International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 22 (2018) pp. 15688-15695

Radioelectric Spectrum Prediction based in ARIMA and SARIMA Time Series Models

Cesar Hernandez[1*], Diego Giral[1], Fredy Martinez[1]

[1] Universidad Distrital Francisco José de Caldas, Technological Faculty, Bogotá, Colombia.

[*]Corresponding author, [1]Orcid: 0000-0001-9409-8341*

Abstract

Predicting the behavior of the primary user in wireless networks enables significant reduction of the interference level caused by the secondary user during his change of channel. Therefore, the purpose of this article is to present a comparative evaluation of the models for time series: ARIMA and SARIMA, that can predict the behavior of the primary user as well as the spectral opportunities for wireless networks in the Wi-Fi frequency band. The performance of the two models for time series will be contrasted with seven evaluation metrics. The results obtained show that the SARIMA model has the best performance in general.

Keyword: Cognitive radio, handoff, prediction, spectrum opportunities, time series, wireless networks

INTRODUCTION

Currently the amount of traffic transported by wireless networks is growing constantly due to the increase in the number of users and the increase in wireless applications. The above, together with a policy of fixed allocation of frequencies, has produced a shortage in available frequency bands. However, the results of studies carried out show that certain bands, such as those from 50 MHz to 700 MHz, are being underutilized, since their useful cycles are practically nil and that in some cases the spectral utilization times are less than 10% [1], in contrast to other bands such as those assigned to the cellular network that are currently saturated. The federal communications commission (FCC) has informed of temporal and geographic variations in the use of the spectrum in a 15 to 85% range [2].

With the purpose of achieving a more efficient use of the spectrum, cognitive radio (CR) technology proposes a dynamic spectrum assignment (DSA) [3]. It consists on non-licensed users also known as secondary users (SU) or cognitive radio users, use SO within licensed frequency bands which are assigned to licensed users also known as primary users (PU) without interrupting any process in those bands. To achieve this, the CR interacts dynamically with the environment and modifies the necessary operation parameters with the purpose of harnessing the unused spectrum and not interfering with the PU [4], [5].

The purpose of this article is to present a comparative evaluation of two models based on time series: ARIMA and SARIMA, in order to predict spectral opportunities for cognitive radio networks in the Wi-Fi frequency band. The performance of the two models will be contrasted later on with seven evaluation metrics: number of total handoffs, number of failed handoffs, number of handoffs with interference, number of perfect handoffs and number of anticipated handoffs.

The article is made up of five sections including the introduction. The second section describes the generalized mathematical model of each method. In the third section the used methodology is presented. The fourth section presents the results. Finally, in the fifth section, the conclusions are drawn.

TIME SERIES MODELS

These methods model time series by studying the structure of correlation that the time, index or distance induce in the random variables originating the series. The strategy in these models consists on: 1) Stabilizing the variance and eliminating the tendency and stationality of the series through transformations and/or differences which leads to a stationary series. 2) For the resulting series, a model is estimated with the purpose of explaining the correlation structure of the time series. 3) Inverse transformations are applied to the model obtained in step 2 so the variance, tendency and stationality of the original series can be established [6]–[8].

The three fundamental models based on time series that are autoregressive Integrated Moving Average (ARIMA) are: The Auto-Regressive (AR), the Moving Average (MA) and the Auto-Regressive of Moving Average (ARMA). These time series have been studied quite a lot in the current literature, which is why in this research more sophisticated time series such as ARIMA and SARIMA will be used.

ARIMA Model

The AR model considers that the value of the stationary series in present time t depends on all past values that the series has taken, pondered by a weight factor φ_j. The latter measures the present influence of the past value; and of a present random perturbation [9].

The AR model is described in equation (1) where φ_j correspond to the parameters of the model and a_t is an error term (or white Gaussian noise process term), i.e., random variables with a null average, constant variance, uncorrelated between them and the series' past values.

$$Z_t = \sum_{j=1}^{\infty} \varphi_j Z_{t-j} + a_t \tag{1}$$

The AR process is a regression model where the explicative variables are the same delayed dependent variable. A condition for the AR model being stationary is that $\varphi_{j<1}$ [10]. Only when the last past values p of the series affect significantly the present value, the model is called AR of order p, AR (p) and in this case, the upper limit of the sum in equation (1) is p. To determine the value of p, the Partial Auto-Correlation Function (PACF) is used.

The MA model considers that the value of the stationary series oscillates or moves around the average called μ. Additionally, it assumes that the displacement of μ in present time t is caused by infinite perturbations occurred in the past pondered by a factor θ_j that measures the influence of such perturbation in the present of the series [9].

The MA model is described in equation **Error! Reference source not found.**2) where φ_j correspond to the parameters of the model and a_t is an error term (or white Gaussian noise process term), i.e., random variables with a null average, constant variance, uncorrelated between them and the series' past values.

$$Z_t = \mu + \sum_{j=0}^{\infty} \theta_j a_{t-j} \tag{2}$$

The MA model assumes that all observations of the time series are equally important for estimating the predicted parameter. Only when the last past perturbations affect significantly the present value of the series is the model called MA of order q noted MA (q) and in this case the sum in equation (2) has q as upper limit. The average of the most recent data values q of the time series are used to forecast during the next period. To determine the value of q, the Auto-Correlation Function (ACF) is used.

The ARMA model corresponds to the combination of the AR (p) and MA (q) models to produce the ARMA (p, q) model. The ARMA model is described by equation (3).

$$Z_t = \mu + \sum_{j=0}^{\infty} \theta_j a_{t-j} + \sum_{j=1}^{\infty} \varphi_j Z_{t-j} + a_t \tag{3}$$

In general, time series are not stationary but can be transformed into stationary with the use of transformations of variance and differences.

The ARIMA (p, d, q) models are the result of integrating into the ARMA (p, q) the differences and transformations that were necessary to convert the initial series into a stationary one. The number of differences and transformations of the series define the parameter d of the model [9].

SARIMA Model

In general, if a time series exhibits potential seasonality indexed by s, then using a multiplied seasonal ARIMA(p,d,q)(P,D,Q)s model is advantageous, where d is the level of non-seasonal differencing, p is the autoregressive (AR) non-seasonal order, q is the moving average (MA) non-seasonal order, P is the number of seasonal autoregressive terms, D is the number of seasonal differences, and Q is the number of seasonal moving average terms. The seasonal autoregressive integrated moving average model of Box and Jenkins [10] is given in the equation (4) [11].

$$\phi_p(B)\varphi_P\left(B^s\right)\nabla^d\nabla_s^D x_t = \theta_q\left(B\right)\Theta_Q\left(B^s\right)e_t \tag{4}$$

Where B is the backward shift operator, xt is the observed time series of load at time t, et is the independent, identical, normally distributed error (random shock) at period t; $\nabla_S^D = \left(1-B^S\right)^D x_t$, $\varphi_P\left(B^S\right)$ and $\Theta_Q\left(B^S\right)$ are the seasonal *AR(p)* and *MA(q)* operators, respectively, which are defined in equations (5) and (6).

$$\phi_p\left(B^S\right) = 1 - \phi_1 B^s - \phi_2 B^{2s} - \dots - \phi_p B^{Ps} \tag{5}$$

$$\Theta_Q\left(B^S\right) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs} \tag{6}$$

where ϕ1, ϕ2,…, ϕp are the parameters of the seasonal AR(p) model, Θ1, Θ2,…, ΘQ are the parameters of the seasonal MA(q).

The Box-Jenkins methodology consists of four iterative steps [12], [13]:

- Step 1: Identification. This step focuses on the selection of d, D, p, P, q and Q. The number of order can be identified by observing the sample autocorrelations (ACF) and sample partial autocorrelations (PACF).
- Step 2: Estimation. The historical data is used to estimate the parameters of the tentative model in Step 1.
- Step 3: Diagnostic checking. Diagnostic test is used to check the adequacy of the tentative model.
- Step 4: Forecasting. The final model in Step 3 is used to forecast the values.

METHODOLOGY

To evaluate the performance of the proposed algorithms: ARIMA and SARIMA, seven evaluation metrics are described in Table 1.

Table 1. Evaluation Metrics Used for the Evaluation of the Proposed Algorithms.

Name	Description	Type of EM
Number of total handoffs	It corresponds to the total handoffs during the 10-minute transmission.	Cost
Number of failed handoffs	It is the number of Handoffs that the SU could not materialize because he found the respective targeted SO occupied.	Cost
Number of handoffs with interference	It is the total number of reactive handoffs carried out once the PU arrives, during the 10 minutes of transmission of the SU.	Cost
Number of anticipated handoffs	It is the number of AAPH carried out way before the PU's arrival during the 10 minutes of transmission of the SU.	Cost
Number of perfect handoffs	It is the number of AAPH carried out very closely to the PU's arrival but without interfering on him during the 10 minutes of transmission of the SU.	Benefit
Average bandwidth	It is the average bandwidth of the communication during the 10 minutes of transmission of the SU.	Benefit
Accumulative delay	It is the accumulative delay of the communication during the 10 minutes of transmission of the SU.	Cost

In order to assess the performance of each developed handoff, a simulation environment progressively reconstructs the behavior of the spectrum occupancy with the use of the captured data traces in the frequency Wi-Fi band. These allows to accurately evaluate the behavior of the PUs and also, to assess and validate the performance of each handoff. The spectral occupancy data corresponds to a week-long observation captured at Bogota City in Colombia [14].

RESULTS

The Figure 1 to Figure 7 show the performance of the metrics for the Wi-Fi network: number of total handoffs, number of failed handoffs, number of handoffs with interference, number of perfect handoffs, number of anticipated handoffs, average bandwidth and average delay.

Table 2 shows the values obtained for each of the seven metrics. The model with the best performance corresponds to the one that obtained the lowest values for each of the cost metrics and the highest for the benefit metrics. In accordance with the above, the SARIMA model presents the best performance because the cost metrics compared to the ARIMA model are, 23% of the number of handoffs, 8% of the failed handoffs, 80% of the interference handoffs, and 13% of the delay, only for the anticipated handoffs, ARIMA has the best performance with 54% of those that SARIMA has. With respect to the benefit metrics, SARIMA has 77% more perfect handoffs and a very similar bandwidth.

Table 2. Evaluation Metrics.

Model	Handoffs	Failed handoffs	Anticipated handoffs	Interference handoffs	Perfect handoffs	Average Bandwidth (kHz)	Average Delay (s)
ARIMA	7217	6615	296	247	59	1164	791,1
SARIMA	1673	826	544	198	105	1172	261,3

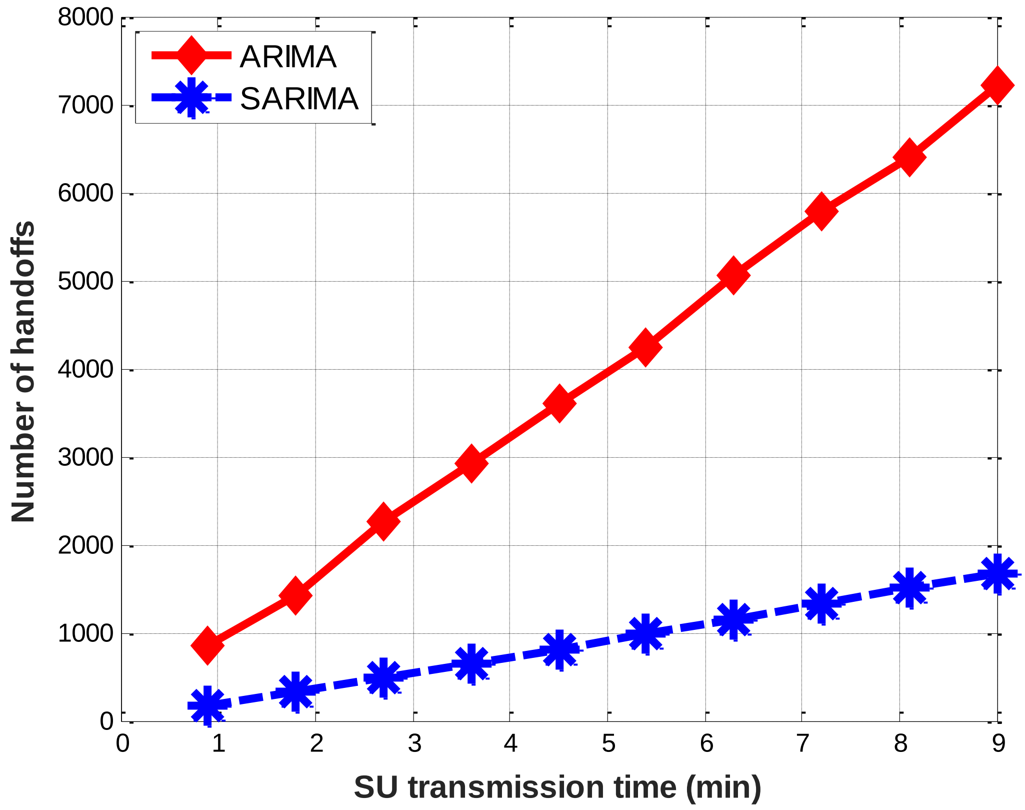

Figure 1. Number of handoff for ARIMA and SARIMA

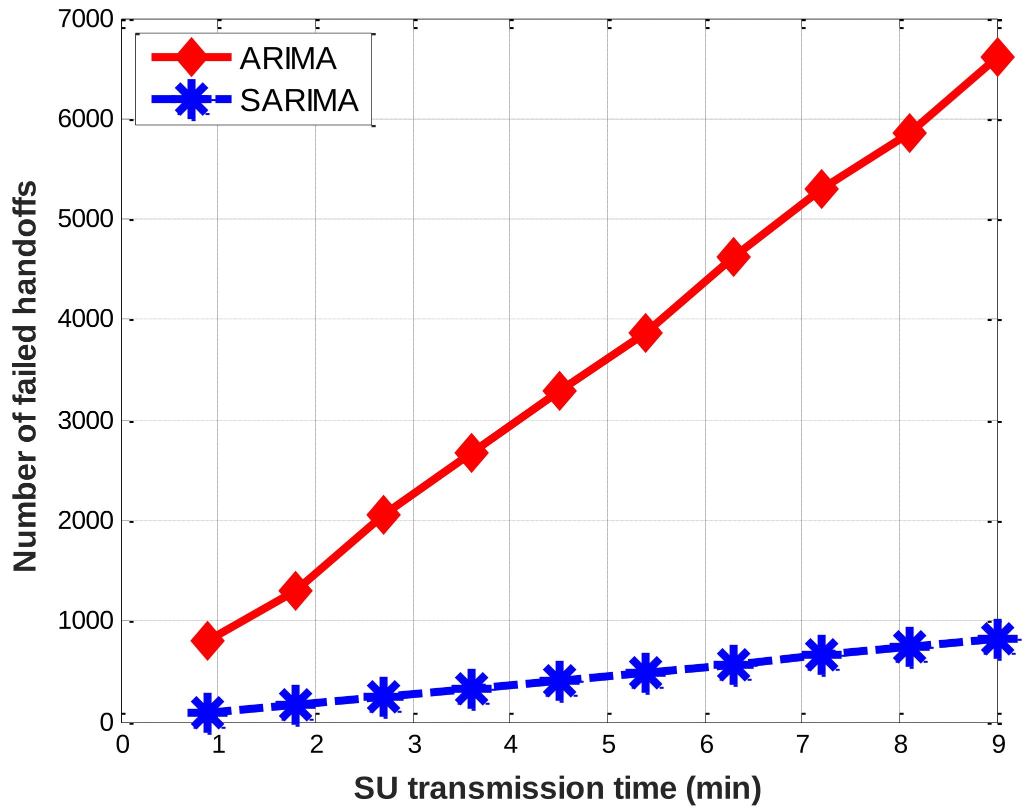

Figure 2. Number of failed handoff for ARIMA and SARIMA

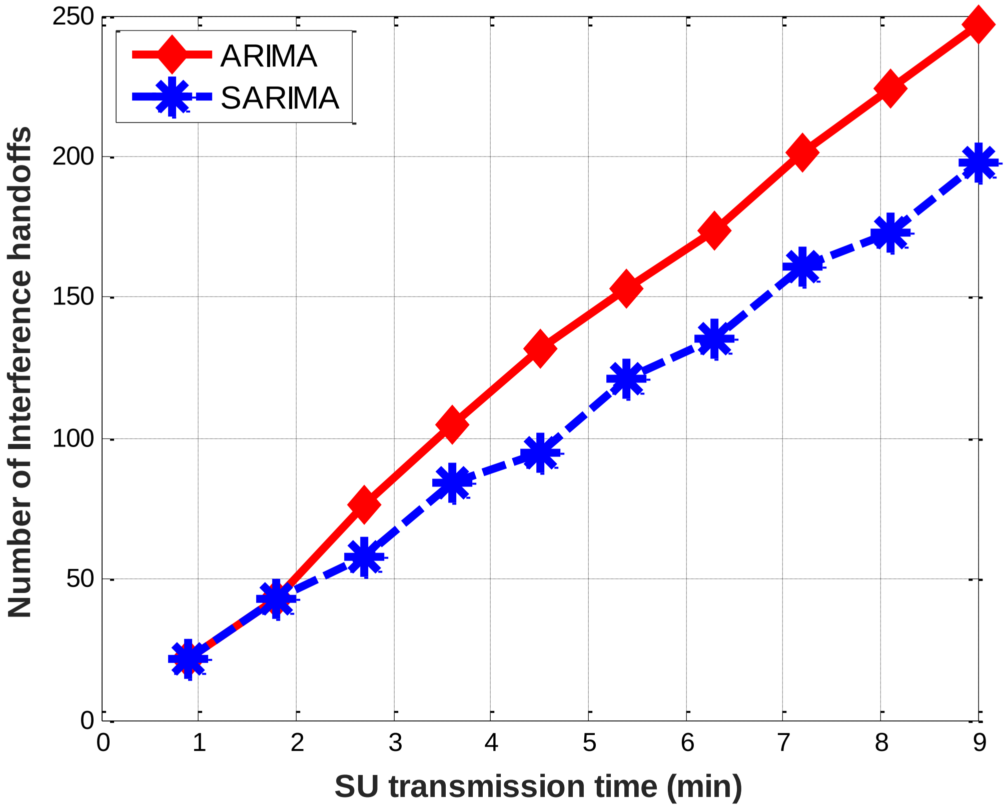

Figure 3. Number of interference handoff for ARIMA and SARIMA

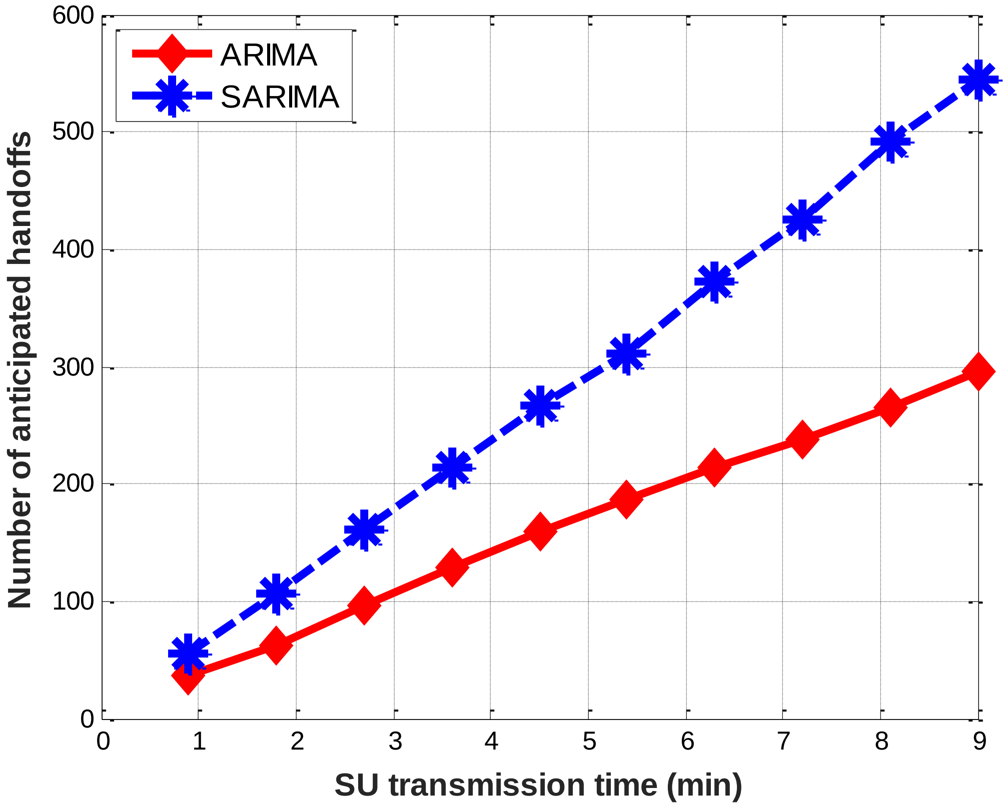

Figure 4. Number of anticipated handoff for ARIMA and SARIMA

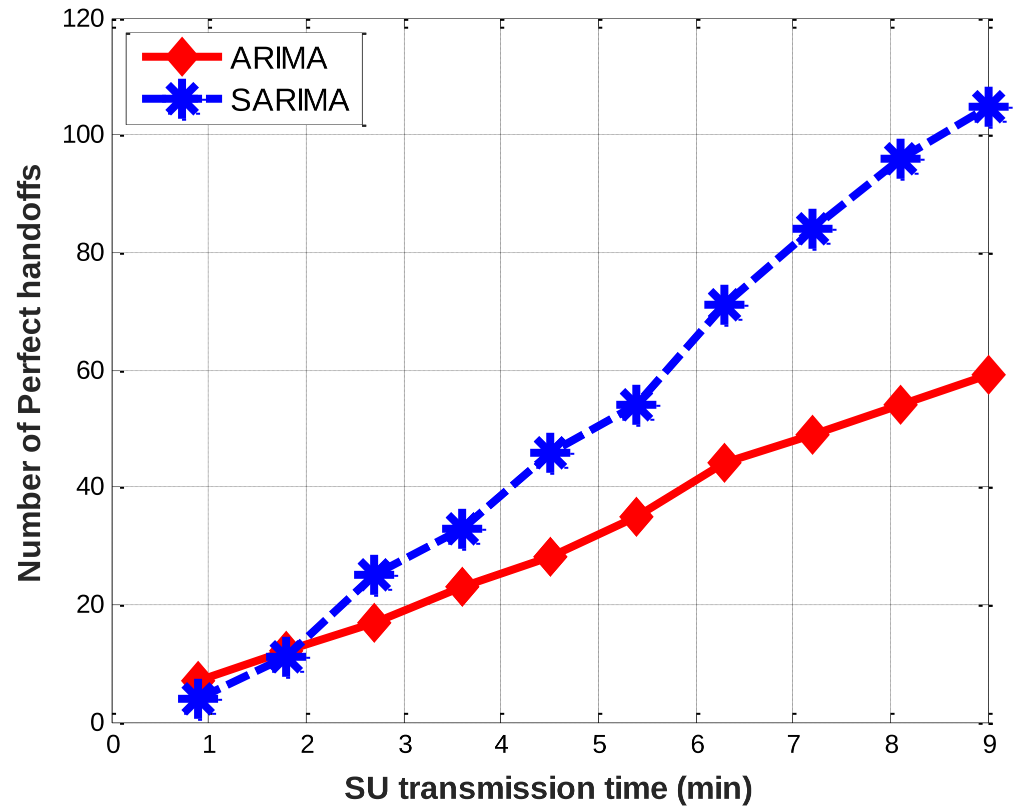

Figure 5. Number of perfect handoff for ARIMA and SARIMA

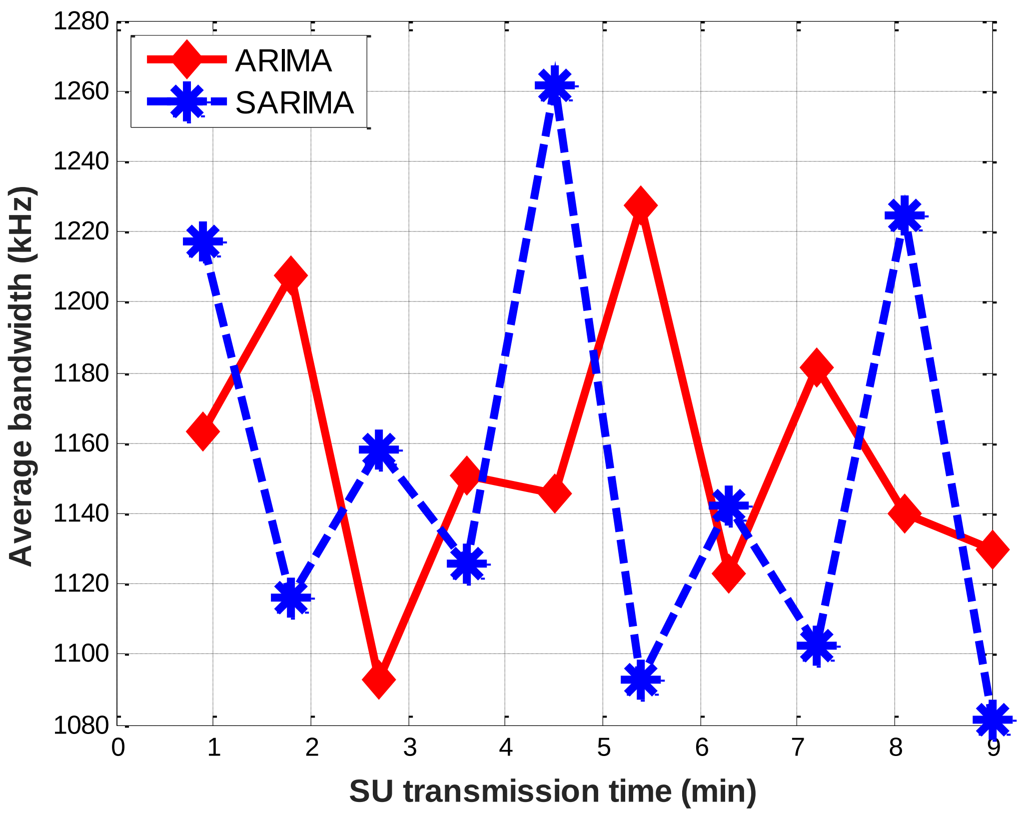

Figure 6. Average bandwidth for ARIMA and SARIMA

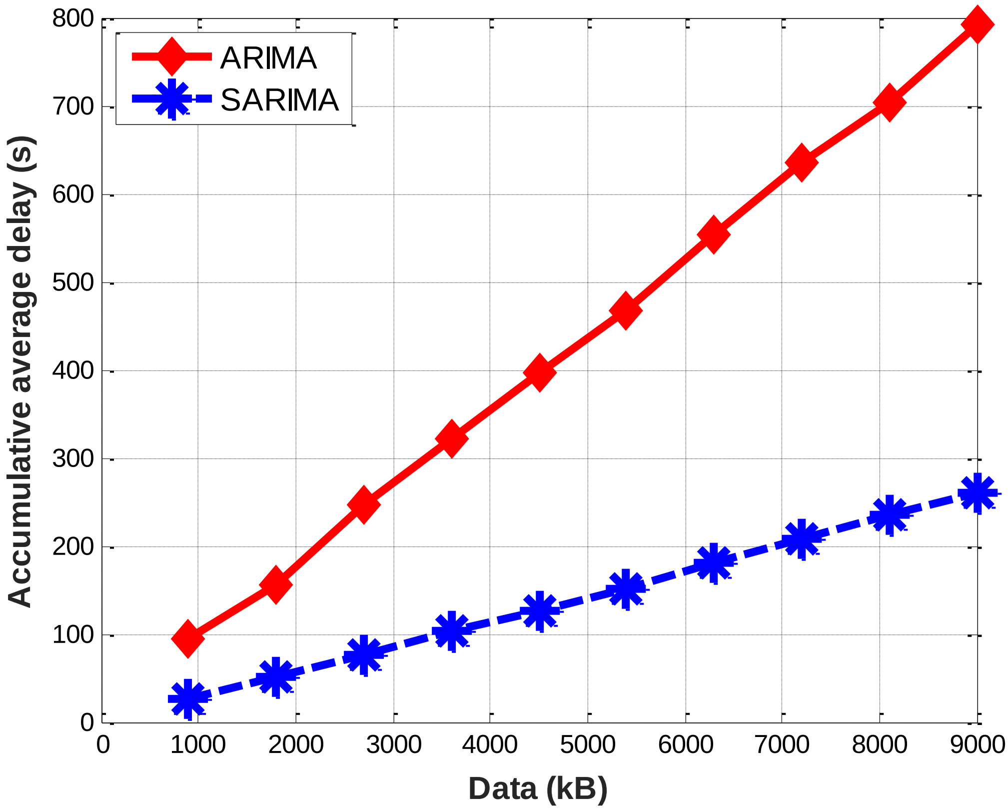

Figure 7. Average delay for ARIMA y SARIMA

CONCLUSION

The ARIMA and SARIMA algorithms have been evaluated in this paper in order to forecast the spectrum occupancy of the primary user in a Wi-Fi band. The SARIMA algorithm has better performance and is more convenient for a cognitive radio network, because it has higher precisions with respect to availability and occupancy times, with which the use of spectrum efficiency is improved and the interference level and collisions between PUs and SUs will be reduced.

ACKNOWLEDGMENT

The authors wish to thank Universidad Distrital Francisco José de Caldas for support during the course of this research work.

REFERENCES

[1] L. F. Pedraza, F. Forero, and I. Paez, "Evaluación de ocupación del espectro radioeléctrico en Bogotá-Colombia," *Ing. y Cienc.*, vol. 10, no. 19, pp. 127–143, 2014.

[2] Federal Communications Commission, "Spectrum Policy Task Force," 2015.

[3] C. Hernández, I. Páez, and D. Giral, *Modelo adaptativo multivariable de handoff espectral para incrementar el desempeño en redes móviles de radio cognitiva*. Bogotá: Editorial UD, 2017.

[4] J. Mitola and G. Q. Maguire, "Cognitive radio: making software radios more personal," *IEEE Pers. Commun.*, vol. 6, no. 4, pp. 13–18, 1999.

[5] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, 2005.

[6] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *International symposium on information theory*, 1973, pp. 267–281.

[7] G. E. P. Box and D. R. Cox, "An analysis of transformations," *J. R. Stat. Soc.*, vol. 26, no. 2, pp. 211–252, 1964.

[8] V. M. Guerrero, *Análisis estadístico de series de tiempo económicas*, Segunda Ed. México: Thomson, 2003.

[9] P. J. Brockwell, "On continuous-time ARMA processes," in *Handbook of statistics*, Amsterdam: Elsevier, 2001, pp. 249–276.

[10] M. W. Garrett and W. Willinger, "Analysis, Modeling and Generation of Self-Similar VBR Video Traffic," in *ACM Sigcomm*, 1994, pp. 269–280.

[11] L. F. Pedraza, C. Hernandez, and C. Salgado, *Modelo de predicción de la ocupación espectral para el análisis y diseño de redes de radio cognitiva*, Primera.

Bogotá: Universidad Distrital Francisco José de Caldas, 2018.

[12] G. Box, G. Jenkins, and C. Reinsel, *Time Series Analysis: Forecasting and Control (4th ed)*. New Jersey: Wiley, 2008.

[13] C. Hernández, O. Salcedo, and L. F. Pedraza, "An ARIMA model for forecasting Wi-Fi data network traffic values," *Ing. e Investig.*, vol. 29, no. 2, pp. 65–69, 2009.

[14] L. F. Pedraza, C. Hernández, K. Galeano, E. Rodríguez-Colina, and I. P. Páez, *Ocupación espectral y modelo de radio cognitiva para Bogotá*, Primera. Bogotá: Editorial UD, 2016.