)$$

$i = 1, 2, \ldots k$

where k_m is the maximum number of change points permitted in the model. We use independent conjugate priors for the mean and variance parameters:

$$\mu_i, \sigma_i^2, i = 1, 2, \cdots, k+1$$

$$\mu_i \sim N(\mu_0, \sigma_0^2), \sigma_i^2 \sim Ig(\alpha_0, \beta_0) \quad 2$$

where μ_0 and σ_0^2 are the mean and variance in the normal density, and α_0 and β_0 are the shape and scale hyperparameters of the inverse gamma density. To ensure stationarity in each segment, the following priors are assumed for ϕ_i, $i = 1, 2, \ldots, k+1$.

$$\phi_i \sim U(-1, 1) \phi_i \sim U(-1, 1)$$

K and $\theta_k = (\tau_1, \tau_2, \cdots, \tau_k, \mu_1, \mu_2, \cdots, \mu_{k+1}, \sigma_1^2, \sigma_2^2, \cdots, \sigma_{k+1}^2, \phi_1, \phi_2, \cdots, \phi_{k+1})$ Bayesian inference is based on the distribution $f(y_c, \theta_k, k | y_o)$, where $y_o := \{y_i | y_i \leq c_i, i = 1, 2, \cdots, n\}$ and $y_c := \{c_i | y_i > c, i = 1, 2, \cdots, n\}$, which can be factorized by:

$$f(y_c, \theta_k, k | y_o) = f(y_c | y_o, \theta_k, k) f(\theta_k, k | y_o)$$

$$\propto f(y_c | y_o, \theta_k, k) f(y_o | \theta_k, k) f(\theta_k, k)$$

$$\propto f(y | \theta_k, k) f(\theta_k, k)$$

Reversible jump Markov chain Monte Carlo (RJMCMC) algorithm

Since its introduction, the reversible jump Markov chain Monte Carlo (RJMCMC) by Green (1995) has been widely recognized as encompassing many common MCMC algorithms, including standard ones. However, what is not often appreciated is that RJMCMC is merely a tightening of the terms in the Metropolis–Hastings algorithm to allow consideration of problems involving general state spaces, including those comprised of subspaces of different dimensions. This algorithm is based on creating an irreducible and aperiodic Markov chain that can alternate jumps among various models with parameter spaces of different dimensions while retaining detailed balance, ensuring convergence to correct the limiting distribution. As a result, the RJMCMC algorithm has been widely used.

Green (1995) applied this algorithm to Poisson processes with rate $x(t)$, where $x(t)$ is constrained to be a step function with an unknown number of steps. Troughton and Godsill (1997) addressed the problem of Bayesian inference in autoregressive processes where the correct model order was unknown. As noted by many authors, this type of problem is particularly appropriate due to its nested structure, which facilitates the construction across dimension jumps. Richardson and Green (1997) used the RJMCMC algorithm to carry out Bayesian inference for mixture models with an unknown number of mixture components.

Typically, the RJMCMC algorithm is used when the sampler target distributions are defined over a union of subspaces $\{C_r\}$ of different n_r dimensions. It requires the design of different types of moves between the subspaces. These will be combined to form what Tierney (1994) calls a hybrid sampler by randomly choosing between available moves at each transition to traverse freely across the combined parameter space C. The algorithm achieves model space moves by employing a proposal distribution and acceptance probability designed to preserve detailed balance and, hence, ensuring convergence to the correct invariant distribution. Let M_0, M_1, ..., M_m be a collection of candidate models, and assume that M_r is a model with a vector of unknown parameters of size n_r, which may vary from model to model. Assume we wish to move from the current model M_i to the new model M_j.

A general version of the RJMCMC algorithm is created by proposing a new model M_j with $r_{ij}(\theta_i)$ probability, moving model M_i to M_j, and generating a vector of random variables $\boldsymbol{u}$ from a specified proposal density $q(u)$. A reversible move between models M_i and M_j is established by a bijective function g that transforms the parameters $g: (\theta_i, \boldsymbol{u}) \rightarrow (\theta_j, \boldsymbol{u'})$ and retains the dimensions $dim(\theta_i) + dim(\boldsymbol{u}) = dim(\theta_j) + dim(\boldsymbol{u'})$. The function g^{-1} gives the move to the other direction, where the vectors u and u' are used to make the dimensions of the parameter spaces of M_i and M_j equal. Note that the function g and its inverse must be differentiable. The proposed move to model M_j is accepted with probability

$$\alpha = min\left(1, \frac{f(M_j, \theta_j|y) r_{ji}(\theta_j) q'(\boldsymbol{u}')}{f(M_i, \theta_i|y) r_{ij}(\theta_i) q(\boldsymbol{u})} \left|\frac{\partial g(\theta_j, \boldsymbol{u})}{\partial(\theta_i, \boldsymbol{u})}\right|\right)$$

where $q(u')$ is the proposal density independent of θ_j (dimension m_j) and $|\cdot|$ is the Jacobian determinant. If the move is not accepted, stay in model M_i with parameter vector θ_i.

In the inverse direction, the move from M_j to M_i has an acceptance probability α' and is given by:

$$\alpha' = min\left(1, \frac{f(M_i, \theta_i|y) r_{ij}(\theta_i) q(\boldsymbol{u})}{f(M_j, \theta_j|y) r_{ji}(\theta_j) q'(\boldsymbol{u}')} \left|\frac{\partial g^{-1}(\theta_i, \boldsymbol{u})}{\partial(\theta_j, \boldsymbol{u}')}\right|\right)$$

The simulation of reversible jumps provides a sample of values among a set of statistical models based on the probability of each of these models; for example, the Markov chain generated by a reversible jump algorithm will move between the different models that are considered to describe the data, allowing the chain to stay longer in those models that better describe them. As noted by Green (1995), the RJMCMC algorithm usually consists of different types of moves that change the state of the Markov chain from one model to another.

The Markov chain has movements that occur across each model visited by the simulation process. These movements are known as within-model moves and are carried out by the Gibbs sampler and the Metropolis algorithm. The other type of move involves a change in the dimension of the parameter spaces and is known as between-models moves. To achieve this, a vector of random variables u is generated from a specified proposal density $g(u)$. The proposed new state is then given by $g(\theta_i, u) = (\theta_j, u')$, where g and its inverse are differentiable, and finally, the acceptance probability is calculated.

In the implementation of the RJMCMC algorithm, several difficulties are often encountered, particularly in making the chain jump from one model to another. There are two choices to be made when constructing moves between models of different dimensions: the proposal density and the bijective function g. To design more efficient problems, it has become standard practice to tune proposals. Tuning is the process of performing several short runs of an RJMCMC algorithm, each time changing certain aspects of a proposal. The specification associated with the run that maximizes acceptance rates may then be chosen as the one to be used for the main RJMCMC analysis.

There are several suggestions on how to select a proposal density; for example, Brooks *et al*. (2003) introduced a framework for selecting an efficient proposal density q for the implementation of the RJMCMC algorithm and then building a bijective function g.

Estimation of the number and locations of change points

To implement the RJMCMC algorithm for our change-points problem and censored observations, four move types are proposed: move Type 1: generate a new change point; move Type 2: eliminate one change point; move Type 3: update the autoregressive parameters; and move Type 4: update the censored observations. In each iteration, one move is chosen randomly. Notice that move types 1 and 2 involve a change in dimension. A Gibbs sampler is used in move Type 3, and the Park *et al*. (2007) method is used in move Type 4.

Move type 1: Generating a new change point

When a new change point occurs, the model transitions from k to $k+1$. Following the procedure proposed by Green (1995), a segment is randomly chosen, followed by a change point within that segment.

The proposal is to generate u from the standard normal truncated in the interval (-1, 1).

$$u \sim N_{(-1,1)}(0,1)$$

where a reversible move between a model with k change points and a model with $k+1$ change points is required. This is accomplished by the bijective function g that follows:

$$g\left(\tau_{i_s} - \tau_{i_s-1}, \mu_{i_s}, \sigma_{i_s}^2, u_1, u_2, u_3\right) = \left(\tau'_{i_s} - \tau_{i_s-1}, \tau_i - \tau'_{i_s}, \mu_{i_s,1}, \mu_{i_s,2}, \sigma_{i_s,1}^2, \sigma_{i_s,2}^2\right)$$

$$(\tau'_{i_s} - \tau_{i_s-1}) = (\tau_i - \tau_{i_s-1})u_1$$

$$(\tau_{i_s} - \tau'_{i_s}) = (\tau_{i_s} - \tau_{i_s-1})(1-u_1)$$

$$\mu_{i_s,1} = \mu_{i_s} - u_2\,\sigma_i\sqrt{\frac{(\tau_{i_s} - \tau'_{i_s})}{(\tau_i - \tau_{i_s-1})}}$$

$$\mu_{i_s,2} = \mu_{i_s} - u_2\sigma_i\sqrt{\frac{(\tau_{i_s} - \tau_{i_s-1})}{(\tau_i - \tau'_{i_s})}}$$

$$\sigma_{i_s,1}^2 = u_3(1-u_2^2)\sigma_i^2\,\frac{(\tau_{i_s} - \tau_{i_s-1})}{(\tau_{i_s} - \tau_{i_s-1})}$$

$$\sigma_{i_s,2}^2 = (1-u_3)(1-u_2^2)\sigma_i^2\,\frac{(\tau_{i_s} - \tau_{i_s-1})}{(\tau_{i_s} - \tau'_{i_s})}$$

.

such that $dim\ (u) + dim\ (\theta_i) = \dim\ (u') + dim\ (\theta_j)$. The function g was used by Richardson and Green (1997) to solve the distribution mixture problem. The function g^{-1} gives the move to the other direction.

The probability of acceptance (from k to $k + 1$) is given by:

$$\alpha = min\left(1, \frac{f(M_j, \theta_j|y) r_{ji}(\theta_j) q'(\boldsymbol{u}')}{f(M_i, \theta_i|y) r_{ij}(\theta_i) q(\boldsymbol{u})} \left|\frac{\partial g(\theta_j, \boldsymbol{u})}{\partial(\theta_i, \boldsymbol{u})}\right|\right)$$

Move type 2: Eliminate a change point

We consider a transition from a model with k change points to another model with k - 1 change points. To accomplish this, a random change point is selected, and the inverse transform g^{-1} is applied in the previous step. The probability of accepting this move is:

$$\alpha' = min\left(1, \frac{\mathrm{f(M_i, \theta_i|y) r_{ij}(\theta_i) q(\mathbf{u})}}{\mathrm{f(M_j, \theta_j|y) r_{ji}(\theta_j) q'(\mathbf{u}')}} \left|\frac{\partial \mathrm{g^{-1}(\theta_i, \mathbf{u})}}{\partial(\mathrm{\theta_j, \mathbf{u}'})}\right|\right)$$

Move type 3: Update the autoregressive parameters using the Gibbs sampler

The vector parameters of the autoregressive model are updated sequentially, and these updates are obtained by μ, τ, and ϕ in the i_s segment with the ends at the change points in $\tau_{i_{s-1}}$ and τ_{i_s} for each autoregressive segment i = 1, 2, ..., k + 1, using the Gibbs sampler algorithm. The conditional densities of μ_i, ϕ_i, and σ_i^2 are obtained.

Conditional density of μ is explained by:

$$f(\mu|\tau, \phi, \sigma^2, y_{1,} y_{2,\ldots,} y_n) = \frac{f(\mu, \tau, \phi, \sigma^2, y_c|y_0)}{f(\tau, \phi, \sigma^2, y_c|y_0)}$$

$$\propto f(\mu, \tau, \phi, \sigma^2, y_{c,}|y_0)$$

$$\propto f(y_{c,} y_0|\mu, \tau, \phi, \sigma^2) f(\mu, \tau, \phi, \sigma^2)$$

$$= \prod_{i=1}^{k+1} \prod_{j=\tau_{i-1}+1}^{\tau_i} f(y_j|\mu, \tau, \phi, \sigma^2, y_{j-1}) \prod_{i=1}^{k+1} f(\mu_i|\mu_0, \sigma_0^2)$$

Then,

$$f\left(\mu_{i_s}|\tau_{i_s}, \phi_{i_s}, \sigma_{i_s}^2, y_{\tau_{i_{s-1}+1}}, \ldots, y_{\tau_{i_s}}\right) = \prod_{j=\tau_{i_{s-1}+1}}^{\tau_{i_s}} f(y_j|\mu_{i_s}, \tau_{i_s}, \phi_{i_s}, \sigma_{i_s}^2, y_{j-1}) f(\mu_{i_s}|\mu_0, \sigma_0^2)$$

by (2)

$$= exp\left\{\frac{-1}{2\sigma_{i_s}^2}\sum_{j=\tau_{i_{s-1}+1}}^{\tau_{i_s}}\left(y_j-\mu_{i_s}-\phi_{i_s}(y_{j-1}-\mu_{i_s})\right)^2\right\}exp\left\{\frac{-1}{2\sigma_0^2}(\mu_{i_s}-\mu_0)^2\right\}$$

$$= exp\left\{\frac{-1}{2\sigma_{i_s}^2}\sum_{j=\tau_{i_{s-1}+1}}^{\tau_{i_s}}\left(y_j-\phi_{i_s}y_{j-1}-\mu_{i_s}(1-\phi_{i_s})\right)^2-\frac{1}{2\sigma_0^2}(\mu_{i_s}-\mu_0)^2\right\}$$

Then

$$\left(\mu_{i_s}|\tau_{i_s},\phi_{i_s},\sigma_{i_s}^2,y_{\tau_{i_{s-1}+1}1},\dots,y_{\tau_{i_s}}\right)\sim N\left(\frac{A(1-\phi_{i_s})^2\sigma_0^2+\mu_0\sigma_{i_s}^2}{A\bar{v}_{i_s}(1-\phi_{i_s})^2\sigma_0^2+\sigma_{i_s}^2},\frac{\sigma_0^2\sigma_{i_s}^2}{A(1-\phi_{i_s})^2\sigma_0^2+\sigma_{i_s}^2}\right)$$

where $\bar{v}_{i_s}=\frac{1}{n}\sum_{j=\tau_{i_{s-1}}+1}^{\tau_{i_s}}\left(y_j-\phi_{i_s}y_{j-1}\right)$, and $A=\left(\tau_{i_s}-\tau_{i_{s-1}}\right)$

Conditional density of σ^2 is explained by:

$$f(\sigma^2|\mu,\tau,\phi,y_0)=\frac{f(\mu,\tau,\phi,\sigma^2,y_c|y_0)}{f(\tau,\phi,\mu,y_c|y_0)}$$

$$\propto f(\mu,\tau,\phi,\sigma^2,y_c|y_0)$$

$$\propto f\left(y_{c,}y_0|\mu,\tau,\phi,\sigma^2\right)f(\mu,\tau,\phi,\sigma^2)$$

$$=\prod_{i=1}^{k+1}\prod_{j=\tau_{i-1}+1}^{\tau_i}f\left(y_j|\mu,\tau,\phi,\sigma^2,y_{j-1}\right)\prod_{i=1}^{k+1}f(\sigma_i^2|\alpha,\beta)$$

Then by (2),

$$f\left(\sigma_{i_s}^2|\mu_{i_s},\tau_{i_{s-1}}\phi_{i_s},y_{\tau_{i_{s-1}}+1},\ldots,y_{\tau_{i_s}}\right)=\prod_{j=\tau_{i_{s-1}}+1}^{\tau_{i_s}} f\left(y_j|\mu,\tau,\phi,\sigma^2,y_{j-1}\right)Ig\left(\sigma_{i_s}^2|\alpha,\beta\right)$$

$$=\sigma_{i_s}{}^{(\tau_{i_s}-\tau_{i_s-1})}exp\left\{\frac{-1}{2\sigma_{i_s}^2}\sum_{j=\tau_{i_{s-1}}+1}^{\tau_{i_s}}\left(y_j-\mu_{i_s}-\phi_{i_s}\left(y_{j-1}-\mu_{i_s}\right)\right)^2\right\}\sigma_{i_s}^{-2(\alpha-1)}exp\left\{\frac{-\beta}{\sigma_{i_s}^2}\right\}$$

Thus

$$\left(\sigma_{i_s}^2|\mu_{i_s},\tau_{i_{s-1}}\phi_{i_s},y_{\tau_{i_{s-1}}+1},\ldots,y_{\tau_{i_s}}\right)\sim Ig\left(0.5A+\alpha,0.5\sum_{j=\tau_{i_{s-1}}+1}^{\tau_{i_s}}\left(y_j-\mu_{i_s}-\phi_{i_s}B^2\right)+\beta\right)$$

where

$$A=(\tau_{i_s}-\tau_{i_s-1})$$

$$B=\left(y_{j-1}-\mu_{i_s}\right)$$

Finally, conditional density of ϕ is explained by:

$$f\left(\phi|\mu,\tau,\sigma^2,y_{1,}y_{2,\ldots,}y_n\right)=\frac{f\left(\mu,\tau,\phi,\sigma^2,y_{c,}|y_0\right)}{f\left(\mu,\phi,\sigma^2,y_{c,}|y_0\right)}$$

$$\propto f\left(\mu,\tau,\phi,\sigma^2,y_{c,}|y_0\right)$$

$$\propto f\left(y_{c,}y_0|\mu,\tau,\phi,\sigma^2\right)f\left(\mu,\tau,\phi,\sigma^2\right)$$

$$=\prod_{i=1}^{k+1}\prod_{j=\tau_{i-1}+1}^{\tau_i} f\left(y_j|\mu,\tau,\phi,\sigma^2,y_{j-1}\right)\prod_{i=1}^{k+1} f(\phi)$$

Since we previously assumed that $\phi_i \sim U$ (-1, 1),

$$f\left(\phi_{i_s}|\mu_{i_s},\tau_{i_s},\sigma_{i_s}^2,y_{\tau_{i_s-1}+1},\ldots,y_{\tau_{i_s}},k\right)$$
$$\propto \sigma_{i_s}^{-(\tau_{i_s}-\tau_{i_s-1})} exp\left\{\frac{-1}{2\sigma_{i_s}^2}\sum_{j=\tau_{i_s-1}+1}^{\tau_{i_s}}\left(y_j-\mu_{i_s}-\phi_{i_s}(y_{j-1}-\mu_{i_s})\right)^2\right\}\frac{1}{2}$$
$$\propto exp\left\{\frac{-1}{2\sigma_{i_s}^2}\sum_{j=\tau_{i_s-1}+1}^{\tau_{i_s}}(y_{j-1}-\mu_{i_s})\left(\phi_{i_s}-\frac{(y_j-\mu_{i_s})}{(y_{j-1}-\mu_{i_s})}\right)^2\right\}$$

Finally,

$$\left(\phi_{i_s}|\mu_{i_s},\tau_{i_s},\sigma_{i_s}^2,y_{\tau_{i_s-1}+1},\ldots,y_{\tau_{i_s}},k\right)\sim N\left(\frac{\sum_{j=\tau_{i_s-1}+1}^{\tau_{i_s}}(y_{j-1}-\mu_{i_s})(y_j-\mu_{i_s})}{\sum_{j=\tau_{i_s-1}+1}^{\tau_{i_s}}(y_{j-1}-\mu_{i_s})},\frac{\sigma_{i_s}^2}{\sum_{j=\tau_{i_s-1}+1}^{\tau_{i_s}}(y_{j-1}-\mu_{i_s})}\right)$$

Move type 4: Data augmentation

This is solved by using the method proposed by Park *et al*. (2007), which consists of imputing the censored values with a random vector from a multivariate normal distribution given the observed part.

Consider the time series $y_1, y_2, \ldots, y_n$, where y_i is observed or censored to the right and $y_o=\{y_i|y_i \leq c_i, i = 1,2,\cdots,n\}$, $y_c=\{c_i|y_i > c_i, i = 1,2,\cdots,n\}$. The basic idea is to replace X_c by sampling values from the conditional distribution $f(Y_c|Y_o,\theta_k,k)$ of Y_c given by Anderson (2003):

$$y_c|y_o,\theta_k,k\sim N_T(\mu_c^o+\Sigma_{co}^o(\Sigma_{oo}^o)^{-1}(y_o-\mu_o^o),\Sigma_{cc}^o-\Sigma_{co}^o(\Sigma_{oo}^o)^{-1}\Sigma_{oc}^o)$$

where $\Sigma_{oo}^o = P_o\ \Sigma\ P_o^t, \Sigma_{oc}^o = P_o\ \Sigma\ P_c^t, \Sigma_{co}^o = P_c\ \Sigma\ P_o^t$, and $\Sigma_{cc}^o = P_c\ \Sigma\ P_c^t$ (3)

Here, P is selected such that $PY=(Y_c,Y_o)^t$, P_o, and P_c are defined by the permutation $P=(P_o,P_c)^t$

$$P_{o,kj}=\begin{cases}1, y_j = y_o\ k = 1,2,\ldots,n.\\ 0, y_j \neq y_o\ j = 1,2,\ldots,n.\end{cases}$$

$$P_{c,ij}=\begin{cases}1, y_j = y_c\ i = 1,2,\ldots,n_c.\\ 0, y_j \neq y_c\ j = 1,2,\ldots,n.\end{cases}$$

The vectors X_c and X_o are obtained as follows:

$$\boldsymbol{X_o} = \begin{cases} y_k, y_k \leq c_k \; k = 1,2, \ldots, n. \\ 0, y_k > c_t \; k = 1,2, \ldots, n. \end{cases}$$

$$\boldsymbol{X_c} = \begin{cases} c_t, y_i > c_i \; i = 1,2, \ldots, n. \\ 0, y_i \leq c_i \; i = 1,2, \ldots, n. \end{cases} \quad (4)$$

Applying the permutation matrix P to the observed data $\boldsymbol{X} = (X_1, \cdots, X_n)^t$, we obtain $PX = (\boldsymbol{X_c}, \boldsymbol{X_o})^t =^d (\boldsymbol{X_c}, \boldsymbol{Y_o})^t$.

The Park *et al*. (2007) algorithm, used to simulate censored observations, is built on the permutation matrices P_o and P_c and the vectors X_o and X_c by means of the expressions given in (3) and (4), respectively. The initial values for $\hat{\mu}_0$, $\hat{\Phi}_0$, and $\hat{\sigma}_0^2$ are obtained to calculate the mean and the covariance matrix:

$$\mu^{(0)} = \hat{\mu}_0 \mathbf{1_n} \text{ and } \Sigma_{ij}^{(0)} = \frac{\hat{\sigma}_0^2}{\left(1 - \hat{\Phi}_0^{\,2}\right)} \hat{\Phi}_0^{\,|i-j|}, i, j = 1,2, \ldots, n$$

where Σ_{oo}^o, Σ_{oc}^o, Σ_{co}^o, and Σ_{cc}^o are obtained by using equation (3).

To simulate the censored values:

$$\boldsymbol{y_c} \sim N_T(\mu_c^{(o)} + \Sigma_{co}^{(o)}(\Sigma_{oo}^{(o)})^{-1}(\boldsymbol{y}_o - \mu_o^{(o)}), \Sigma_{cc}^{(o)} - \Sigma_{co}^{(o)}(\Sigma_{oo}^{(o)})^{-1}\Sigma_{oc}^{(o)})$$

To obtain the vector of complete observations:

$\boldsymbol{y} = P^{-1}(\boldsymbol{y_c}, \boldsymbol{y_o})^t$, where P=$(P_o, P_c)^t$

Finally, μ, ϕ, and σ^2 are estimated.

Convergence

To monitor the convergence of a reversible jump Markov chain Monte Carlo, multiple chains of the same length but with different starting points are generated and the moment when the chains forget the starting points is checked. There are several methods for comparing parallel chains.

The Castelloe (1999) method to monitor convergence in a Markov chain Monte Carlo with reversible jumps is suitable in situations where the different parameter spaces of the models are indexed by some parameter in the chain. It entails monitoring various sources of variation, including variations in the chains, interactions between model

chains, and significant differences in the frequencies of model visits from one chain to another. Any of these conditions may indicate that the chains no longer come from the same stationary distribution and that they have not reached convergence to the stationary distribution.

Let $\theta_1, \theta_2, \ldots \theta_p$, be the parameters, assuming that there are C parallel Markov chains, each with a length of qb, with $\theta_i^{(j)}, i = 1, 2, \ldots, C, j = qb + 1, qb + 2, \ldots 2qb$, for some q and a base size of b. The notation used for the implementation of the convergence assessment is:

$$R_{\cdot m} = \sum_{c=1}^{C} R_{cm}$$

$$\bar{\theta}_{cm}^{\cdot} = \frac{1}{R_{cm}} \sum_{r=1}^{R_{cm}} \theta_{cm}^{r}$$

$$\bar{\theta}_{c\cdot}^{\cdot} = \frac{1}{T} \sum_{r=1}^{R_{cm}} \theta_{cm}^{r}$$

$$\bar{\theta}_{\cdot m}^{\cdot} = \frac{1}{R_{\cdot m}} \sum_{c=1}^{C} \sum_{r=1}^{R_{cm}} \theta_{cm}^{r}$$

$$\bar{\theta}_{\cdot\cdot}^{\cdot} = \frac{1}{CT} \sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} \theta_{cm}^{r}$$

where θ refers to the parameter vector with the same interpretation in the models; C is the number of chains; τ is the chain size; M is the number of different models visited by the chain; θ_{cm}^{r} is the value of θ for the rth occurrence of the model m in the chain C; and R_{cm} is the number of times that the model m appears in the chain C.

The Castelloe (1999) method to monitor convergence in RJMCMC is based on the estimates of the following variation sources:

$$V = \frac{1}{CT - 1} \sum_{cm=1}^{C} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} (\theta_{cm}^{r} - \bar{\theta}_{\cdot\cdot}^{\cdot})^2,$$

$$W_c = \frac{1}{C(T - 1)} \sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} (\theta_{cm}^{r} - \bar{\theta}_{c\cdot}^{\cdot})^2,$$

$$W_m = \frac{1}{C(T-M)} \sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} (\theta_{cm}^{r} - \bar{\theta}_{\cdot m}^{\cdot})^2,$$

$$W_m W_c = \frac{1}{C(T-M)} \sum_{c=1}^{C} \sum_{m=1}^{M} \sum_{r=1}^{R_{cm}} (\theta_{cm}^{r} - \bar{\theta}_{cm}^{\cdot})^2,$$

where V is the total variation, W_c is the variation within the chains, W_m is the variation within the models, and $W_m W_c$ is the variation between the chains-models. The statistics for the convergence assessment are: $R_1 = \frac{V}{W_c}$ and $R_2 = \frac{W_m}{W_m W_c}$. If the chains converge, then the estimates of V and W_c are similar, as are the estimations of W_m and $W_m W_c$. Values higher than 1 for R_1 and R_2 indicate convergence.

RESULTS AND DISCUSSION

In this section, the model is exemplified using a simulated dataset with 600 observations, change points at $\tau_1 = 200$ and $\tau_1 = 400$ and the parameters of the three AR(1) segments $\mu = (12, 12, 12)$, $\sigma^2 = (3, 1.5, 3)$ and $\phi = (0.5, 0.79, 0.5)$. The process can be written as:

$$Y_t = \begin{cases} 12 + 0.50(Y_{t-1} - 12) + \varepsilon_t, 1 \leq t \leq 200 \\ 12 + 0.79(Y_{t-1} - 12) + \varepsilon_t, 201 \leq t \leq 400 \\ 12 - 0.50(Y_{t-1} - 12) + \varepsilon_t, 401 \leq t \leq 600 \end{cases}$$

Three levels of censoring were considered: 10, 40, and 0 % (Figure 1). The RJMCMC algorithm generated two chains of 50 000 iterations each. The first 10 000 iterations were removed, which correspond to the burning period, and the remaining were used to obtain the Bayesian estimates for the model parameters. With 50 000 iterations, the statistics R1 and R2 were always greater than one. According to Castelloe (1999), the chains converged to their statio.

Regarding the values of the parameters and their estimates (Table 1), the estimated point values and their standard deviations were obtained by using the mean of the subsequent distribution. The estimates are based on the path of the RJMCMC for which $k = 2$. The estimations of the number of change points and their localizations have little bias.

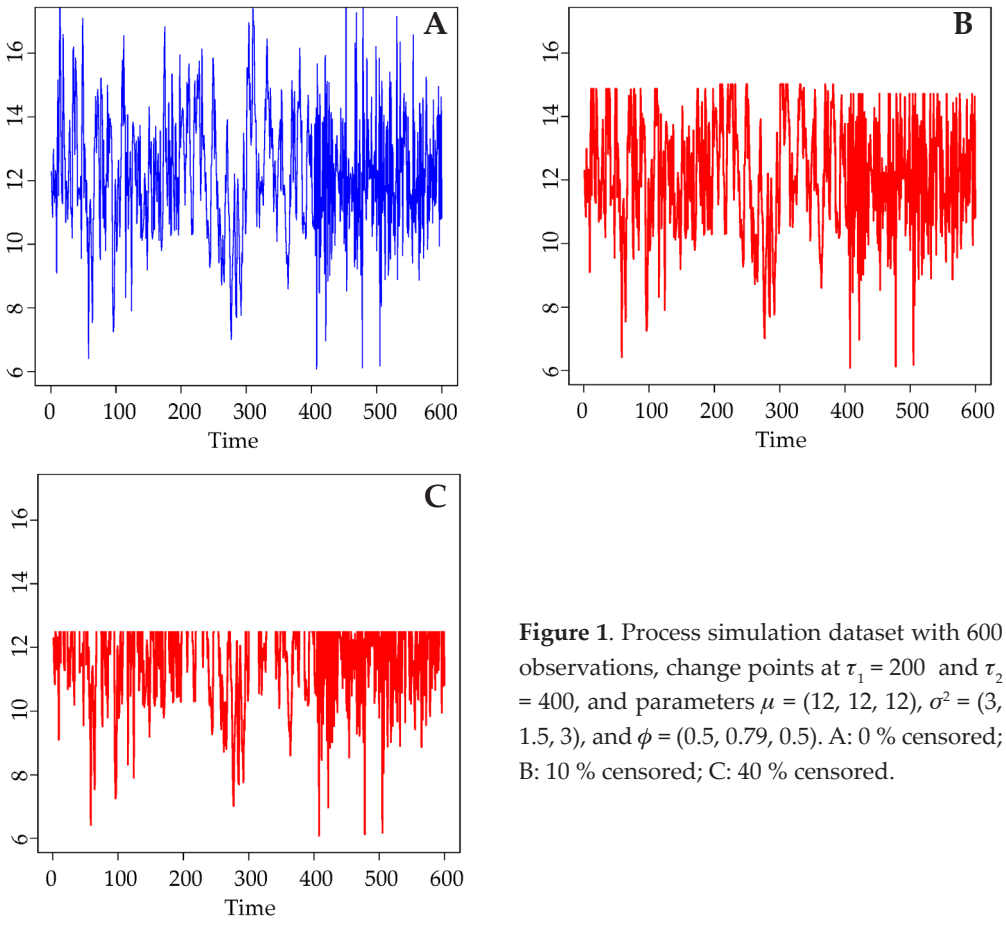

Figure 1. Process simulation dataset with 600 observations, change points at $\tau_1 = 200$ and $\tau_2 = 400$, and parameters $\mu = (12, 12, 12)$, $\sigma^2 = (3, 1.5, 3)$, and $\phi = (0.5, 0.79, 0.5)$. A: 0 % censored; B: 10 % censored; C: 40 % censored.

Table 1. Simulation parameter and estimation results.

Parameters	True Value	Censoring		
		10 % Estimated ± SD	40 % Estimated ± SD	0 % Estimated ± SD
τ_1	200	203 ± 37.80	199 ± 41.38	209 ± 41.24
τ_1	400	400 ± 15.09	398 ± 12.47	391 ± 13.01
μ_1	12	12.19 ± 0.17	11.69 ± 0.12	12.28 ± 0.21
μ_2	12	12.18 ± 0.37	11.76 ± 0.27	12.22 ± 0.41
μ_3	12	11.98 ± 0.06	11.47 ± 0.06	12.10 ± 0.10
σ_1	1.73	1.36 ± 0.16	1.02 ± 0.09	1.58 ± 0.18
σ_2	1.22	1.01 ± 0.15	0.67 ± 0.1	1.21 ± 0.20
σ_3	1.73	1.37 ± 0.18	1.20 ± 0.12	1.60 ± 0.21
Φ_1	0.50	0.41 ± 0.06	0.34 ± 0.079	0.44 ± 0.06
Φ_2	0.79	0.77 ± 0.12	0.76 ± 0.137	0.77 ± 0.15
Φ_3	-0.50	-0.50 ± 0.05	-0.35 ± 0.059	-0.52 ± 0.05

The histograms obtained from samples from the posterior distribution for the number of change points (Figure 2) show the value with highest probability is k=2. Thus, the estimated number of change points is 2.

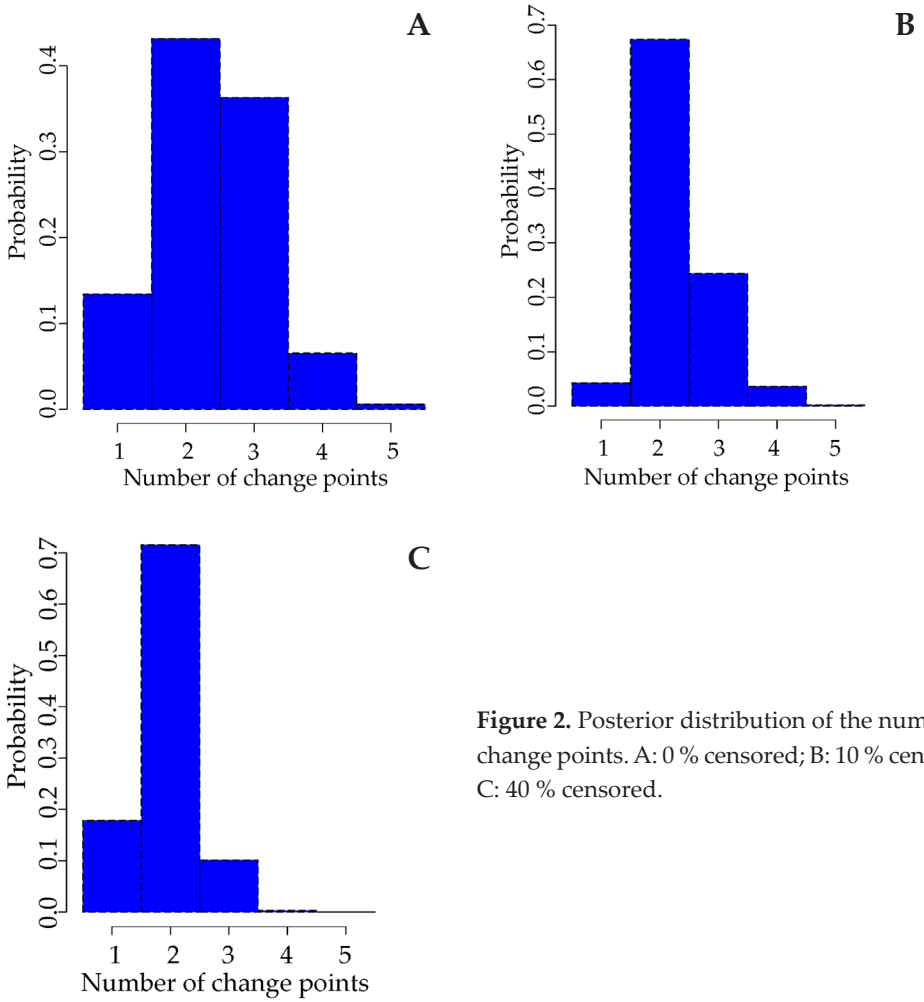

Figure 2. Posterior distribution of the number of change points. A: 0 % censored; B: 10 % censored; C: 40 % censored.

The histograms obtained from samples from the posterior distribution for the locations of the change points (Figure 3) are shown (conditioned on the value k=2).

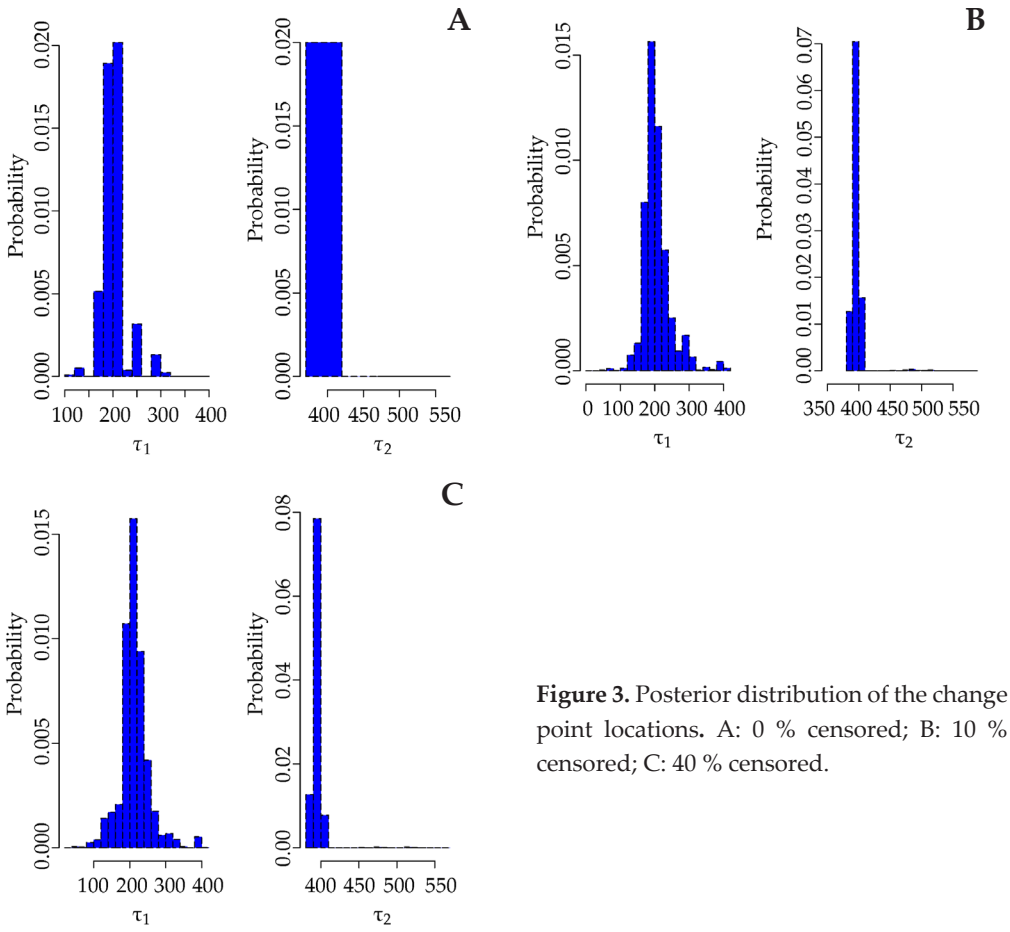

Figure 3. Posterior distribution of the change point locations**.** A: 0 % censored; B: 10 % censored; C: 40 % censored.

CONCLUSIONS

This paper proposes a Bayesian model for a time series with structural or parameter changes that may include censoring in its observations. The number of change points and their positions were assumed to be unknown. In each segment, autoregressive processes of order one were assumed. Except for the autoregressive coefficients, initial conjugated distributions were used to calculate the means and variances of each segment. Classic Markov chain Monte Carlo methods cannot be directly used to analyze the proposed Bayesian model. The reversible jump Markov chain Monte Carlo algorithm is a generalization of the MCMC techniques in which the parametric space can change between iterations. It is used to obtain a sample of the final distribution of change points and their localizations. The numerical example shows that the estimates of the number of change points and their localizations have little bias. Additionally, the estimates are robust to the percentage of censoring.

REFERENCES

Anderson TW. 2003. An introduction to multivariate statistical analysis (Third edition). John Wiley and Sons: Hoboken, NJ, USA. 457 p.

Barbieri MM, O'Hagan A. 1996. A reversible jump MCMC sampler for Bayesian analysis of ARMA time series. Technical report. Dipartimento di Statistica, Probabilità e Statistica Applicata, Università La Sapienza: Rome, Italy.

Brooks SP, Giudic P, Roberts GO. 2003. Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions. Journal of the Royal Statistical Society. Series B Statistical Methodology 65 (1): 3–39. https://doi.org/10.1111/1467-9868.03711

Castelloe J. 1999. Reversible jump Markov Chain Monte Carlo analysis of spatial Poisson cluster processes with bivariate normal displacement. Computing Science and Statistics 31: 306–315.

Chib S. 1986. Estimation and comparison of multiple change-point models. Journal of Econometrics 86 (2): 221–241. https://doi.org/10.1016/S0304-4076(97)00115-2

Davis RA, Lee TCM, Rodriguez-Yam GA. 2006. Structural breaks estimation for non-stationary time series models. Journal of the American Statistical Association 101 (473): 223–239. https://doi.org/10.1198/016214505000000745

Gijbels I, Gürler Ü. 2003. Estimation of a change point in a hazard function based on censored data. Lifetime Data Analysis 9 (4): 395–411. https://doi.org/10.1023/B:LIDA.0000012424.71723.9d

Giordani P, Khon R. 2008. Efficient Bayesian inference for multiple change-point and mixture innovation models. Journal of Business and Economic Statistics 26 (1): 66–77. https://doi.org/10.1198/073500107000000241

Gombay E, Liu S. 2000. A nonparametric nest for change in randomly censored data. The Canadian Journal of Statistics 28 (1): 113–121. https://doi.org/10.2307/3315885

Green P. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82 (4): 711–732. https://doi.org/10.2307/2337340

Hušková M, Neuhaus G. 2004. Change point analysis for censored data. Journal of Statistical Planning and Inference 126 (1): 207–223. https://doi.org/10.1016/j.jspi.2003.07.018

Lavielle M, Teyssière G. 2006. Detection of multiple change points in multivariate time series. Lithuanian Mathematical Journal 46 (3): 287–306. https://doi.org/10.1007/s10986-006-0028-9

Park JW, Genton MG, Ghosh SK. 2007. Censored time series analysis with autoregressive moving average models. The Canadian Journal of Statistics 35 (1): 151–168. https://doi.org/10.1002/cjs.5550350113

Richardson S, Green PJ. 1997. On Bayesian analysis of mixtures with an unknown number of components (with discussion). Journal of the Royal Statistical Society: Series B Statistical Methodology 59 (4): 731–792. https://doi.org/10.1111/1467-9868.00095

Robinson PM. 1980. Estimation and forecasting for time series containing censored or missing observations. *In* Anderson OD. (ed.), Time Series: Proceedings of the International Conference Held at Nottingham University, March 1979. North-Holland: Amsterdam, Netherlands.

Tierney L. 1994. Markov Chains for exploring posterior distributions. The Annals of Statistics 22 (4): 1701–1728. https://doi.org/10.1214/aos/1176325750.

Troughton PT, Godsill SJ. 1997. A reversible jump sampler for autoregressive time series. *In* Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181). Institute of Electrical and Electronics Engineers: Seattle, WA, USA. https://doi.org/10.1109/icassp.1998.681598.