

Comparative Evaluation of Prediction Models for Forecasting Spectral Opportunities

Cesar Hernandez^{#1}, Hans Marquez^{*2}, Diego Giral^{#3}

[#] Technical faculty, Universidad Distrital Francisco José de Caldas
Cl. 68D Bis A Sur N° 49F – 70, Bogotá, Colombia

¹ cahernandezs@udistrital.edu.co

² hrmarquezz@correo.udistrital.edu.co

³ dagiralr@correo.udistrital.edu.co

Abstract— Predicting the behavior of the primary user in cognitive radio 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: AR, MA and ARMA that can predict the behavior of the primary user as well as the spectral opportunities for cognitive radio networks in the GSM frequency band. The performance of the three models for time series will be contrasted with a purely reactive model (non-predictive) under two scenarios, two traffic levels and six evaluation metrics. The results obtained show that the moving average model has the best performance in general. However, it is not the best in all four testing scenarios.

Keyword - cognitive radio networks, handoff, prediction, spectrum opportunities, time series

I. INTRODUCTION

The great diversity of wireless networks and the high demand for their implementation, especially in the mobile communications field, along with the fixed assignment policies of the radio-electric spectrum have reduced significantly the amount of frequency bands available for their licensing [1]. However, some studies [2]–[4] have shown that most of the licensed bands within the radio-electric spectrum are underused in the time and space domain, which results in spectral opportunities (SO) which are available frequency channels that are not being used by users licensed in the time-space domain. The use of the spectrum is carried out mainly around certain bands while a considerable amount of the spectrum is underused. The federal communications commission (FCC) has informed of temporal and geographic variations in the use of the spectrum in a 15 to 85% range [3].

With the purpose of achieving a more efficient use of the spectrum, cognitive radio (CR) technology proposes a dynamic spectrum assignment (DSA). 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 [5], [6].

The purpose of this article is to present a comparative evaluation of three models based on time series: AR, MA and ARMA in order to predict spectral opportunities for cognitive radio networks in the GSM frequency band. The performance of the three models will be contrasted later on with a purely reactive model (non-predictive). In the performance evaluation two types of applications were considered: Real-Time (RT) and Best-Effort (BE), two levels of traffic: High Traffic (HT) and Low Traffic (LT) and six evaluation metrics (EM): number of Average Accumulated Handoffs (AAH), number of Average Accumulated Failed Handoffs (AAFH), number of Average Accumulated Handoffs without interference (AAPH), number of Average Accumulated Interfered Handoffs (AAIH), number of Average Accumulated Perfect Handoffs (AAEH) and number of Average Accumulated Anticipated Handoffs (AAUH).

After reviewing the current literature, it is evidenced that most proactive spectral handoff models are not pure since they do not comply with the requirement of temporary prediction of the PU except for the two that are now presented.

The authors in [7], propose a transmission strategy for CRN with several SO based on a predictive SH through dynamic programming. In this method, the SU predicts the future state of the spectrum and decides whether it is maintained inactive, kept in the current SO or changes to a new one to continue its transmission. The results show that the proposed scheme achieves a high efficiency in data transmission in comparison to the traditional schemes of always waiting and always changing.

The authors in [8] propose a model with diverse variables to perform the prediction of the channel including a prediction model for the behavior of the PU to avoid interference and a multi-user model to control the collision between SU. Collision control between SU is one of the most complicated areas when performing assertive models due to their random behavior. To solve this [8], a Common Hopping coordination scheme is proposed for the SH protocol's design; in this case, all SU are synchronized to the Hop through channels with the same Sequence Hopping. To perform the detection of the spectrum, it is assumed that the cognitive device has two antennas, one for transmission and control and another one exclusively dedicated to the detection of the spectrum. The results show that the proactive strategy is efficient when the load of the PU is low, reducing the number of handoffs and collisions; however, if the demand is high on behalf of the PU, the collision control is maintained but the number of handoffs increases.

II. MODELS BASED ON TIME SERIES

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 [9]–[11].

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).

A. AR Model

This 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 [12].

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.

$$Z_t = \sum_{j=1}^{\infty} \phi_j Z_{t-j} + a_t \quad (1)$$

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 $\sum_{j=1}^{\infty} \phi_j < 1$ [13]. 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.

B. MA Model

This 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 [12].

The MA model is described in equation (2) where θ_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.

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

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.

C. ARMA Model

This 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).

$$Z_t = \mu + \sum_{j=0}^{\infty} \theta_j a_{t-j} + \sum_{j=1}^{\infty} \phi_j Z_{t-j} + a_t \quad (3)$$

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 [12].

III. EVALUATION METHODOLOGY

To evaluate the performance of the proactive predictive algorithms: PPTS, also proposed in this investigation. The six evaluation metrics (EM) are described in Table 1.

TABLE I. EM Used for the Evaluation of the PPTS Algorithm.

Initials	Name	Description	Type of EM
AAH	Number of average accumulated handoffs	It corresponds to the total handoffs during the 10-minute transmission.	Cost
AAFH	Number of average accumulated failed handoffs	It is the number of Handoffs that the SU could not materialize because he found the respective targeted SO occupied.	Cost
AAPH	Number of average accumulated handoffs without interference	It is the total number of predictive handoffs carried out before the arrival of the PU, during the 10 minutes of transmission of the SU.	Benefit-Cost
AAIH	Number of average accumulated 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
AAEH	Number of average accumulated 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
AAUH	Number of average accumulated 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

To perform a fair comparative evaluation, each absolute value of the AAPH, AAIH, AAEH and AAUH metrics was divided by the absolute value of the AAH corresponding to each evaluation scenario, i.e., the respective values were taken with respect to AAH.

AAPH represents the SH carried out before the PU's arrival; while the AAIH represents the SH carried out after his arrival; therefore $AAH = AAPH + AAIH$. The EM APPH is dual: it is of the benefit type since it is desirable that the SH is performed before the PU's arrival to avoid the interference between the PU and the SU and it is of the cost type when the prediction is imprecise and is performed in too much anticipation before the PU's arrival, causing an increase in the AAH.

Due to the previous statements, it was decided to create the AAEH and AAUH evaluation metrics which are subsets of AAPH. AAEH represents the SH that are performed very closely to the PU's arrival; however, before this metric, the perfect SH are considered since they optimize the SO available time that they are using. AAUH represents the SH that are performed way before the PU's arrival which produces an increase in the AAH.

To determine whether a prediction is classified as AAEH it is verified that the SH was performed afterwards at 80% of the SO availability time in that moment; in the AAUH case, it is verified that the SH is performed 20% before the SO availability time in that moment. This does not only imply that the AAPH is not equal to the sum of AAEH and AAUH but also that there is a number of intermediate SH which are performed between 20% and 80% of the availability time of the SO; this can be calculated as $AAPH - AAEH - AAUH$.

With the purpose of facilitating the comparative analysis for each algorithm, the relative values (in percentage) were calculated for each EM. For the benefit-type metrics, the relative value (Rel) of the algorithm i was calculated from the absolute value (Abs) and the maximum value (Max) of the EM as described in equation (4). For the cost-type metrics, the relative value (Rel) of the algorithm i was calculated from the absolute value (Abs) and the minimum value (Min) of the EM as described in equation (5).

$$X_i^{Rel} = \frac{X_i^{Abs}}{X_{Max}^{Abs}} \times 100\% \quad (4)$$

$$X_i^{Rel} = \frac{X_i^{Min}}{X_i^{Abs}} \times 100\% \quad (5)$$

To calculate the global scores, the ponderation on equation (6) was used:

$$PG_i = 0.02 \times AAFH_i + 0.18 \times AAPH_i + 0.2 \times (AAH_i + AAIH_i + AAEH_i + AAUH_i) \quad (6)$$

IV. SIMULATION

In order to assess the performance of each developed VHDA, a simulation environment progressively reconstructs the behavior of the spectrum occupancy with the use of the captured data traces in the frequency GSM band. These allows to accurately evaluate the behavior of the PUs and also, to assess and validate the performance of each VHDA. The spectral occupancy data corresponds to a week-long observation captured at Bogota City in Colombia [14]. The energy detection technique was used to determine the occupation or availability of each of the 124 channels of the analyzed GSM band, with a decision threshold for the power of 5 dBm above the noise power. To determine whether a frequency channel is busy or not, the proposed decision threshold is based on the average noise floor for the frequency band used. We consider the specifications of the GSM band, the standard configuration of the spectrum analyzer and the measurements to establish the noise floor and the guard level. The average noise floor is obtained via spectrum analyzer measurements. The guard level was fixed at +5 dBm above the noise floor, in order to minimize false alarms. Thus, the average noise floor is -113 dBm and the decision threshold is set to $-113 + 5 = -108$ dBm.

V. RESULTS

The aforementioned EM were calculated for the HT and LT trace as well as the RT and BE approaches, which led to four evaluation scenarios for each metric: GSM-RT-LT, GSM-RT-HT, GSM-BE-LT, GSM-BE-HT (see Fig. 1 to 6).

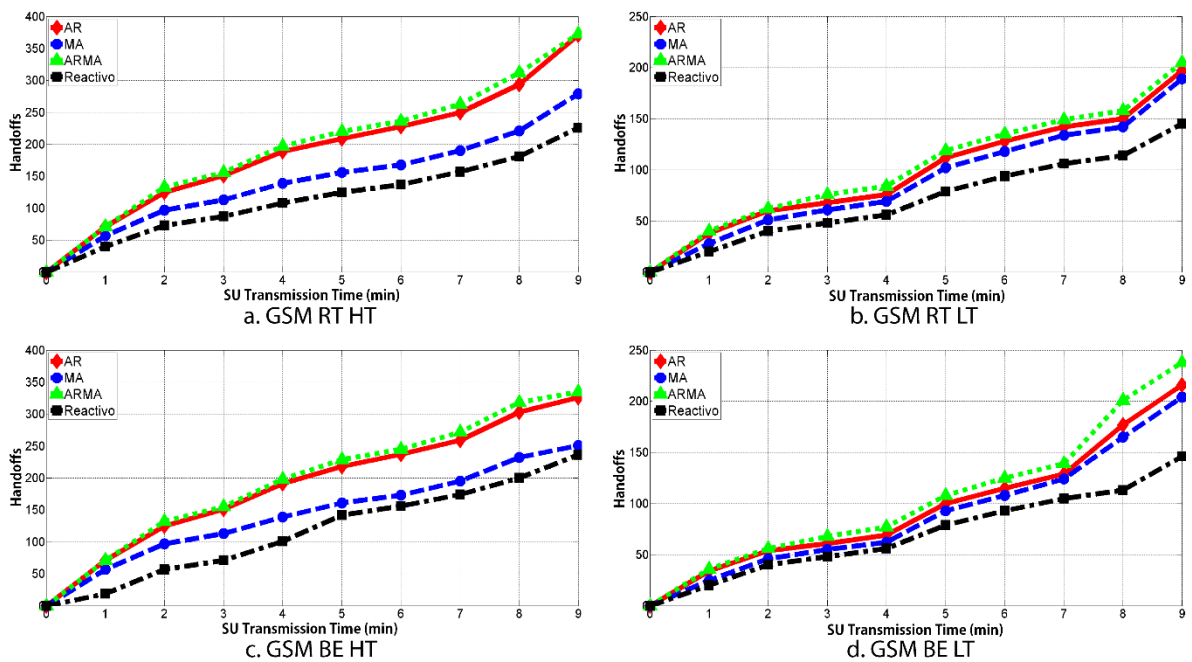


Fig. 1. AAH a. GSM RT HT, b. GSM RT LT, c. GSM BE HT, d. GSM BE LT

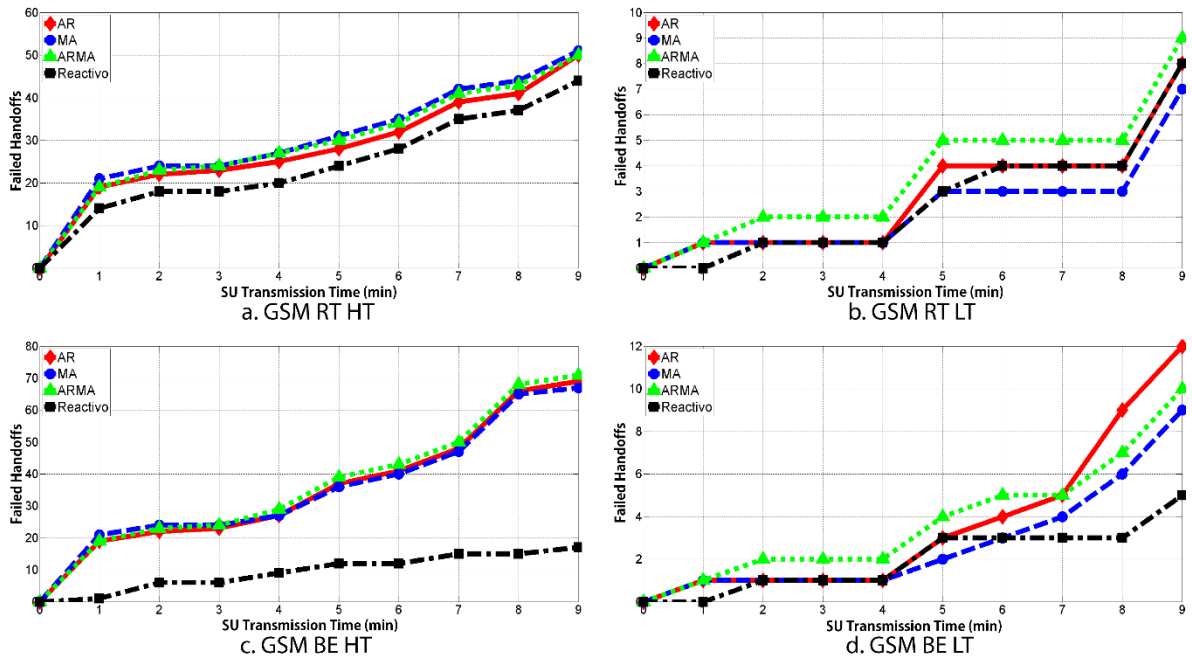


Fig. 2. AAFH a. GSM RT HT, b. GSM RT LT, c. GSM BE HT, d. GSM BE LT

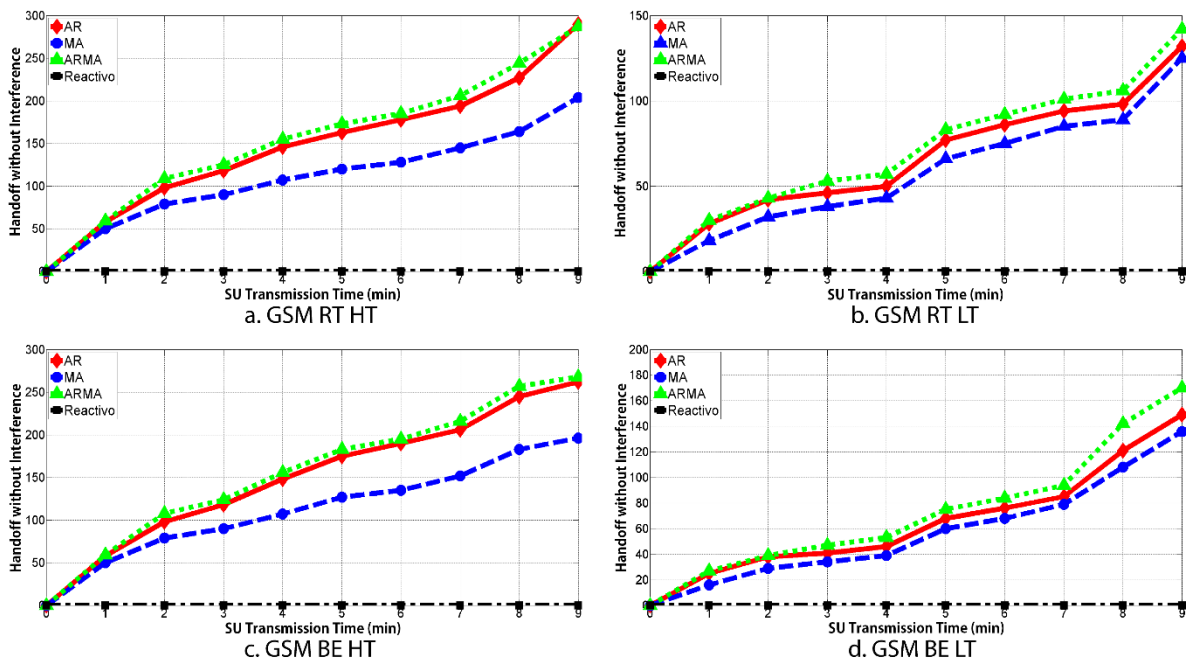


Fig. 3. AAPH a. GSM RT HT, b. GSM RT LT, c. GSM BE HT, d. GSM BE LT

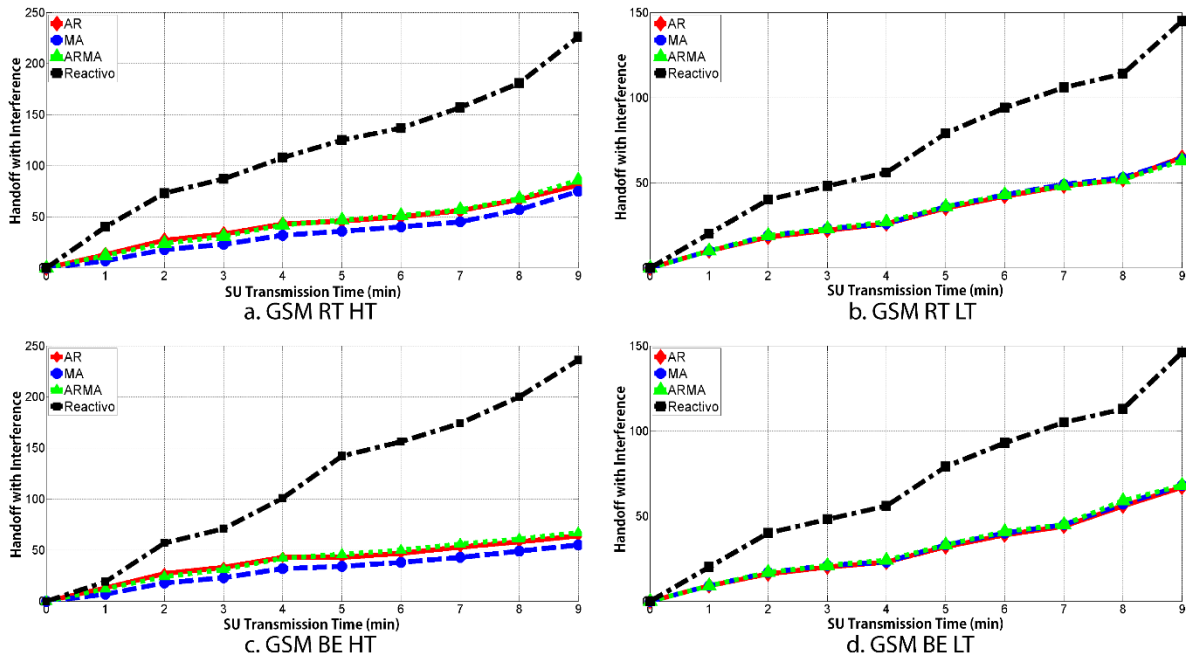


Fig. 4. AAIH a. GSM RT HT, b. GSM RT LT, c. GSM BE HT, d. GSM BE LT

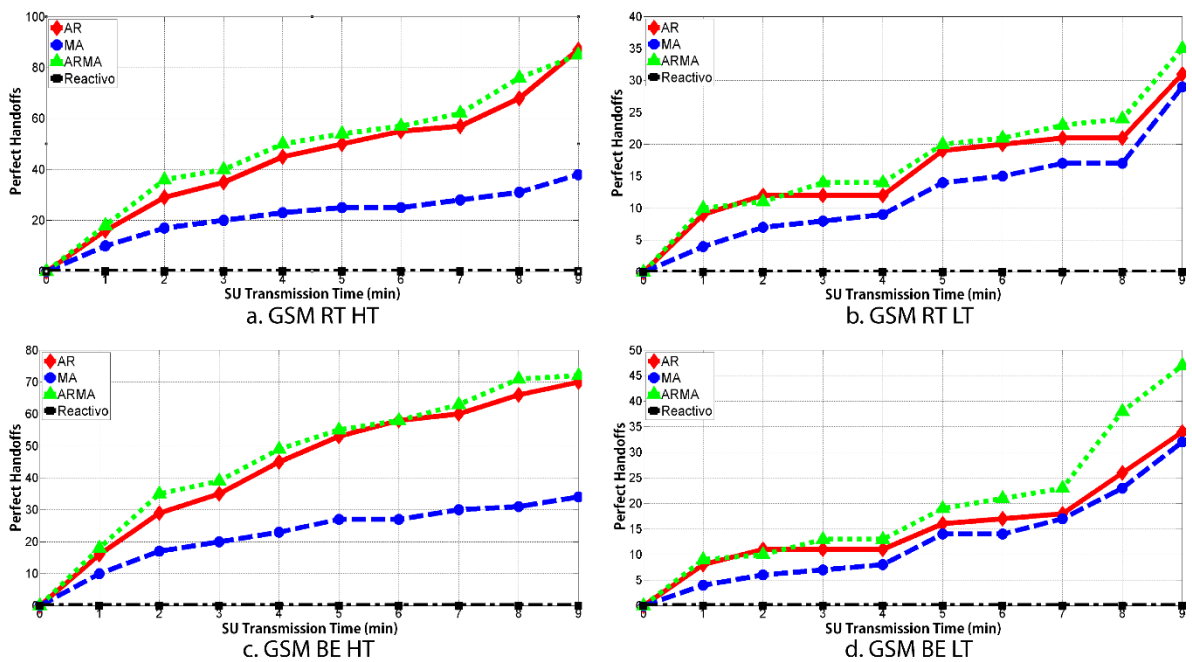


Fig. 5. AAEH a. GSM RT HT, b. GSM RT LT, c. GSM BE HT, d. GSM BE LT

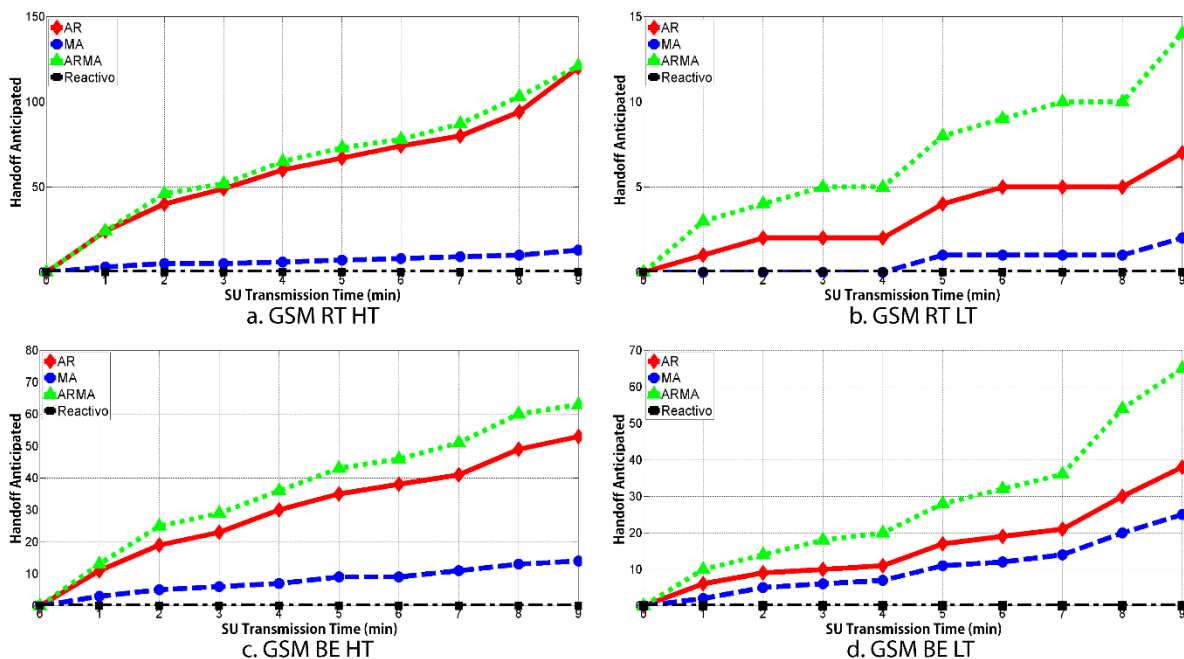


Fig. 6. AAUH a. GSM RT HT, b. GSM RT LT, c. GSM BE HT, d. GSM BE LT

Analyzing the performance of the SH predictive algorithm based on time series such as AR, MA and ARMA along with the reactive version, the following was observed: with respect to AAH, it is noted that the reactive model has the best performance followed by MA. With respect to AAFH, the reactive model has the best performance followed by MA. With respect to AAPH, the ARMA model has the best performance followed by AR. With respect to AAIH, the ARMA model has the best performance. With respect to AAUH, the reactive model has the best performance followed by MA.

When comparing globally each SH algorithm in the four scenarios defined in the methodology for the GSM network, the general global score shows that the MA model has the best performance with a 0.73% margin compared to the second. Therefore, it is interesting to analyze which algorithms are the best in each scenario: in the RT case in HT and BE in HT, the AR model has the best performance for RT in LT, the best model is MA and finally for BE in LT the ARMA model is the best one. If the results are averaged, it can be concluded that the AR model is the best one for HT with a 2.03% margin compared to the second and the MA model is the best in LT with a 5.6% margin compared to the second one.

VI. CONCLUSION

The most significant advantage of the prediction models is their capacity to reduce the level of interference; in the GSM network, the ARMA model has the best performance in this aspect with a margin of only 1,97% with respect to the AR model.

The spectrum assignment algorithms are the tools that give solution to the problem of using efficiently the radio-electric spectrum and contribute in different matters such as: channel characterization, local policies, user requirements, etc. The advantages and disadvantages in the adoption of one algorithm or the other for spectrum assignment are in function of the specific needs of its purpose; hence, its implementation depends on the needs in terms of signal processing, time responses, data availability, storage capacity, learning capacity, robustness, among other factors.

ACKNOWLEDGMENT

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

REFERENCES

- [1] Hernández, L. F. Pedraza, I. Páez, and E. Rodríguez-Colina, "Análisis de la Movilidad Espectral en Redes de Radio Cognitiva," *Inf. tecnológica*, vol. 26, no. 6, pp. 169–186, 2015.
- [2] D. Cabric, S. M. Mishra, and R. W. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," in *Conference on Signals, Systems and Computers*, 2004, vol. 1, pp. 772–776.
- [3] Federal Communications Commission, "Spectrum Policy Task Force," 2015.
- [4] V. Valenta, R. Maršálek, G. Baudoin, M. Villegas, M. Suarez, and F. Robert, "Survey on spectrum utilization in Europe: Measurements, analyses and observations," in *International Conference on Cognitive Radio Oriented Wireless Networks*, 2010, no. 230126, pp. 2–6.
- [5] J. Mitola and G. Q. Maguire, "Cognitive radio: making software radios more personal," *IEEE Pers. Commun.*, vol. 6, no. 4, pp. 13–18, 1999.
- [6] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, 2005.
- [7] Y. Wu, Q. Yang, X. Liu, and K. Kwak, "Delay-Constrained Optimal Transmission with Proactive Spectrum Handoff in Cognitive Radio Networks," *IEEE Trans. Commun.*, 2016.
- [8] Y. Song and J. Xie, "Proactive Spectrum Handoff in Cognitive Radio Ad Hoc Networks based on Common Hopping Coordination," in *INFOCOM*, 2010, pp. 1–2.
- [9] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *International symposium on information theory*, 1973, pp. 267–281.
- [10] G. E. P. Box and D. R. Cox, "An analysis of transformations," *J. R. Stat. Soc.*, vol. 26, no. 2, pp. 211–252, 1964.
- [11] V. M. Guerrero, *Análisis estadístico de series de tiempo económicas*, Segunda Ed. México: Thomson, 2003.
- [12] P. J. Brockwell, "On continuous-time ARMA processes," in *Handbook of statistics*, Amsterdam: Elsevier, 2001, pp. 249–276.
- [13] M. W. Garrett and W. Willinger, "Analysis, Modeling and Generation of Self-Similar VBR Video Traffic," in *ACM Sigcomm*, 1994, pp. 269–280.
- [14] L. F. Pedraza, C. Hernández, K. Galeano, E. Rodríguez-Colina, and I. Páez, *Ocupación espectral y modelo de radio cognitiva para Bogotá*, Primera. Bogotá: Universidad Distrital Francisco José de Caldas, 2016.

AUTHOR PROFILE

Cesar Augusto Hernandez Suarez Electronic Engineer, Master in Information Science and Communications, Doctor of Philosophy in Systems and Computer Engineering of the Universidad Nacional de Colombia. Professor and researcher at Universidad Distrital Francisco José de Caldas, Bogotá, Colombia.

Hans Raul Marquez Ramos Electronic Engineer, Master in Information Science and Communications. Researcher at Universidad Distrital Francisco José de Caldas, Bogotá, Colombia.

Diego Armando Giral Ramirez Electrical Engineer, Master in electric engineering. Professor at Universidad Distrital Francisco José de Caldas, Bogotá, Colombia.