

Hierarchical Bayesian of ARMA Models Using Simulated Annealing Algorithm

Suparman^{*1}, Michel Doisy²

¹ Department of Mathematical Education, University of Ahmad Dahlan, Jl. Prof. Dr. Soepomo, SH., Yogyakarta, Indonesia

² ENSEEIHT/TESA, 2 Rue Camichel, Toulouse, France

*Corresponding author, email: e-mail: suparmancict@yahoo.co.id

Abstract

When the Autoregressive Moving Average (ARMA) model is fitted with real data, the actual value of the model order and the model parameter are often unknown. The goal of this paper is to find an estimator for the model order and the model parameter based on the data. In this paper, the model order identification and model parameter estimation is given in a hierarchical Bayesian framework. In this framework, the model order and model parameter are assumed to have prior distribution, which summarizes all the information available about the process. All the information about the characteristics of the model order and the model parameter are expressed in the posterior distribution. Probability determination of the model order and the model parameter requires the integration of the posterior distribution resulting. It is an operation which is very difficult to be solved analytically. Here the Simuated Annealing Reversible Jump Markov Chain Monte Carlo (MCMC) algorithm was developed to compute the required integration over the posterior distribution simulation. Methods developed are evaluated in simulation studies in a number of set of synthetic data and real data.

Keywords: Simulated Annealing, ARMA model, order identification, parameter estimation

1. Introduction

Suppose x_1, x_2, \dots, x_n is a time series data. Time series x_1, x_2, \dots, x_n is said to be an ARMA(p, q) model, if x_1, x_2, \dots, x_n satisfy following stochastic equation [3] :

$$x_t = z_t + \sum_{j=1}^q \theta_j z_{t-j} - \sum_{i=1}^p \phi_i x_{t-i}, \quad t = 1, 2, \dots, n \quad (1)$$

where z_t is the random error at the time t , ϕ_i ($i = 1, 2, \dots, p$) and θ_j ($j = 1, 2, \dots, q$) is the coefficient-coefficient. Here z_t is assumed to have the normal distribution with mean 0 and variance σ^2 . ARMA model $(x_t)_{t \in \mathbb{Z}}$ is called stationary if and only if the equality of polynomial

$$1 + \sum_{i=1}^p \phi_i^{(p)} b^i = 0$$

must lie outside the unit circle. Next ARMA model $(x_t)_{t \in \mathbb{Z}}$ is called invertible if and only if the equality of polynomial

$$1 + \sum_{j=1}^q \theta_j^{(q)} b^j = 0$$

must lie outside the unit circle [4].

The order (p, q) is assumed to be known, parameter estimation of the ARMA model has been examined by several researchers, for example [3], [4] and [12]. In practice when we match the ARMA model to the data, in general order (p, q) is not known.

While for the order (p, q) is not known, the order identification and the parameters estimation are done in two stages. The first stage is to estimate the coefficients and error variance with the assumption of order (p, q) is known. Based on the parameter estimation in the first stage, the second stage is to identify the order (p, q) . A criterium used to determine order (p, q) has been proposed by many researchers, among others are the Akaike Information Criteria Criterion (AIC), Bayesian Information Criterion (BIC), and the Final Prediction Error Criterion (FPE).

Based on the data $x_t, t = 1, 2, \dots, n$, this research proposes a method to estimate the value $p, q, \phi^{(p)}, \theta^{(q)}$, and σ^2 simultaneously. To do that, we will use a hierarchical Bayesian approach [10], which will be described below.

2. Research Method

2.1 Hierarchical Bayesian

Let $s = (x_{p+q+1}, x_{p+q+2}, \dots, x_n)$ be a realization of the ARMA (p, q) model. If the value $s_0 = (x_1, x_2, \dots, x_{p+q})$ is known, then the likelihood function of s can be written approximately as follows :

$$\ell(s|p, q, \phi^{(p)}, \theta^{(q)}, \sigma^2) = \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n-p}{2}} \exp - \frac{1}{2\sigma^2} \sum_{t=p+q+1}^n g^2(t, p, q, \phi^{(p)}, \theta^{(q)}) \quad (2)$$

where $g(t, p, q, \phi^{(p)}, \theta^{(q)}) = x_t + \sum_{i=1}^p \phi_i^{(p)} x_{t-i} - \sum_{j=1}^q \theta_j^{(q)} z_{t-j}$ for $t = p+q+1, p+q+2, \dots, n$ with initial value $x_1 = x_2 = \dots = x_{p+q} = 0$ [11]. Suppose S_p and I_q are respectively stationary region and invertible region. By using the transformation

$$F : \phi^{(p)} = (\phi_1^{(p)}, \phi_2^{(p)}, \dots, \phi_p^{(p)}) \in S_p \mapsto r^{(p)} = (r_1, r_2, \dots, r_p) \in (-1, 1)^p \quad (3)$$

$$G : \theta^{(q)} = (\theta_1^{(q)}, \theta_2^{(q)}, \dots, \theta_q^{(q)}) \in I_q \mapsto \rho^{(q)} = (\rho_1, \rho_2, \dots, \rho_q) \in (-1, 1)^q$$

the ARMA model $(x_t)_{t \in \mathbb{Z}}$ stationery if and only if $r^{(p)} = (r_1, r_2, \dots, r_p) \in (-1, 1)^p$ [2] and the ARMA model $(x_t)_{t \in \mathbb{Z}}$ invertible if and only if $\rho^{(q)} = (\rho_1, \rho_2, \dots, \rho_q) \in (-1, 1)^q$ [1]. Then the likelihood function can be rewritten as:

$$\ell(s|p, q, r^{(p)}, \rho^{(q)}, \sigma^2) = \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n-p}{2}} \exp - \frac{1}{2\sigma^2} \sum_{t=p+q+1}^n g^2(t, p, q, F^{-1}(\phi^{(p)}), G^{-1}(\theta^{(q)})) \quad (4)$$

The prior distribution determination for parameters are as follows:

a) Order p have Binomial distribution with parameters λ :

$$\pi(p|\lambda) = \binom{p_{\max}}{p} \lambda^p (1-\lambda)^{p_{\max}-p}$$

b) Order q have Binomial distribution with parameters μ :

$$\pi(q|\lambda) = \binom{q_{\max}}{q} \mu^q (1-\mu)^{q_{\max}-q}$$

- c) Order p is known, the coefficient vectors $r^{(p)}$ have uniform distribution on the interval $(-1,1)^p$.
- d) Order q is known, the coefficient vectors $\rho^{(q)}$ have uniform distribution on the interval $(-1,1)^q$.
- e) Variance σ^2 have inverse gamma distribution with parameter $\frac{\alpha}{2}$ and $\frac{\beta}{2}$:

$$\pi(\sigma^2|\alpha, \beta) = \frac{\left(\frac{\beta}{2}\right)^{\frac{\alpha}{2}}}{\Gamma\left(\frac{\alpha}{2}\right)} (\sigma^2)^{-\left(1+\frac{\alpha}{2}\right)} \exp - \frac{\beta}{2\sigma^2}$$

Here, the parameters λ and μ are assumed to have uniform distribution on the interval $(0,1)$, the value of α is 2 and the parameters β is assumed have Jeffrey distribution. So the prior distribution for the parameters $H_1 = (p, q, r^{(p)}, \rho^{(q)}, \sigma^2)$ and $H_2 = (\lambda, \beta)$ can be expressed as:

$$\pi(H_1, H_2) = \pi(p|\lambda) \pi(r^{(p)}|p) \pi(q|\mu) \pi(\rho^{(q)}|q) \pi(\sigma^2|\alpha, \beta) \pi(\lambda) \pi(\beta) \quad (5)$$

According to Bayes theorems, then the a posteriori distribution for the parameters H_1 and H_2 can be expressed as:

$$\pi(H_1, H_2|s) \propto \ell(s|H_1) \pi(H_1, H_2) \quad (6)$$

A posteriori distribution is a combination of the likelihood function and prior distribution. The prior distribution is determined before the data is taken. The likelihood function is objective while this prior distribution is subjective. In this case, the a posteriori distribution $\pi(H_1, H_2|s)$ has the form of a very complex, so it can not be solved analytically. To handle this problem, reversible jump MCMC method is proposed.

2.2 Reversible Jump MCMC Method

Suppose $M = (H_1, H_2)$. In general, the MCMC method is a method of sampling, as how to create a homogeneous Markov chain that meet aperiodic nature and irreducible ([9]) M_1, M_2, \dots, M_m to be considered such as a random variable following the distribution $\pi(H_1, H_2|s)$. Thus M_1, M_2, \dots, M_m it can be used to estimate the parameter M . To realize this, the Gibbs Hybrid algorithm ([9]) is adopted, which consists of two phases:

1. Simulation of the distribution $\pi(H_2|H_1, s)$
2. Simulation of the distribution $\pi(H_1|H_2, s)$

Gibbs algorithm is used to simulate the distribution $\pi(H_2|H_1, s)$ and the hybrid algorithm, which combine the reversible jump MCMC algorithm [5] and the Gibbs algorithm,

used to simulate the distribution $\pi(H_1|H_2, s)$. The reversible jump MCMC algorithm is generally of the Metropolis-Hastings algorithm [6], [8].

2.2.1 Simulation of the distribution $\pi(H_2|H_1, s)$

The conditional distribution of H_2 given (H_1, s) , written $\pi(H_2|H_1, s)$, can be expressed as

$$\pi(H_2|H_1, s) \propto \lambda^p (1-\lambda)^{p_{\max}-p} \mu^q (1-\mu)^{q_{\max}-q} \left(\frac{\beta}{2}\right)^{\frac{\alpha}{2}} \exp - \frac{\beta}{2\sigma^2} \frac{1}{\beta}.$$

This distribution is inversion gamma distribution with parameters $\frac{\alpha}{2}$ and $\frac{1}{2\sigma^2}$. So the Gibbs algorithm is used to simulate it.

2.2.2 Simulation of the distribution $\pi(H_1|H_2, s)$

If the conditional distribution of H_1 given (H_2, s) , written $\pi(H_1|H_2, s)$, is integrated with respect to σ^2 , then we get

$$\pi(p, q, r^{(p)}, \rho^{(q)}|H_2, s) = \int_{\mathbb{R}^+} \pi(H_1|H_2, s) d\sigma^2$$

$$\text{Let } v = \frac{\alpha}{2} + \frac{n - p_{\max}}{2} \text{ and } w = \frac{\beta}{2} + \frac{1}{2} \sum_{t=p_{\max}+1}^n g^2(t, p, q, F^{-1}(r^{(p)})G^{-1}(\rho^{(q)}))$$

and we use

$$\int_{\mathbb{R}^+} (\sigma^2)^{-(1+v)} \exp - \frac{w}{\sigma^2} d\sigma^2 = \frac{\Gamma(v)}{w^v}$$

then we get

$$\pi(p, q, r^{(p)}, \rho^{(q)}|H_2, s) \propto \binom{p_{\max}}{p} \lambda^p (1-\lambda)^{p_{\max}-p} \binom{q_{\max}}{q} \pi^q (1-\pi)^{q_{\max}-q} \left(\frac{1}{2}\right)^{p+q} \frac{\left(\frac{\beta}{2}\right)^{\frac{\alpha}{2}}}{\Gamma\left(\frac{\alpha}{2}\right)} \frac{1}{\beta} \frac{\Gamma(v)}{w^v}$$

On the other hand, we have also

$$\pi(\sigma^2|p, q, r^{(p)}, \rho^{(q)}, H_2, s) \propto (\sigma^2)^{-(v+1)} \exp - \frac{w}{\sigma^2}$$

So that we can express the distribution $\pi(H_1|H_2, s)$ as a result of multiplication of the distribution $\pi(p, q, r^{(p)}, \rho^{(q)}|H_2, s)$ and the distribution, $\pi(\sigma^2|p, q, r^{(p)}, \rho^{(q)}, H_2, s)$:

$$\pi(H_1|H_2, s) = \pi(p, q, r^{(p)}, \rho^{(q)}|H_2, s) \times \pi(\sigma^2|p, q, r^{(p)}, \rho^{(q)}, H_2, s)$$

Next to simulate the distribution $\pi(H_1|H_2, s)$, we use a hybrid algorithm that consists of two phases:

- Phase 1: Simulate the distribution of $\pi(\sigma^2|p, q, r^{(p)}, \rho^{(q)}, H_2, s)$
- Phase 2: Simulate the distribution $\pi(p, q, r^{(p)}, \rho^{(q)}|H_2, s)$

The Gibbs algorithm is used to simulate the distribution $\pi(\sigma^2|p, q, r^{(p)}, \rho^{(q)}, H_2, s)$. Conversely, the distribution $\pi(p, q, r^{(p)}, \rho^{(q)}|H_2, s)$ has a complex form. The reversible jump MCMC is used to simulate it.

When the order (p, q) is determined, we can use the Metropolis Hastings algorithm. Therefore, in the case that this order is not known, Markov chain must jump from the order (p, q) with parameters $(r^{(p)}, \rho^{(q)})$ to the order (p^*, q^*) with the parameter $(r^{(p^*)}, \rho^{(q^*)})$. To solve this problem, we use the Reversible Jump MCMC algorithm.

2.2.3 Type of jump selection

Suppose (p, q) represent actual values for the order, we will write: η_p^{AR} the probability to jump from the p to $p+1$, δ_p^{AR} the probability to jump from the p to $p-1$, ζ_p^{AR} the probability to jump from p to p , η_q^{MA} the probability to jump from q to $q+1$, δ_q^{MA} the probability to jump from the q to $q-1$ and ζ_q^{MA} the probability to jump from q to q . For each component, we will choose the uniform distribution on the possible jump. As an example for the AR, this distribution depends on p and satisfy

$$\eta_p^{AR} + \delta_p^{AR} + \zeta_p^{AR} = 1$$

We set $\delta_0^{AR} = \zeta_0^{AR} = 0$ and $\delta_{p_{\max}}^{AR} = 0$ Under this restriction, the probability will be

$$\eta_p^{AR} = c \min \left\{ 1, \frac{\pi(p+1)}{\pi(p)} \right\} \text{ and } \delta_p^{AR} = c \min \left\{ 1, \frac{\pi(p)}{\pi(p+1)} \right\}$$

with constant c , as much as possible so that $\eta_p^{AR} + \delta_p^{AR} \leq 0.9$ for $p = 0, 1, \dots, p_{\max}$. The goal is to have

$$\eta_p^{AR} \pi(p) = \delta_{p+1}^{AR} \pi(p+1)$$

2.2.4 Birth / Death of Order

As for the AR example, suppose that p is the actual value for the order of the ARMA model, $r^{(p)} = (r_1, r_2, \dots, r_p)$ is the coefficient value. Consider that we want to jump from p to $p+1$. We take the random variable u according to the triangular distribution with mean 0

$$g(u) = \begin{cases} u+1, & -1 < u < 0 \\ 1-u, & 0 < u < 1 \end{cases}$$

We complete the vector $r^{(p)}$ random variables with u . So the new coefficient vector is proposed

$$r^{(p+1)} = (r_1, r_2, \dots, r_p, u)$$

Note that this transformation will change the total value of all. It is clearly seen that the Jacobian of the transformation of value is 1.

Instead, to jump from $p+1$ to p is done by removing the last coefficients in $r^{(p+1)} = (r_1, r_2, \dots, r_p, r_{p+1})$. So the new coefficient vectors that is proposed become $r^{(p)} = (r_1, r_2, \dots, r_p)$. The probability of acceptance / rejection respectively is

$$\alpha_N = \min\{1, r_N\} \text{ and } \alpha_D = \min\{1, r_N^{-1}\}$$

where

$$r_N = \frac{\pi(p+1, q, r^{(p+1)}, \rho^{(q)} | H_2, s) q(p+1, r^{(p+1)}; p, r^{(p)})}{\pi(p, q, r^{(p)}, \rho^{(q)} | H_2, s) q(p, r^{(p)}; p+1, r^{(p+1)})}$$

We have

$$\begin{cases} q(p+1, r^{(p+1)}; p, r^{(p)}) = \delta_{p+1}^{AR} \\ q(p, r^{(p)}; p+1, r^{(p+1)}) = \eta_p^{AR} g(r_{p+1}) \end{cases}$$

Finally, we get

$$r_N = \frac{w(\beta, p+1, q, r^{(p+1)}, \rho^{(q)})^{-v(\alpha)}}{w(\beta, p, q, r^{(p)}, \rho^{(q)})^{-v(\alpha)}} \frac{p_{\max} - p}{p+1} \frac{\lambda}{1-\lambda} \frac{1}{2} \frac{\eta_{p+1}^{AR}}{\delta_p^{AR}} \frac{1}{g(r_{p+1})}$$

2.2.5 Changes in coefficients

Suppose now that the AR part is selected to jump from p to p without a order change, but only the coefficient is changed. If $r^{(p)} = (r_1, r_2, \dots, r_p)$ is the coefficient vector, we modify the coefficient vector. Consider that r_1, r_2, \dots, r_p is courant point and supposing that u_1, u_2, \dots, u_p new point, we define the point u_i in the following way:

$$u_i = \sin(r_i + s)$$

s is taken with the uniform distribution on the interval $\left[-\frac{\pi}{10}, \frac{\pi}{10}\right]$. Then u_i is selected with the distribution

$$f(u_i | r_i) = \frac{5}{\pi \sqrt{1 - u_i^2}}$$

in the interval $\left[\sin(r_i - \frac{\pi}{10}), \sin(r_i + \frac{\pi}{10})\right]$.

If $r^{*(p)} = (r_1^*, r_2^*, \dots, r_{i-1}^*, r_i, r_{i+1}, \dots, r_p)$ and $\tilde{r}^{*(p)} = (r_1^*, r_2^*, \dots, r_{i-1}^*, u_i, r_{i+1}, \dots, r_p)$, then the probability acceptance / rejection can be written by

$$\alpha_C = \min\{1, r_C\}$$

Where

$$r_C = \frac{\pi(p, q, r^{*(p)}, \rho^{(q)}, H_2 | s) q(p, r^{*(p)}; p, \tilde{r}^{*(p)})}{\pi(p, q, \tilde{r}^{*(p)}, \rho^{(q)}, H_2 | s) q(p, \tilde{r}^{*(p)}; p, r^{*(p)})}$$

because

$$\frac{q(p, r^{*(p)}; p, \tilde{r}^{*(p)})}{q(p, \tilde{r}^{*(p)}; p, r^{*(p)})} = \left(\frac{1+r_i}{1+u_i} \frac{1-r_i}{1-u_i} \right)^{\frac{1}{2}}$$

So

$$r_C = \frac{w(\beta, p, q, r^{*(p)}, \rho^{(q)})^{-v(\alpha)}}{w(\beta, p, q, \tilde{r}^{*(p)}, \rho^{(q)})^{-v(\alpha)}} \left(\frac{1+r_i}{1+u_i} \frac{1-r_i}{1-u_i} \right)^{\frac{1}{2}}$$

2.3 Simulated Annealing Algorithm

Simulated Annealing algorithm [7] is obtained by adding a line in the temperature T_1, T_2, \dots, T_m at the top of the MCMC method. Next simulated annealing algorithm will produce a Markov chain $M(T_1), M(T_2), \dots, M(T_m)$ which is no longer homogeneous. With a hypothetical on a certain T_1, T_2, \dots, T_m [14] will be convergent to maximize the value of a posteriori distribution $\pi(H_1, H_2 | s)$.

3. Results and Analysis

In As an illustration, we will apply this method to identify the order and estimate the parameter synthesis ARMA data and real ARMA data. Simulation studies are done to confirm that the performance of simulated annealing algorithm is able to work well. While case studies are given to exemplify the application of research in solving problems in everyday life.

For both synthesis ARMA data and real ARMA data, we will use the simulated annealing algorithm to identify order and estimate the parameters of the ARMA model. For this purpose, the simulated annealing algorithm is implemented for 70000 iterations with a value of initial temperature $T_0 = 10$ and the temperature is derived with the temperature factor 0.995 up to the end temperature $T_{1400} = 0.01$. Value of order p and q is limited to a maximum of 10. So that $p_{\max} = q_{\max} = 10$.

3.1 Synthetic ARMA data

Figure 1 shows a synthetic ARMA data. The data are made according to the equation (1) above, with the number of data $n = 250$, order $p = 2$, order $q = 1$, $\phi^{(2)} = (-1.36, 0.7)$, $\theta^{(1)} = (0.7)$, and $\sigma^2 = 1$.

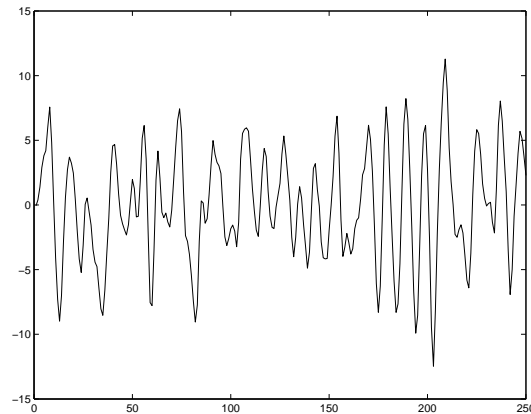


Figure 1. ARMA Synthetic Data

Based on the synthetic data in Figure 1, next order p , q order, ARMA model parameter and variance σ^2 are estimated by using the SA algorithm. The order p , q order, ARMA model parameter and variance σ^2 produced by the simulated annealing algorithm are $\hat{p} = 2$, $\hat{q} = 1$, $\hat{\phi}^{(2)} = (0.41, 0.75)$, $\hat{\theta}^{(1)} = (0.72)$ and $\hat{\sigma}^2 = 1.06$. When we compare between the actual value and the estimator value, it shows that simulated annealing algorithm can work well.

3.2 Real ARMA Data

The real data in Figure 4 below is a passenger service charge (PSC) at the Adisutjipto International Airport in Yogyakarta Indonesia for the period 55 from January 2001 to July 2005.

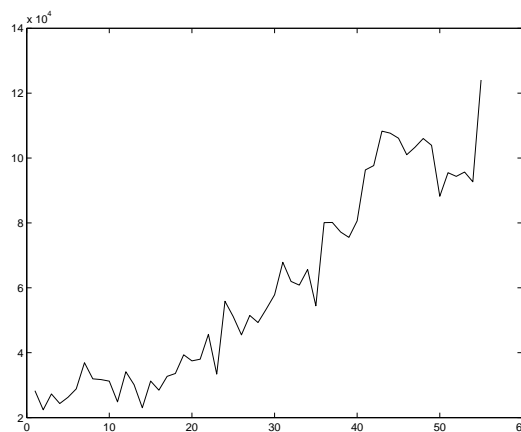


Figure 2. First distinction of PSC data at the Adisutjipto International Airport Yogyakarta.

Clearly visible in Figure 4, the data are not stationary very day. To get stationary data the first distinction is made and the results shown in Figure 3.

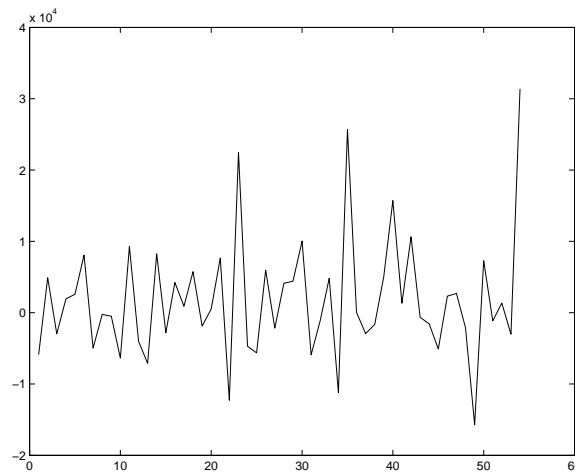


Figure 3. Second distinction of PSC data at the Adisutjipto International Airport.

Based on the data in Figure 3, next order p , q order, ARMA model parameter and variance are estimated by using the simulated annealing algorithm. The results are $\hat{p} = 1$, $\hat{q} = 0$, $\hat{\phi}^{(1)} = (0.38)$ dan $\hat{\sigma}^2 = 6.75 \times 10^7$.

4. Conclusion

The description above is a study of the theory of simulated annealing algorithms and its application in the identification of order p and q , coefficient vectors estimation $\phi^{(r)}$ and $\theta^{(q)}$, and variance estimation σ^2 from the ARMA model. The results of the simulation show that the simulated annealing algorithm can estimate the parameters well. Simulated annealing algorithm can also be implemented with good results on Synthetic Aperture Radar image segmentation [13].

As the implementation, the simulated annealing algorithm is applied to the PSC data at the Adisutjipto International Airport. Its result is that the PSC data can be modeled with the ARIMA model (1,0). The model can be used to predict the number of PSC at the Adisutjipto International Airport in the future.

References

- [1] Barndorff-Nielsen O, and Schou G. On the parametrization of autoregressive models by partial autocorrelation. *J. Multivar. Anal.* 1973; 3: 408-419.
- [2] Bhansali RJ. The inverse partial correlation function of a time series and its applications. *J. Multivar. Anal.* 1983; 13: 310-327.
- [3] Box GEP, Jenkins GM, Reinsel GC. *Time Series Analysis: Forecasting and Control*. New Jersey: Prentice Hall. 2013.
- [4] Brockwell PJ, Davis RA. *Times Series: Theory and Methods*. Springer: New York. 2009.
- [5] Green PJ. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian model determination. *Biometrika*. 1995; 82: 711-732.
- [6] Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*. 1970; 57: 97-109.
- [7] Kirpatrick S. Optimization by Simulated Annealing: Quantitative Studies. *Journal of Statistical Physics*. 1984; 16:975-986.
- [8] Metropolis N, Rosenbluth AW, Teller AH, and Teller E. Equations of state calculations by fast computing machines. *Journal Chemical Physics*. 1953; 21: 1087-1091.
- [9] Robert CP. *Méthodes de Monte Carlo par Chaînes de Markov*. Economica. 1996.
- [10] Robert CP. *The Bayesian Choice. A Decision-Theoretic Motivation*. Springer Texts in Statistics. 1994.
- [11] Shaarawy S, Broemeling L. Bayesian inferences and forecasts with moving averages processes. *Commun. Statist. - Theory Meth.* 1984; 13(15): 1871-1888.

- [12] Suparman, Soejoeti Z. Bayesian estimation of ARMA Time Series Models. *WKSJ Journal*. 1999; 2(3): 91-98.
- [13] Suparman. Segmentasi Bayesian Hierarki dalam Citra SAR dengan Menggunakan Algoritma SA. *Jurnal Pakar*. 2005; 6: 227-235.
- [14] Winkler G. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods : A Mathematical Introduction*. New York: Springer Verlag. 2012.