International Journal of GEOMATE, March, 2022, Vol.22, Issue 91, pp.38-45
ISSN: 2186-2982 (P), 2186-2990 (O), Japan, DOI: https://doi.org/10.21660/2022.91.gxi270
Geotechnique, Construction Materials and Environment

DATA MODELING WITH AUTOREGRESSIVE BASED ON REVERSIBLE JUMP MCMC SIMULATION: COMPARING GAUSSIAN AND LAPLACIAN NOISE

*Suparman[1], Mahyudin Ritonga[2], Ahmad Muhammad Diponegoro[3], Mohamed Nor Azhari Azman[4], Eviana Hikamudin[5]

[1]Faculty of Teacher Training and Education, Universitas Ahmad Dahlan, Indonesia
[2]Faculty of Islamic Studies, Muhammadiyah University of West Sumatera, Indonesia
[3]Faculty of Psychology, Universitas Ahmad Dahlan, Indonesia
[4]Faculty of Technical and Vocational Education, Universiti Pendidikan Sultan Idris, Malaysia
[5]Faculty of Educational Science, Universitas Pendidikan Indonesia, Indonesia

*Corresponding Author, Received: 30 Nov. 2021, Revised: 29 Dec. 2021, Accepted: 30 Jan. 2022

ABSTRACT: The autoregressive model (AR) is one of the stochastic models in the time series that is used for forecasting. The AR model is affected by noise which has a distribution. The accuracy in choosing the noise distribution has an impact on the fit of the AR model to the data. This paper presents an AR model in which the noise has a Laplace distribution. And also, the Laplacian AR model is compared with the Gaussian AR model. The Bayesian approach was adopted to estimate the AR model parameters. The Binomial distribution was chosen as the prior distribution for the older model, the uniform distribution was chosen as the prior distribution for the AR model coefficients. The Bayesian estimator for the AR model parameters is calculated based on the posterior distribution with the help of the reversible jump algorithm Markov Chain Monte Carlo (MCMC). The results in this paper indicate that the reversible jump MCMC algorithm is categorized as valid in estimating the parameters of the AR model. Based on a simulation study, this paper shows that the Laplacian AR model can be used as an alternative to approximate an AR model that contains non-Gaussian noise. To support this finding, the research can be studied further from a theoretical point of view. With the help of the reversible jump MCMC algorithm, the Bayesian estimator for the AR model parameters is computed based on the posterior distribution. According to the findings of this paper, the reversible jump MCMC algorithm is suitable for estimating the parameters of the AR model. This research illustrates that the Laplacian AR model can be utilized as an alternative to approximate an AR model with non-Gaussian noise, based on a simulation analysis. The findings can be investigated further from a theoretical standpoint to support this finding.

Keywords: Autoregressive processes, Bayes methods, Gaussian noise, Laplacian noise, Monte Carlo methods

1. INTRODUCTION

The Autoregressive (AR) model is one of the stochastic models in the time series. Forecasting, model determination, estimation, and process management can all be done with stochastic models, including AR models [1]. In engineering areas as well as other subject areas, data often appear as time series. Along with the times, the area of application of statistical models is increasing. Therefore, the development of AR models is a topic that is still open for research.

Like the stochastic model in general, the AR model contains noise. The noise selection correlates with the accuracy of the AR model in matching the data. In the theory and methods related to the AR model, various kinds of literature assume that the noise has a Gaussian (normal) distribution, for example [2]. In various applications, some data do not meet the assumption of normality. Thus, some literature proposes non-Gaussian AR models such as [3-4]. In addition, Gaussian noise is used as an approximation when the related theory of non-Gaussian noise has not been found, for example [5]. On the other hand, several types of noise distribution are used in data modeling, for example [6-10]. In [6], Laplace noise is used in the sparse representation. Gamma noise and Gaussian noise are employed in the picture model in [7]. The Signal model in [8] uses both Rayleigh and Gaussian noise. The pictured model in [9] uses Cauchy noise. Meanwhile, the Internal HIV model in [10] employs Gaussian noise. Furthermore, the literature has examined the differences between non-Gaussian and Gaussian noise, for example [11]. However, there hasn't been much research on the comparison of Laplacian and Gaussian noise in the AR model. As a result, the goal of this research is to compare Laplacian and Gaussian noise in the AR model. In addition, this article also aims to determine whether Laplacian noise can be used as an alternative in approximating non-Gaussian AR models.

The order of the AR model, the AR model coefficients, and the AR model noise variance are all AR model parameters. There are two techniques to estimate the parameters of the AR model: the Bayesian approach and the non-Bayesian approach. The Bayesian statistical model is created using a parametric statistical model and the prior distribution of parameters in the Bayesian technique [12]. In this paper, parameter estimation uses the Bayesian approach. The AR model parameter space is a combination of parameter spaces that have different dimensions because the model order is part of the parameters. Green [13] suggested the reversible jump MCMC algorithm, which is an extension of the metropolis-Hasting algorithm. The reversible jump MCMC algorithm has an advantage over the Metropolis-Hasting algorithm in that it generates a Markov chain that can hop across parameter spaces of different dimensions. To estimate the AR model parameters with both Gaussian and Laplacian noise, this paper uses the reversible jump MCMC approach.

2. RESEARCH SIGNIFICANCE

In modeling time series data using the AR model, the literature often assumes that the noise has a Gaussian distribution. And the AR modeled time series analysis was developed based on the assumption of normality. Sometimes, the Gaussian AR model is used to approximate the non-Gaussian AR model when the method for the non-Gaussian AR model is not available. This paper proposes a Laplacian AR model. This paper also compares the Laplacian AR model and the Gaussian AR model in approximating the AR model with different distributed noise. The significance of this paper is to provide a Laplacian AR model as an alternative in time series analysis data modeling. In addition, this paper contributes to providing an alternative in approximating non-Gaussian AR models.

3. METHODOLOGY

The AR model parameters are estimated using a Bayesian technique. The binomial distribution is the prior distribution for the older model. The AR model parameters' prior distribution is a uniform distribution. The inverse Gamma distribution is the prior distribution for a noise variance. The reversible jump Markov Chain Monte Carlo (MCMC) process is used to generate the Bayes estimator, which is based on the posterior distribution. Reference [2] discusses the theory behind the reversible jump MCMC approach for parameter estimation of the Gaussian AR model. Meanwhile, [14] discusses the theory behind the reversible jump MCMC technique for parameter estimation of the Laplacian AR model.

The performance of the reversible jump MCMC algorithm was validated by a simulation study using six artificial data sets. Theories related to modeling and simulation can be studied in various literature, for example [15]. The first three artificial data sets were created following the Gaussian AR model. The next three artificial data sets were created following the Laplacian AR model. Furthermore, a comparison between the Gaussian AR model and the Laplacian AR model was carried out using three artificial data sets. These three artificial data sets were created following the uniform AR model.

4. RESULTS AND DISCUSSION

This section discusses the validation of the reversible jump MCMC algorithm [13] in estimating the parameters of the AR model, both Gaussian noise and Laplacian noise. Then, the Laplacian AR model is compared with the Gaussian AR in approximating the AR model containing different noise.

4.1 Algorithm Validation

A simulation study is used to validate the algorithm. The Gaussian AR model was used to produce the artificial data set. The reversible jump MCMC technique for the Gaussian AR model is then applied to this fabricated data set as input. A model parameter estimator is the algorithm's output. The algorithm is categorized as valid if the parameter estimator is close to the model parameter value. In the same way, the algorithm for the Laplacian AR model is validated using an artificial data set created using the Laplacian AR model.

3.4.1 Gaussian AR model

Three artificial data sets were created based on the Gaussian AR model. The creation of this artificial data set uses parameter values as shown in Table 1. The first line is the parameter value for the first artificial data, the second line is the parameter value for the second artificial data, and the third line is the parameter value for the third artificial data. Three artificial data sets are presented in Figs. 1-3. These three artificial data sets were made of 250 each.

Using three artificial data sets presented in

Figs. 1-3 as input, the reversible jump MCMC algorithm [13] is used to find the Gaussian AR model parameter estimator. More details regarding the use of algorithms in the estimation of Gaussian AR model parameters can be found in [2]. The output of the MCMC reversible jump algorithm is a Markov chain. This Markov chain is used to estimate the Gaussian AR model parameters. The histogram of the order Gaussian AR model is presented in Figs. 4-6. Fig. 4 is the histogram of the order for the first artificial data (Fig. 1), Fig. 5 is the histogram of the order for the second artificial data (Fig. 2), and Fig. 6 is the histogram of the order for the third artificial data (Fig. 3).

Table 1 Parameter values for the three artificial data sets were created based on the Gaussian AR model

Artificial data	p	ϕ	σ^2
1	4	(0.32, 0.04, -0.41, -0.71)	4
2	5	(1.69, 1.35, -0.32, -1.07, -0.81)	4
3	6	(-0.08, -0.80, -0.06, 0.93, 0.15, -0.80)	4

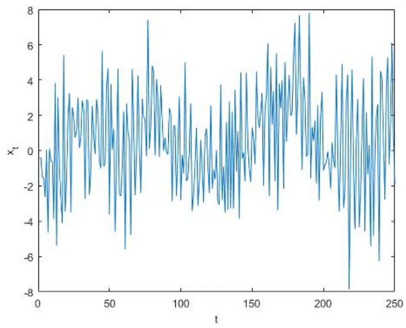

Fig.1 Artificial data with Gaussian AR (p = 4)

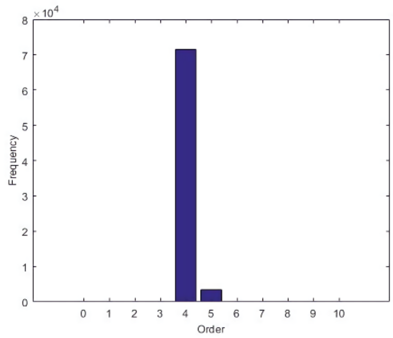

Fig.4 Histogram of the order for the data in Fig.1

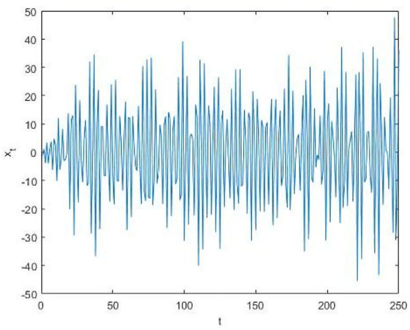

Fig. 2 Artificial data with Gaussian AR (p = 5)

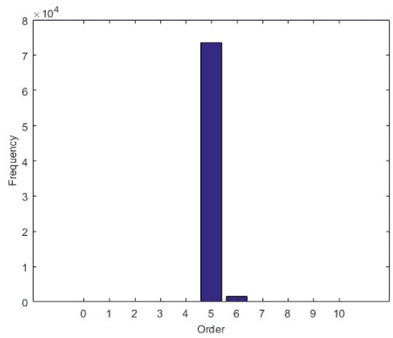

Fig.5 Histogram of the order for the data in Fig.2

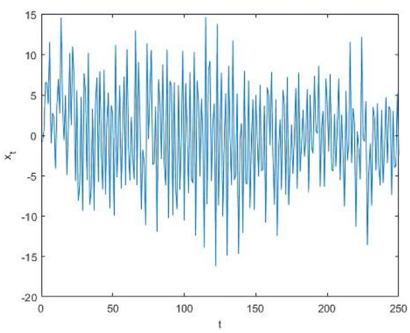

Fig.3 Artificial data with Gaussian AR (p = 6)

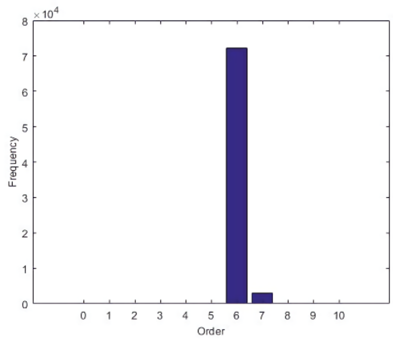

Fig.6 Histogram of the order for the data in Fig.3

Fig. 4 shows that the order reached its maximum on the 4th order. These results indicate that the order model estimator is $\hat{p} = 4$ (for the first artificial data). Fig. 5 shows that the order reached its maximum on the 5th order. These results indicate that the order model estimator is $\hat{p} = 5$ (for the second artificial data). While Fig. 6 indicates that the order reached its maximum on the 6th order. These results indicate that the order model estimator is $\hat{p} = 6$ (for the third artificial data).

With condition at $\hat{p} = 4$, the coefficient estimator of the AR model and the estimator of the noise variance of the AR model (for the first artificial data) are calculated and the results are presented in Table 2 (second row). In the same way, the third row presents the coefficient estimator of the AR model and the estimator of the noise variance of the AR model (for the second artificial data). While the fourth line presents the coefficient estimator of the AR model and the estimator of the noise variance of the AR model (for the third artificial data).

Table 2 Parameter estimator for three artificial data sets (Gaussian AR model)

Artificial data	$\hat{p}$	$\hat{\phi}$	$\hat{\sigma}^2$
1	4	(0.32, -0.02, -0.36, -0.74)	3.94
2	5	(1.73, 1.46, -0.17, -0.96, -0.75)	3.75
3	6	(-0.78, -0.83, -0.04, 0.90, 0.14, -0.84)	4.35

4.2 Laplacian AR model

As for the case of the Gaussian AR model, the results for the Laplacian AR model are presented in Tables 3-4 and Figs. 7-12. Table 3 shows the parameter values for the three artificial data sets with the Laplacian AR model. Figs. 7-9 shows the artificial data created using the parameters in Table 3. The Laplacian AR model of order 4 is presented in Fig. 7. The Laplacian AR model of order 5 is presented in Fig.8. The Laplacian AR model of order 6 is presented in Fig. 9.

While Figs 10-12 show a histogram of the order for the 3 artificial data. Fig. 10 indicates that the maximum value was reached on the 4rd order. Fig. 11 indicates that the maximum value was reached on the 5th order. Fig. 12 indicates that the maximum value was reached on the 6th order.

Table 4 presents parameter estimators for the Laplacian AR model generated by the MCMC reversible jump algorithm. The algorithm is run for 100,000 iterations and the burn-in period is 25,000.

Table 3 Parameter values for the three artificial data sets were created based on the Laplacian AR model

Artificial data	p	ϕ	σ^2
1	4	(1.50, 1.05, 0.77, 0.42)	4
2	5	(-0.42, -0.16, 0.62, -0.27, 0.35)	4
3	6	(1.08, 0.87, 1.01, 0.70, 1.10, 0.69)	4

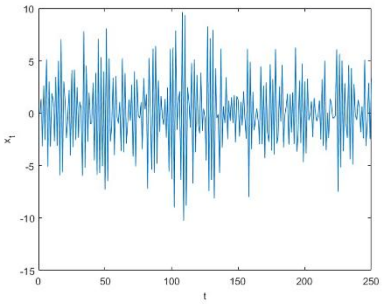

Fig.7 Artificial data with Laplacian AR (p = 4)

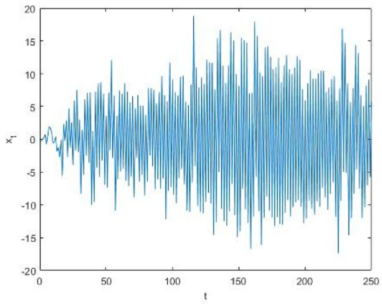

Fig.8 Artificial data with Laplacian AR (p = 5)

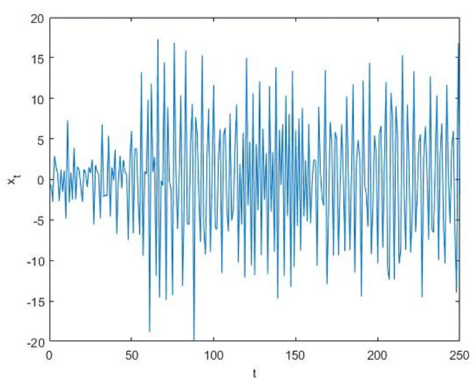

Fig.9 Artificial data with Laplacian AR (p = 6)

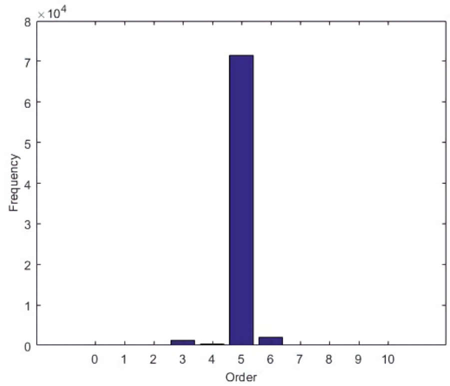

Fig.11 Histogram of the order for the data in Fig.8

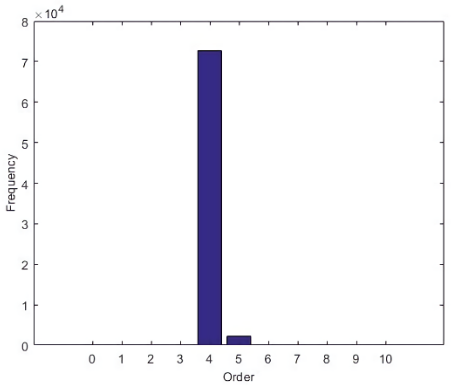

Fig.10 Histogram of the order for the data in Fig.7

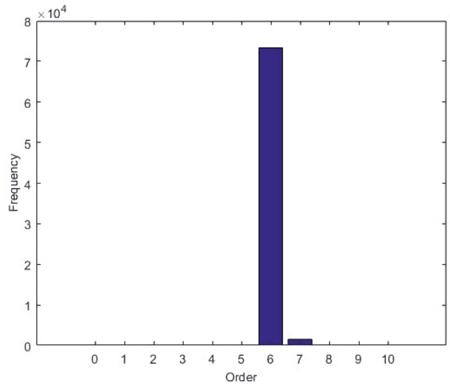

Fig.12 Histogram of the order for the data in Fig.9

Table 4 Parameter estimator for three artificial data sets (Laplacian AR model)

Artificial data	$\hat{p}$	$\hat{\phi}$	$\hat{\sigma}^2$
1	4	(1.31, 0.83, 0.70, 0.45)	4.80
2	5	(-0,40, -0.14, 0.62, -0.25, 0.37)	6.34
3	6	(1.06, 0.85, 0.99, 0.70, 1.12, 0.68)	6.64

4.2 Comparison between Laplacian and Gaussian Noises

The comparison between the Laplacian AR model and the Gaussian AR model was carried out through a simulation study using three artificial data sets. Three artificial data sets were created following the Uniform AR model with parameter values presented in Table 5.

In the first case, these three artificial data sets are approximated using the Gaussian AR model. Using these three artificial data sets as input, the reversible jump MCMC algorithm is used to find the Gaussian AR model parameter estimator. The parameter estimators of the Laplacian AR model are presented in Table 6.

In the second case, these three artificial data sets were approximated using the Laplacian AR model. Similarly, the parameter estimators of the Laplacian AR model are presented in Table 7.

Table 5 Parameter values for three artificial data sets (Uniform AR model)

Artificial data	p	ϕ	σ^2
1	4	(0.00, -0.06, -0.52, 0.52)	0.33
2	5	(-1.58, 1.66, -1.23, 0.51, -0.34)	0.33
3	6	(-0.35, 0.29, -0.72, 0.46, -0.31, 0.89)	0.33

Table 6 Parameter estimators for three artificial data sets (Approximated by Gaussian AR model)

Artificial data	$\hat{p}$	$\hat{\phi}$	$\hat{\sigma}^2$
1	4	(-0.08, -0.15, -0.57, 0.53)	0.32
2	5	(-1.57, 1.70, -1.29, 0.57, -0.38)	0.32
3	6	(-0.33, 0.30, -0.71, 0.44, -0.28, 0.86)	0.33

Table 7 Parameter estimators for three artificial data sets (Approximated by Laplacian AR model)

Artificial data	$\hat{p}$	$\hat{\phi}$	$\hat{\sigma}^2$
1	4	(-0.09, -0.17, -0.58, 0.55)	0.65
2	5	(-1.58, 1.66, -1.23, 0.51, -0.34)	0.61
3	6	(-0.35, 0.30, -0.68, 0.46, -0.30, 0.81)	0.61

4.3 Discussion

For parameter estimation of the Gaussian AR model, the distances between the parameter estimators in Table 9 (third column) and the parameter values in Table 9 (second column) were calculated using the Euclidean distance. These Euclidean distances are presented in Table 9 (fourth column). Similarly, the Euclidean distances between parameter estimators and parameter values for the Laplacian AR model are presented in Table 10 (fourth column).

The Euclidean distance is relatively small, as shown in Tables 8 (fourth column) and 9 (fourth column). This suggests that the Gaussian AR model's parameters can be successfully estimated using the reversible jump MCMC approach. The reversible jump MCMC approach is included in the valid category in estimating the parameters of the Gaussian AR model, according to this simulation. The same results were found when the Laplacian AR model was estimated. As a result, the reversible jump MCMC approach is considered to be valid for estimating both Laplacian and Gaussian noise parameters in the AR model.

Table 10 (fourth column) also shows that the Euclidean distance is relatively small. And in Table 11 (fourth column) it is also seen that the Euclidean distance is relatively small. This means that both Gaussian AR and Laplacian AR models can be used as alternatives to approximate the Uniform AR model.

Table 8 The Euclidean distances for three artificial data sets (Gaussian AR model)

Artificial data	ϕ	$\hat{\phi}$	$\lvert\phi - \hat{\phi}\rvert$
1	(0.32, 0.04, -0.41, -0.71)	(0.32, 0.04, -0.41, -0.71)	0.22
2	(1.69, 1.35, -0.32, -1.07, -0.81)	(1.73, 1.46, -0.17, -0.96, -0.75)	0.47
3	(-0.08, -0.80, -0.06, 0.93, 0.15, -0.80)	(-0.78, -0.83, -0.04, 0.90, 0.14, -0.84)	0.09

Table 9 The Euclidean distances for three artificial data sets (Laplacian AR model)

Artificial data	ϕ	$\hat{\phi}$	$\lvert\phi - \hat{\phi}\rvert$
1	(1.50, 1.05, 0.77, 0.42)	(1.31, 0.83, 0.70, 0.45)	0.46
2	(-0.42, -0.16, 0.62, -0.27, 0.35)	(-0,40, -0.14, 0.62, -0.25, 0.37)	0.08
3	(1.08, 0.87, 1.01, 0.70, 1.10, 0.69)	(1.06, 0.85, 0.99, 0.70, 1.12, 0.68)	0.07

Table 10 The Euclidean distances for three artificial data sets (Approximate by Gaussian noise)

Artificial data	ϕ	$\hat{\phi}$	$\lvert\phi - \hat{\phi}\rvert$
1	(0.00, -0.06, -0.52, 0.52)	(-0.08, -0.15, -0.57, 0.53)	0.21
2	(-1.58, 1.66, -1.23, 0.51, -0.34)	(-1.57, 1.70, -1.29, 0.57, -0.38)	0.01
3	(-0.35, 0.29, -0.72, 0.46, -0.31, 0.89)	(-0.33, 0.30, -0.71, 0.44, -0.28, 0.86)	0.04

Table 11 The Euclidean distances for three artificial data sets (Approximate by Laplacian noise)

Artificial data	$\boldsymbol{\phi}$	$\widehat{\boldsymbol{\phi}}$	$\lvert\boldsymbol{\phi}-\widehat{\boldsymbol{\phi}}\rvert$
1	(0.00, -0.06, -0.52, 0.52)	(-0.09, -0.17, -0.58, 0.55)	0.22
2	(-1.58, 1.66, -1.23, 0.51, -0.34)	(-1.58, 1.66, -1.23, 0.51, -0.34)	0.02
3	(-0.35, 0.29, -0.72, 0.46, -0.31, 0.89)	(-0.35, 0.30, -0.68, 0.46, -0.30, 0.81)	0.03

The findings in this paper are examined in terms of a simulation study using several artificial data sets. The Gaussian AR model and the Laplacian AR model were compared using an artificial data set with the uniform AR model. To support this, further research can be studied theoretical evidence so that these findings can be applied to various types of data.

5. CONCLUSION

For both Gaussian noise and Laplacian noise, the reversible jump MCMC algorithm is categorized as valid in estimating the AR model parameters. Simulation studies using several artificial data sets show that both the Gaussian AR model and the Laplacian AR model can be used to approximate the Uniform AR model. Between the Gaussian AR model and the Laplacian AR model, there is no significant difference in approximating the uniform AR model.

The findings in the study were validated by a simulation study using several artificial data sets. To support these findings, further research can be studied from a theoretical point of view so that the findings can be applied to a wider range of data. In addition, further research can study AR models for other types of noise.

6. ACKNOWLEDGMENTS

The authors would like to thank the Ministry of Education, Culture, Research and Technology (Indonesia) for the research grant (Basic research scheme). In addition, the authors also thank the reviewers of the International Journal of GEOMATE for their suggestions and comments. Their suggestions and comments are very useful as a reference in improving the quality of this paper.

7. REFERENCES

[1] Box G.E.P., Jenkins G.M., Reinsel G.C., and Ljung G.M., Time Series Analysis: Forecasting and Control, John Wiley & Sons, 2015.

[2] Suparman and Rusiman M.S., Hierarchical Bayesian Estimation for Stationary Autoregressive Models Using Reversible Jump MCMC Algorithm, Int. J. Eng. Technol., vol. 7, 2018, pp. 64–67.

[3] VodiČka V., MasŠČuch J., and Kolovratník M., ANNALS of Faculty Engineering Hunedoara, Int. J. Eng. - Ann. Fac. Eng. Hunedoara, vol. XII, no. 2006, 2014, pp. 33–40.

[4] Suparman, A new estimation procedure using a reversible jump MCMC algorithm for ar models of exponential white noise, Int. J. GEOMATE, vol. 15, no. 49, 2018, pp. 85–91.

[5] Vatta F., Soranzo A., Comisso M., Buttazzoni G., and Babich F., A survey on old and new approximations to the function φ(X) involved in ldpc codes density evolution analysis using a Gaussian approximation, Inf., vol. 12, no. 5, 2021.

[6] Miertoiu F.I. and Dumitrescu B., Feasibility pump algorithm for sparse representation under Laplacian noise, Math. Probl. Eng., vol. 2019, 2019.

[7] Ullah A., Chen W., Khan M.A., and Sun H.G., A new variational approach for multiplicative noise and blur removal, PLoS One, vol. 12, no. 1, 2017, pp. 1–27.

[8] Li H., Wu J., Miao A., Yu P., Chen J., and Zhang Y., Rayleigh-maximum-likelihood bilateral filter for ultrasound image enhancement, Biomed. Eng. Online, vol. 16, no. 1, 2017, pp. 1–23.

[9] Jung M. and Kang M., Image restoration under Cauchy noise with sparse representation prior and total generalized variation, J. Comput. Math., vol. 39, no. 1, 2021, pp. 81–107.

[10] Wang X., Li Y., and Wang X., The Stochastic Stability of Internal HIV Models with Gaussian White Noise and Gaussian Colored Noise, Discret. Dyn. Nat. Soc., 2019, vol. 2019, 2019.

[11] Haekal M., Septiawan R.R., Haryanto F., and Arif I., A comparison on the use of Perlin-noise and Gaussian noise-based augmentation on X-ray classification of a lung cancer patient, J. Phys. Conf. Ser., vol. 1951, no. 1, 2021, pp. 1–6.

[12] Robert C.P., The Bayesian Choice. Springer, 2001.

[13] Green P.J., Reversible jump Markov chain

Monte Carlo computation, and Bayesian model determination, Biometrika, vol. 82, no. 4, 1995, pp. 711–743.

[14] Suparman and Diponegoro A.M., Mathematical Model with Laplace Autoregressive Process for Heart Rate Signals, Technol. Reports Kansai Univ., vol. 62, no. 03, 2020, pp. 1001–1011.

[15] Law M. and Kelton W.D., Simulation Modeling and Analysis, Mc Graw Hill, 2000.