Rating Prediction in a Platform IPTV through an ARIMA Model

Ricardo Ferro [#1], Cesar Hernández [#2], Gabriel Puerta [*3]

[#] Master of Science in Information and Communications,
Universidad Distrital Francisco José de Caldas, Carrera 7 No. 40B – 53, Bogotá DC-Colombia
[1] raferrob@udistrital.edu.co
[2] cahernandezs@udistrital.edu.co
[*] Engineering Department, Universidad Minuto de Dios
Calle 81B No. 72Bis - 65, Bogotá DC-Colombia
[3] gpuerta@uniminuto.edu

Abstract— This document describes the architecture and technology of an IPTV network and, present the direct causes that generate considerable in the delay time of channel switching to user perception. Also some techniques that have been developed to mitigate the delay in the zapping are listed. This work propose developing an ARIMA time series model to model and predict the behaviour of the rating in the channels with more audience on an IPTV platform. The validation of the proposed model is performed from real data of the rating provided by UNE Telecommunications. The results show a low percentage of error in the prediction of the rating for one week and mitigation of Channel Zapping Delay in 68% approximately.

Keyword— ARIMA, channel zapping delay, forecasting, IPTV, multicast, network, rating

I. INTRODUCTION

Providers of telecommunications services in countries such as Colombia have developed technological strategies that guarantee the end user a better quality of service in video applications such as IPTV, substantially making use of network resources. Nonetheless, due to various factors such as the complexity of the video, the heterogeneity of different access technologies, among others, it is not possible to ensure total quality in terms of user experience, as occurs in the delay change channel (Channel Zapping Delay) in an IPTV platform. [1].

In traditional analogue broadcast television and cable technology, channel change is almost instantaneous, as it only involves tuning the television receiver to a specific carrier frequency, demodulate the content and display it on the TV screen. The zapping delay in these systems is typically less than 200 ms. Hence viewers sometimes consider that zapping is virtually instantaneous, and become accustomed to this experience. With the digitization and compression of content zapping times have increased significantly. Users no longer experience this today in digital TV networks by cable, but is an even more serious in IPTV, and sometimes zapping is also affected by the delay of the network problem. The zapping IPTV users often experience a few seconds of delay or more. [1]-[4].

Currently there are different methods for evaluating television audience [5]-[7], in which a permanent sample of households participates in the survey. The main device used for this measurement is the people meter, which is connected to every TV panel shows, recording at certain intervals television consumption, taking into account turn on , channel changes and members of the household, throughout the day [8].

One of the IPTV services providers in Colombia is the telco UNE. As for the measurement of audience levels, UNE has implemented an application through the analysis of their devices access network determines the number of users associated with each channel at a given instant of time.

This paper identifies the main causes for the delay channel change, and analysis related to the indirect rating mitigate this parameter [7], [9]. To achieve the goal of mitigating the delay in changing the channel, this paper proposes the development of an ARIMA time series model to estimate and predict the behaviour of rating [7], [9] for the channels with the largest audience, through actual rating data provided by the UNE telecommunications company.

The time series models have been applied in several areas such as econometrics and traffic network [10]-[12], among others. Time series aim at developing statistical models to explain the behaviour of a random variable that varies over time allowing future forecasts estimate that random variable, assuming stable conditions and variations recorded to date, this allows planning and decision making in the short or long term [10].

II. CHANNEL ZAPPING DELAY

In [2], [3], [13] introduce a recent study that concludes that to achieve an acceptable quality of service, channel switching time has to be below 0.43 s when it occurs. Although the study is limited in terms of population size as a test subject, it is clear that the high zapping delay in IPTV degrades the quality of experience perceived by customers of IPTV services.

Within the IPTV transmission there are different systems that provide a degree of delay when performing the channel change. In Fig. 1 proposed in [2], you can watch every stage influencing Channel Zapping Delay: (1) request to change the channel, (2) IGMP router, (3) synchronization delay, (4) video buffer delay, and (5) processing delay in the Set-Top-Box (STB).

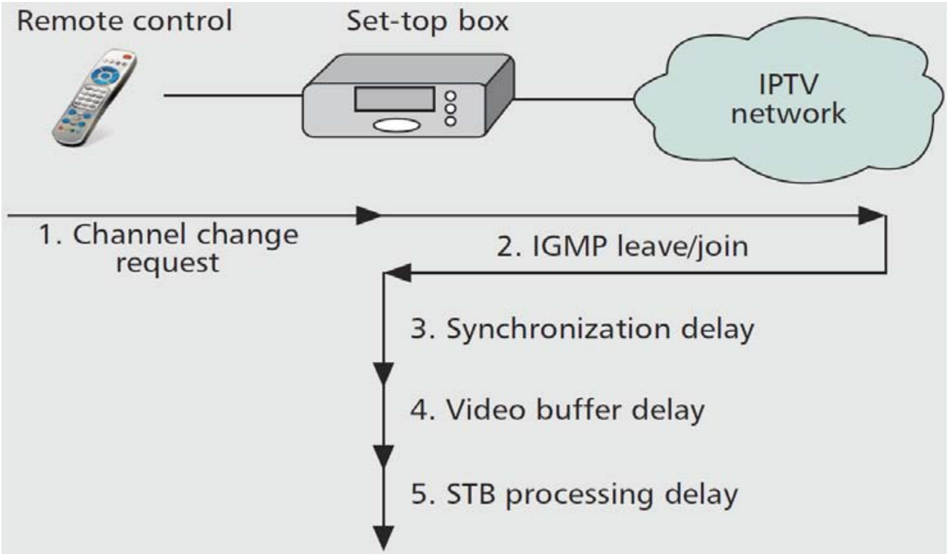

Fig. 1. Main components in the channel change delay. Taken from [2]

The user initiates the process by issuing a request to change channels with the remote control. The application reaches the STB after an estimated delay time of 10.5 ms. In IPTV video delivery, as opposed to typical cable networks, only the channel that the user is viewing is delivered to the STB at any time. This is due to bandwidth limitations in the access network.

When a user switches to a new channel, the STB has to issue a new request to the network channel (step 2). To minimize this delay all television channels are distributed close to the user, particularly to access digital subscriber line multiplexer (DSLAM in DSL networks) or the local router. In addition to the signalling delay, one has to add propagation delay experienced in the access link. The sum of these types of delay, referred to collectively as network delay is usually between 100 and 200 ms [14].

After the STB receives the first packets of the multicast group recently joined, there is still a time delay before one can start using audio-visual data, because the STB must wait until the next I frame before it can start decoding content. This is referred to as the synchronization delay generated (step 3). The maximum synchronization delay is equal to the length of GOP, which occurs when the STB is outside the start of an I frame, so it must wait for the next. Therefore, the synchronization delay by the delay video storage (step 4), required to compensate for lost packets and jitter up a significant portion of delay channel change, which is approximately in a range of 1 to 3 s. [2], [3].

The latest source of delay is the processing delay in the STB and the display device (step 5). This can occur in a number of layers of the system, and is generally a trade-off between terminal resources (memory and processor speed) and cost. The processing time depends largely on the STB, the approximate time in this process is 150 ms. Fig. 2 graphically summarizes the contribution of each component in the total delay IPTV channel change. The main contributors are timing delay and delay stream video storage delay. (Steps 3 and 4). [2], [3].

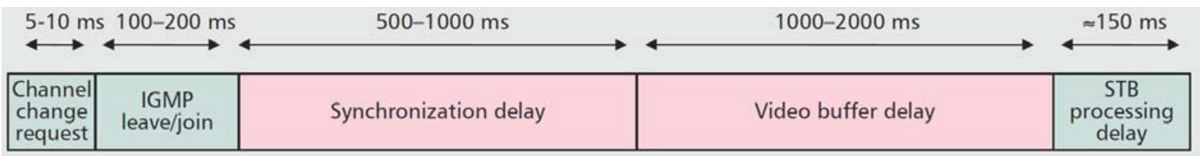

Fig. 2. Contribution of each component (unscaled). Taken from [2]

III. IPTV ARCHITECTURE

For the development of this research it was selected the IPTV's platform of UNE was selected. UNE has chosen IPTV to maximize its copper network and offer TV services in areas where the company has no hybrid networks of fibre and coaxial (HFC).

Listed below are the required parameters of configuration required to determine the scope of investigation and to establish the number of channels that can be preloaded from the head of a single Transport Stream using MPTS encapsulation, based on implemented access technology and bandwidth limitations.

The IPTV header implemented by UNE Company is supplied by Tandberg technology. The developed system has a 13 + 1 configuration, with 13 active and one spare encoders. Each chassis generates an approximately 20 Mbps traffic; likewise, the system is managed by a Cisco C4507 master switch. Each Gigabit port handles 8 stream multicast, from port 2 to port 14 and port 15 is inactive for the spare. Multicast routing is enabled on the switch through IGMP. The switch delivers the stream multicast to the edge router through a fibre link. The First Edge Router (FHR) is implemented in M320, which is a high density edge router, routing and service creation platform based on the Juniper Networks service; the router supports multiple services on a single platform. The router handles a throughput of 320 Gbps, and between the main router and the backup router a 10Gbps link is handled. This device is a layer 3 router which uses the OFSP and PIM dense mode protocol.

The Backbone that manages the IPTV network is supported on IP / MPLS. MPLS is based on the labelling of packages based on priority and / or quality (QoS) criteria. The key element in an MPLS network is the Label Switching Router (LSR), which are classified into core LSR, responsible for mastering a network and MPLS Label Edge Router (LER), which are responsible for interconnecting MPLS network with external networks both to the head, as well as towards the end user. [15].

Most telecom operators use DSL technology (Digital Subscriber Line) to deliver broadband services to households and IPTV in its access network. They can choose from various DSL technologies such as: ADSL, ADSL2, ADSL2 + and VDSL (DSL high speed). Access multiplexers to digital subscriber lines (DSLAM's) transfer the signals from optical to copper wires to deliver DSL to customers' homes. The DSLAM's are often installed within neighborhoods, it is for this reason that these IPTV systems are known with the name of edge fibre deployments or FTTC. [16]-[18].

UNE Telecommunications Company has made use of heterogeneous technologies and has implemented various topologies in the access network, making use of the DSL platform and existing cable networks. Thus UNE has worked on its access network with XDSL variants. Currently ADSL2 + technology is one of the most frequently implemented in the network, which manages download up to 24 Mbps theoretical, without disregarding that this may be attenuated by distance. Thus, and considering the bit rate for a standard TV channel (SD) is about 2.5 to 4 Mbps and an HD channel 7 to 8 Mbps, and leaving sufficient bandwidth resources for voice and data, without saturating the channel it has the option of managing 7-8 SD channels or 2 channels in HD.

Based on this, to develop the research the option to handle 4 SD channels will be taken, (assuming two televisions per household), which may be preloaded in the header implementing MPTS encapsulation.

IV. MEASURING RATING

For measuring the rating, a previous study of the different techniques developed was made. One of the most known is people meter which is implemented by the company TIME IBOPE worldwide, whose function is to offer this service tailored to the different needs of each audio-visual media provider. Because different TV companies around the world pay high costs for access to a study in the measurement of rating, it was difficult to access this information through this application. Because of this, UNE has designed an application to measure levels of rating in the EDATEL network in the north of the country.

The different measurement parameters for the development of research, based on the procedures and guidelines in place in this application will be mentioned. Similarly the number of channels and sample population to the respective measurement is established.

A. Application Developed by UNE to Measure Rating

The application developed by UNE involves the creation of a command, which will be loaded into the access devices (DSLAM).

The function of this command is to ask the DSLAM how many and which IP multicast are connected at that time and through which ports they emerge, emphasizing that each IP multicast represents a channel or service. Fig. 3 depicts the topological application design.

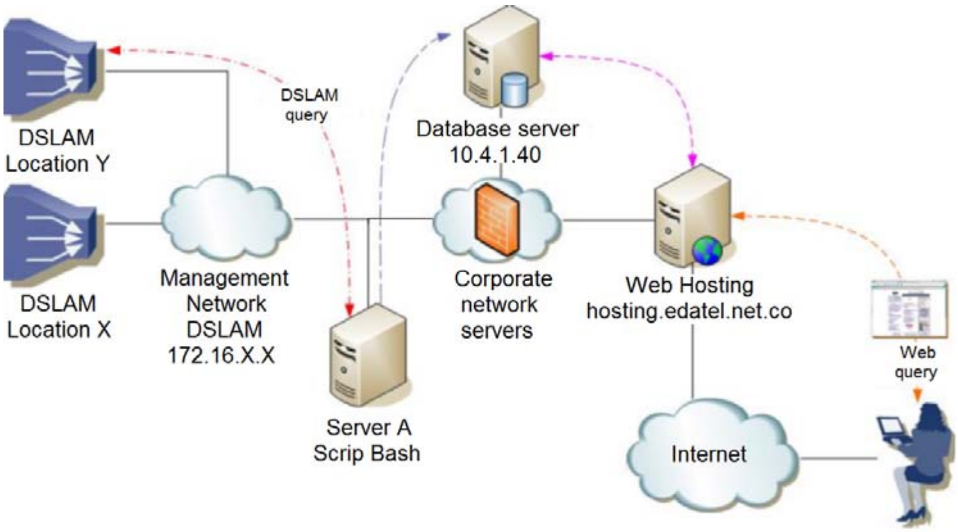

Fig. 3. Topological design for measuring UNE rating. Provided by UNE

B. Initial Measurement Parameters

The monitoring can be performed at intervals of 10 minutes (5 minutes if necessary). The information provided by the interface embraces the location, date, hour / minute, channel name and number of television sets watching each channel. Likewise, the application offers the possibility of making filters to display personalized information. As the aim of the research is to analyse the behaviour of rating over time for a given channel, the filtering tool will be used to take samples by date.

For rating measurement it has been taken a period of 16 weeks, every day. The measurement has been made in the fourth quarter of 2014, from September 1 through December 21, for a total of 112 samples (once daily). Subsequently, a weighted average is performed to determine the most watched channel during the 16 weeks. As a result of a basic statistical analysis, Table I shows the average audience of the most watched channels.

TABLE I. Audience of the Most Watched Channels

Date	Channel	No. TV Connected
Weighted	C1	1357
	C2	1263
	C3	467
	C4	409
	C5	280
	C6	267
	C7	96
	C8	90
	C9	81
	C10	75
	C11	72
	C12	63
	C13	49
	C14	42
	C15	36
	C16	29
	C17	24
	C18	16
	C19	11
	C20	9

For our case it is of interest to consider the four most-watched channels in order to analyse their behaviour over time. Table II shows a comparison between the four most-watched channels.

From Table II it can be seen that C1 and C2 channels handle more than 50% of the audience of the taken sample. Therefore, C1 and C2 would be the two initial channels that can be preloaded into the MPTS group. The application of the time series model on channels C1 and C2 would present no relevance, because regardless of how high the mean error was, it would never be below the average rating of the C5 and C6 channels. Therefore the application of the ARIMA time series model will be made on the C3 and C4 channels whose rating values are not far from the values of the C5 and C6 channels.

TABLE II. Comparison of the Four Channels with Wider Audience

Channel	Users Online	Relative frequency	%
C1	1357	0.2714	27.14
C2	1263	0.2526	25.26
C3	467	0.0934	9.34
C4	409	0.0818	8.18
MPTS Group	3496	0.6992	69.92

V. RATING MODELLING

The primary objective of rating time series modelling is to forecast its future values for the next 7 days, and validate them with real values, obtained from an UNE application. Initially it is extracted the time series independently for channel 3 and channel 4, of the 112 samples of each series, and the last 7 are withdrawn to perform the validation process, and the model is developed with the remaining 105 samples. The development of the model in detail will be made for C3 channel, and this procedure is similar for channel C4.

A. Model Selection

After making a statistical analysis of time series of the rating, through the autocorrelation function (FAC) and partial autocorrelation (FACP), it was determined that there is a high degree of correlation between the taken samples; the incidence of correlation implies that the variability extends to various time scales, with this ruling out the validity of traditional uncorrelated models. In addition to being correlated, time series happen to be the most appropriate to model the behaviour of seasonal time series random variables, which have a cyclical behaviour, as in the case of rating and s make a short-range forecasting (and some sometimes long range).

Thereafter, it was determined through the covariance analysis, that the mean and variance of the time series of rating of channels C3 and C4 are not constant, that is, that these series are not stationary. As the time series are not stationary it is required to transform the series and apply an integrated process, therefore, the time series model that was selected is ARIMA (p, d, q). To develop this model the Box-Jenkins methodology was implemented.

B. Model Identification

The time series of the rating for the defined sample space, was obtained from the relative frequencies of each sample. As mentioned above this series is composed of the first 105 samples, not the total 112 because the last seven were reserved for the validation process. Fig. 4 shows the original data of the channel C3 obtained for a sample space of 15 weeks (105 samples).

The next step is to determine the initial values of the ARIMA model order, the order of the autoregressive process (AR) p, the order of the process of moving averages (MA) q, and the order of the transformation of the series d.

The first step is to transform the series into a stationary series. A time series is stationary when the mean is zero and variance is constant, in Fig. 4 can easily determine that the original series is not stationary. To convert a stationary time series is required to transform the series through a process of differentiation as often as necessary until the series meets the requirements of stationary or through the logarithm function, or a combination thereof. The number of times necessary to transform the series, determines the value of the parameter d. After applying the logarithm function to the original time series and differentiate once the series shown in Fig. 5 is obtained, which is stationary, obtained an ARIMA (p, 1, q). Stationarity is analytically determined through a unit root test. [10].

To determine the order of p (AR), it analysed the FACP, the remnants of stationary time series that exceed the threshold are candidates to join the model. To determine the order of q (MA), a similar but with FAC analysis is performed. [10]. According to the analysis of the FAC and FACP, the initial identification of the model is ARIMA (13,1,19).

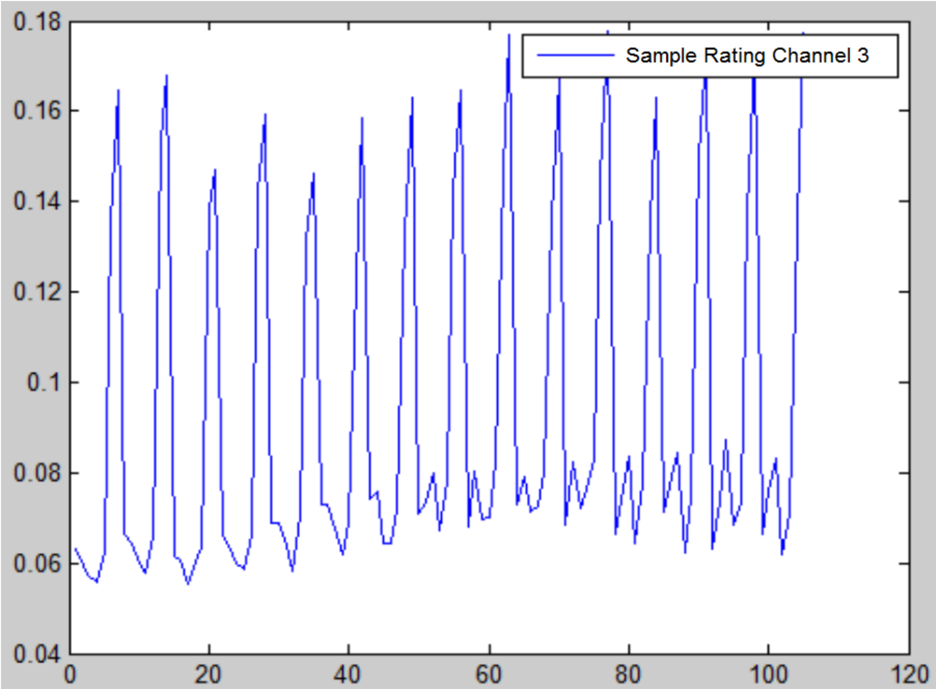

Fig. 4. Time series rating channel 3

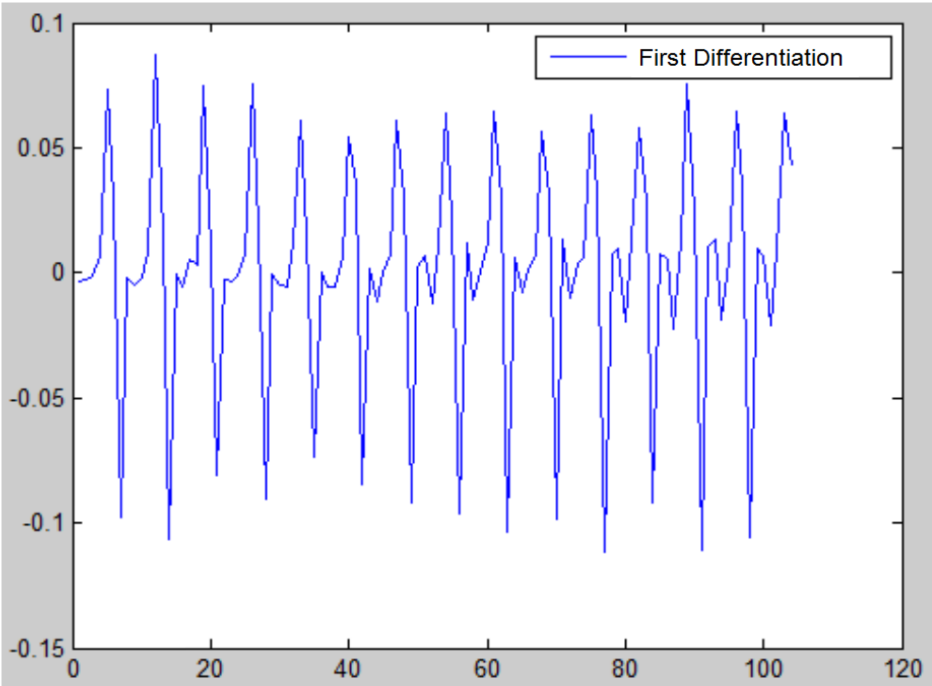

Fig. 5. Differentiation of logarithmic time series C3 channel

C. Estimation and Model Validation

Once the model is identified the next step is to estimate its terms and validate the model estimation. For estimating the terms of the ARIMA model (13,1,19) it was used the STATA v12.0 software. The validation of model estimation is performed through the analysis of FAC and FACP functions of residuals of the model, i.e. the difference values between the transformed time series and those obtained with the estimated model. If a lag of residuals has correlation (exceeds the threshold value), that lag must be included in the model and make a new iteration (estimate). This dynamic re-specification cycle ends when residuals no longer present correlations, that is, they are white noise. [10]. Fig. 6 describes the correlation analysis of residuals on the time series for channel C3.

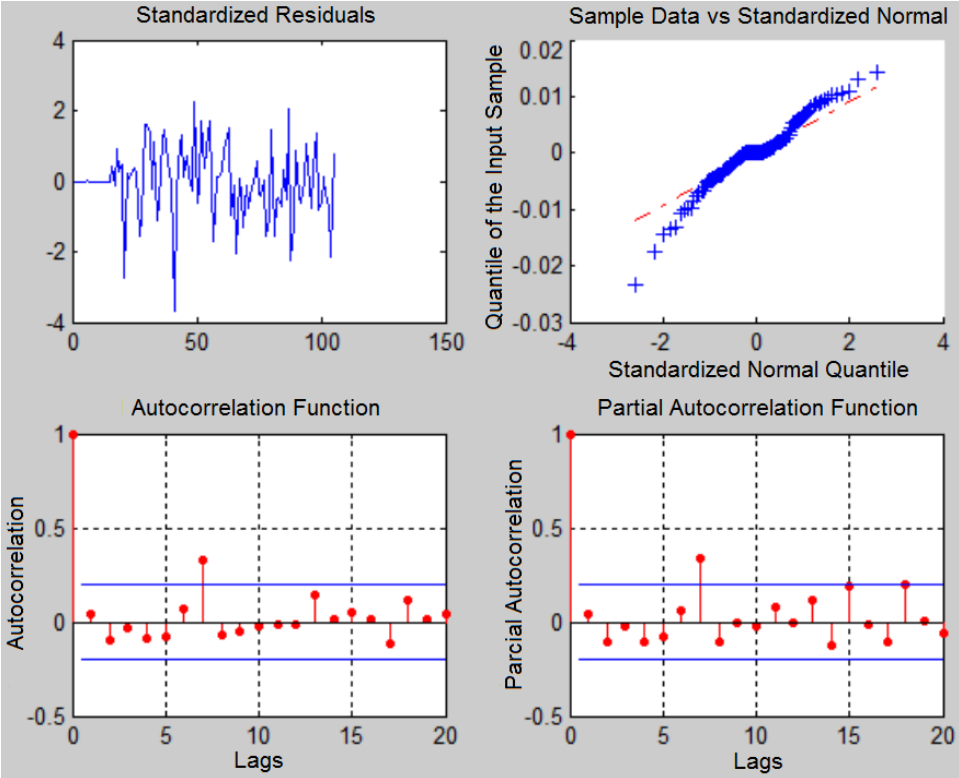

Fig. 6. Residual analysis of channel C3 model

In Fig. 6 it can be established that there is still correlation at lag number 7 and therefore it is necessary to iterate and estimate the coefficients of the model again, including that lag.

After performing 14 iterations, the final estimation of ARIMA (0, 1, and 14) model described by equation (1) is obtained.

$$z_t = \varepsilon_t - 0{,}8872\varepsilon_{t-1} - 0{,}1762\varepsilon_{t-14} \quad (1)$$

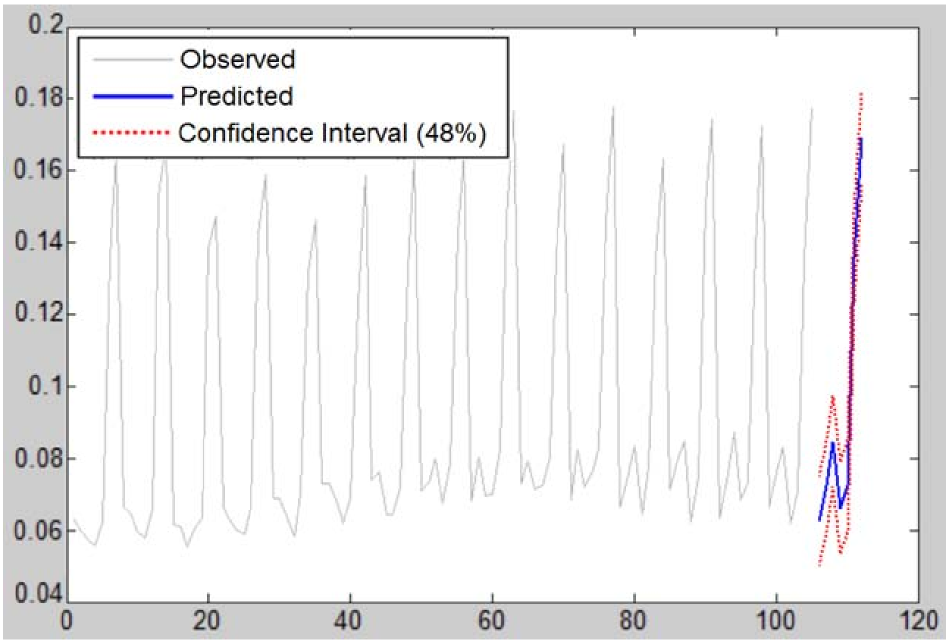

Fig. 7. Prediction of channel C3 rating

D. Prediction Model

The prediction of the time series was performed for the last 7 reserved samples (1 week rating). Fig. 7 shows the predicted data compared to the original time series within a confidence interval of 48%.

E. ARIMA Model for the Channel C4

ARIMA model development (p,d,q) to the channel C4 follows the same procedure described for channel C3. Fig. 8 shows the time series of original channel rating of C4.

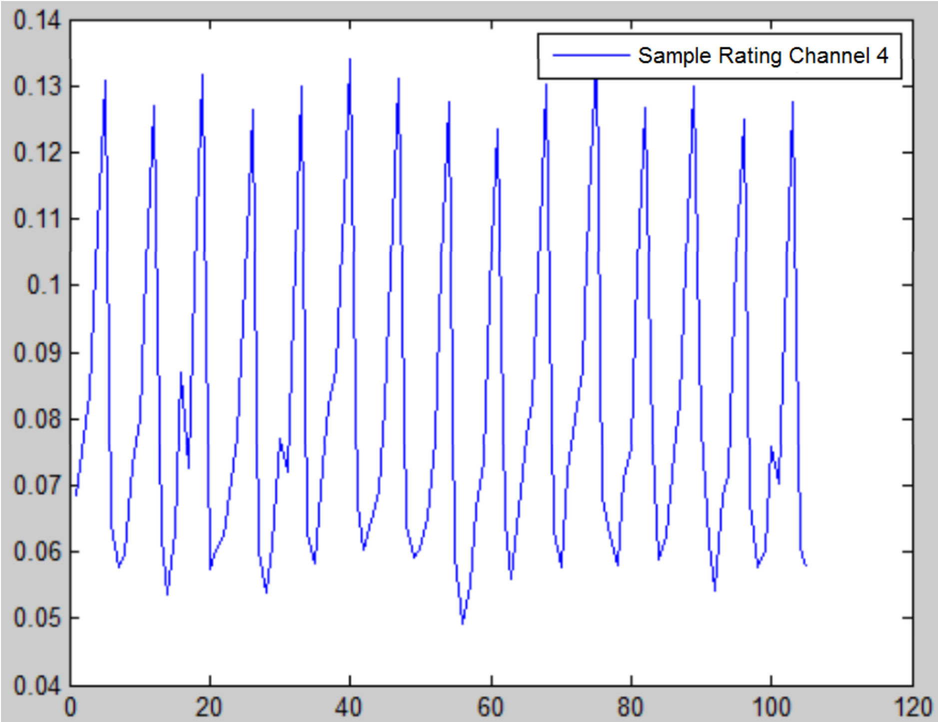

Fig. 8. Time series for channel C4 rating

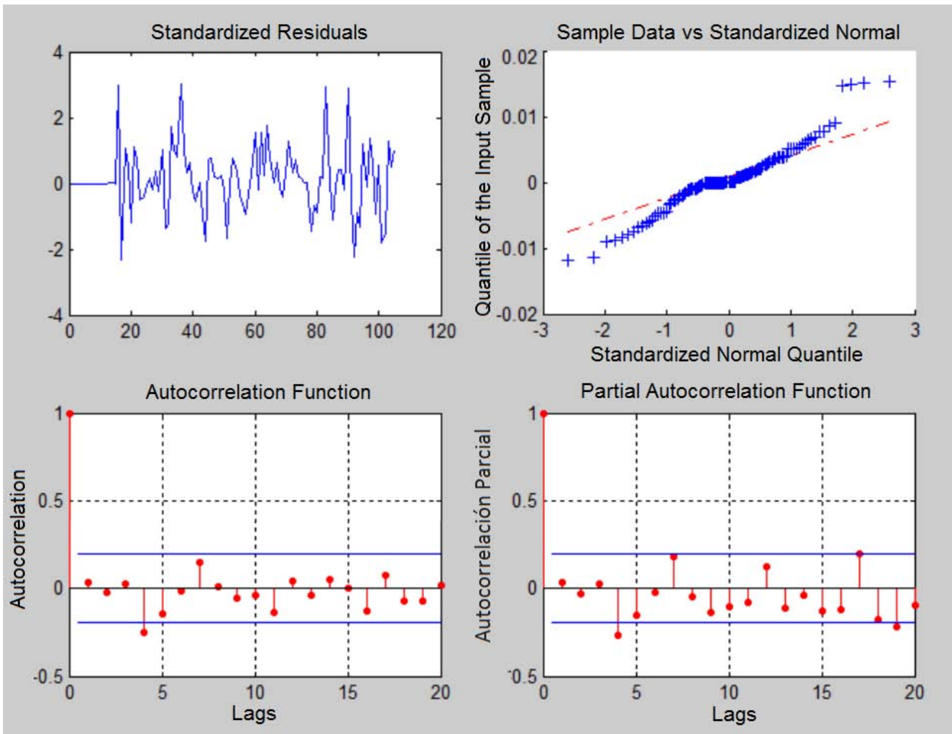

Fig. 9. Residual analysis of C4 channel model

As with the time series of the channel C3, the time series of channel C4 is processed through the logarithmic differentiation once to make it stationary. After the analysis of the FAC and FACP, the model is identified, the result is also an ARIMA (0,1,1). The process model estimation and validation of the estimate by analysing the correlation of the residuals. Fig. 9 describes the correlation analysis waste time series for the channel C4.

In Fig. 9 it can be established that there is still correlation at lag number 4, therefore it is necessary to iterate and estimate again the model coefficients including that lag.

After performing 13 iterations, the final estimation of ARIMA (0,1,14) model described by equation (2) is obtained.

$$z_t = \varepsilon_t - 0{,}5881\varepsilon_{t-1} - 0{,}9356\varepsilon_{t-14} \qquad (2)$$

Fig. 10 shows the predicted data compared to the original time series within a confidence interval of 48% for channel C4. The prediction model was performed in the same way as for channel C3.

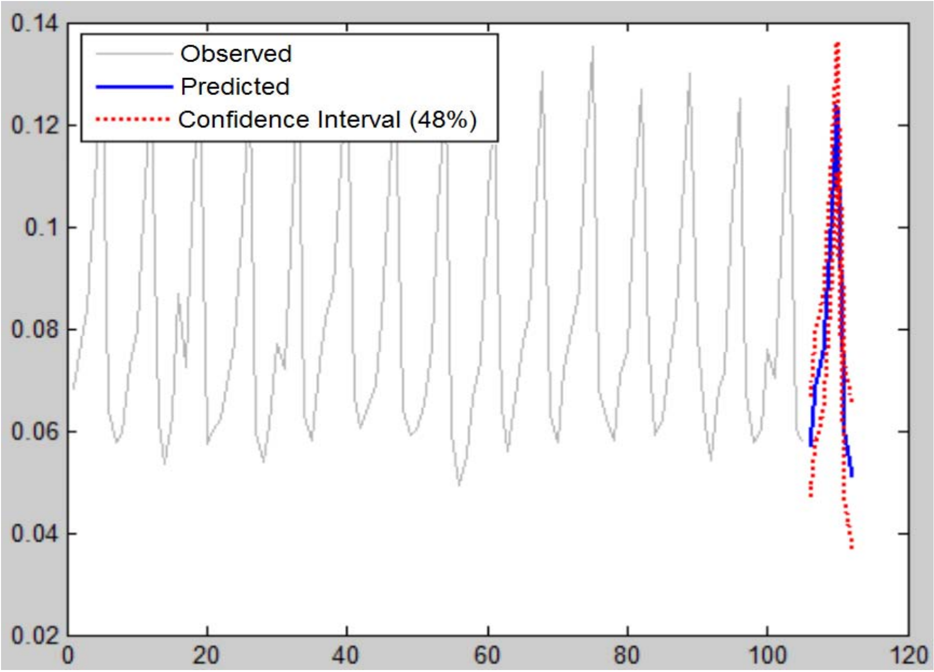

Fig. 10. Forecast of the rating of channel C4

VI. Results analysis

To compare and validate forecast of developed ARIMA models. Table III shows the value of samples rating for the past 7 days, measured by the UNE application and predicted with ARIMA model for C3 channel as well as the number of measured and predicted users, for the same channel, along with its corresponding percentage of error. Fig. 11 describes the behaviour of the data in Table III.

TABLE III. Comparison of Sample Measured vs Predicted for Rating and Number of Users C3 Channel

Measured value Channel C3		Predicted value Channel C3		Error (%)
Rating	No. of Users	Rating	No. of Users	
0.0632	316	0.0622	311	0.10%
0.0744	372	0.0722	361	0.22%
0.0866	433	0.0844	422	0.22%
0.0630	315	0.0661	331	0.31%
0.0758	379	0.0723	362	0.35%
0.1420	710	0.1357	679	0.63%
0.1786	893	0.1691	846	0.95%

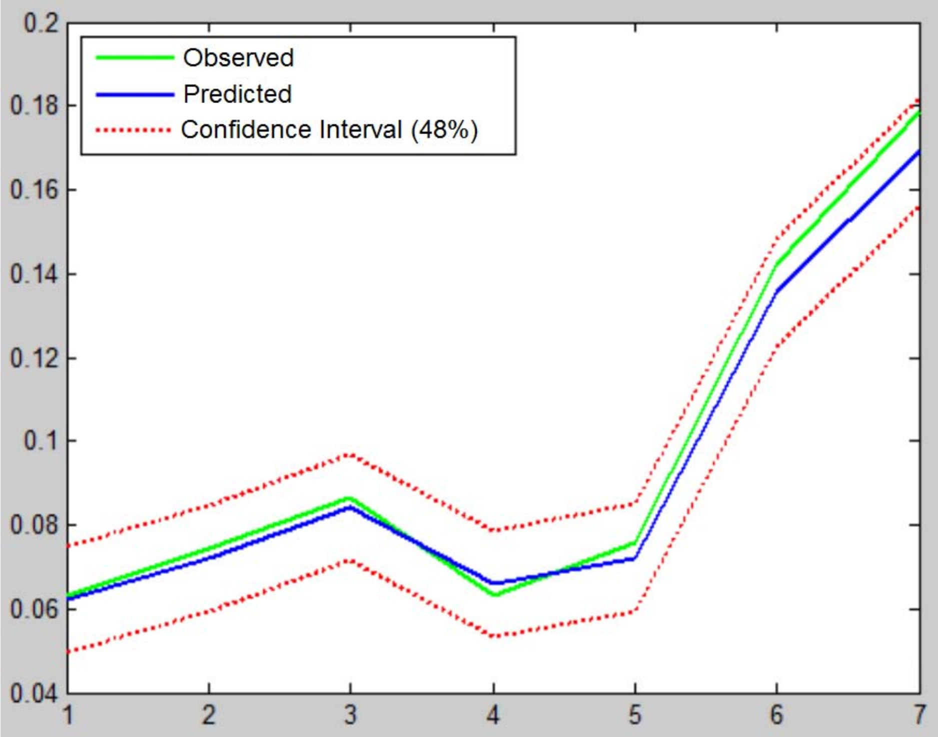

Fig. 11. Samples measures vs predicted samples for channel C3

Table IV shows the same information in Table III but for channel C4. Similarly Fig. 12 describes the behaviour of data in Table IV.

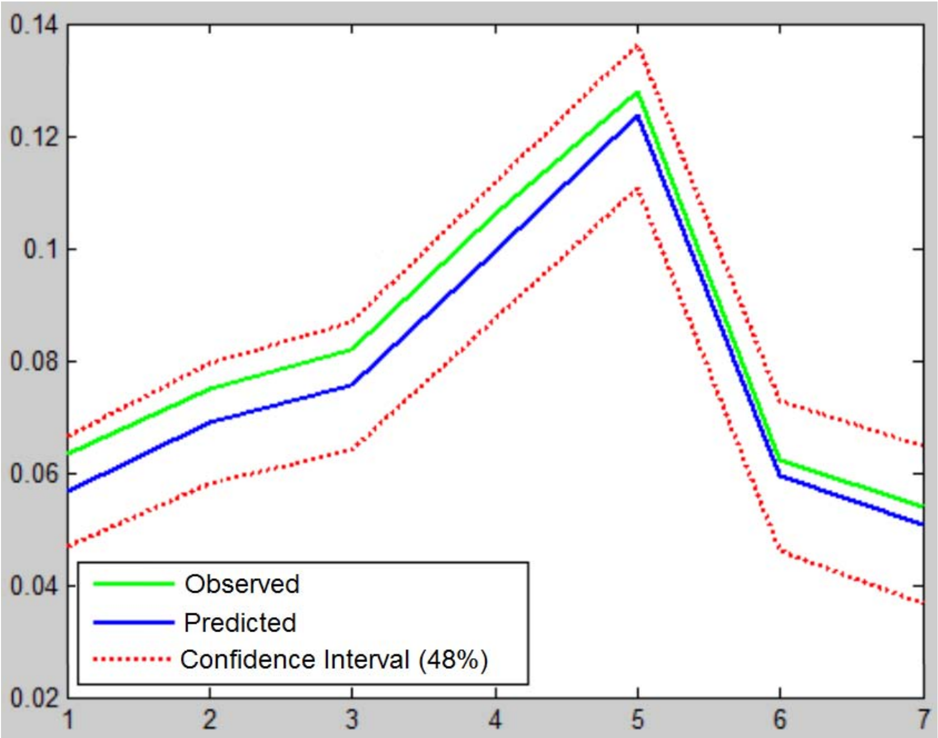

Fig. 12. Measured samples vs predicted samples for channel C4

TABLE IV. Comparison of Measured vs Predicted Sample of Rating and Number of Users of Channel C4.

Measured value Channel C4		**Predicted value Channel C4**		**Error (%)**
Rating	**No. of Users**	**Rating**	**No. of Users**	
0.0632	316	0.0566	283	0.66%
0.0748	374	0.0688	344	0.60%
0.0818	409	0.0755	378	0.63%
0.1062	531	0.0996	498	0.66%
0.1280	640	0.1236	618	0.44%
0.0624	312	0.0595	298	0.29%
0.0540	270	0.0508	254	0.32%

VII. CONCLUSION

A mechanism is proposed to mitigate the delay in changing the channel, consisting of the creation of a group of selected channels from the IPTV head, so that when the user selects one of the channels defined by the group, the delay is minimized at the required decoding time around 100 milliseconds.

The ARIMA time series model, has a satisfactory performance at the time of modelling the behaviour of rating for channels C3 and C4, which was validated through actual rating data corresponding to a week, with an average error less than 1%, in the two selected channels.

The MPTS group consists of the four most popular channels, where it can be attained a mitigation of Channel Zapping Delay of approximately 68%, based on the rating levels of channels C1, C2, C3 and C4 during the period sampled.

The first recommendation proposes to design a mechanism for an IPTV platform that allows sending a request of MPTS group from the head to the end user with the highest-rated channels, from the time when the decoder is turned on.

The second recommendation is to develop a multivariate time series model (VAR or Varma) to analyse rating. The advantage of these models lies in the fact that the dependent variable is not explained solely by a time variable, but through several explanatory variables, such as: slots, the specific time of measurement, and the geographical area of the sample, among others.

ACKNOWLEDGMENT

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

REFERENCES

[1] Van Wallendael, G.; Van Lancker, W.; De Cock, J.; Lambert, P.; Macq, J.-F.; Van de Walle, R., "Fast Channel Switching Based on SVC in IPTV Environments," IEEE Transactions on Broadcasting, vol. 58, no. 1, pp. 57-65, March 2012.

[2] Ramos, F.M.V., "Mitigating IPTV zapping delay," IEEE Communications Magazine, vol. 51, no. 8, pp. 128-133, 2013.

[3] Fernando M.V. Ramos, Jon Crowcroft, Richard J. Gibbens,Pablo Rodriguez, Ian H. White, “Channel Smurfing: Minimising Channel Switching Delay in IPTV Distribution Networks”, IWITMA 2010, Best Paper Award.

[4] Vadakital, V.K.M.; Gabbouj, M., "Prediction and Transmission Optimization of Video Guaranteeing a Bounded Zapping-Delay in DVB-H," IEEE Transactions on Broadcasting, vol. 57, no. 2, pp. 231-245, June 2011.

[5] Chae Young Lee; Chang Ki Hong; Kang Yong Lee, "Reducing Channel Zapping Time in IPTV Based on User's Channel Selection Behaviors," IEEE Transactions on Broadcasting, vol. 56, no. 3, pp. 321-330, Sept. 2010.

[6] Soyoon Lee; Hyewon Moon; Hyokyung Bahn; Taeseok Kim; Ik-Soon Kim, "Popularity and adjacency based prefetching for efficient IPTV channel navigation," IEEE Transactions on Consumer Electronics, vol. 57, no. 3, pp. 1135-1140, August 2011.

[7] Alvarez, F.; Martin, C.A.; Alliez, D.; Roc, P.T.; Steckel, P.; Menendez, J.M.; Cisneros, G.; Jones, S.T., "Audience Measurement Modeling for Convergent Broadcasting and IPTV Networks," IEEE Transactions on Broadcasting, vol. 55, no. 2, pp. 502-515, June 2009.

[8] R. Sotelo, A. Gil Solla, “Utilizando Sistemas Recomendadores para Predecir Ratings en TV”, Informática na Educação: teoria & prática, Porto Alegre, v. 14, n. 1, pp. 11-27, jan./jun. 2011.

[9] Yue Teng; Liang He, "Rating prediction algorithm and recommendation based on user beahavior in IPTV," 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet), pp. 3373-3378, 21-23 April 2012.

[10] Cesar Hernández, Octavio Salcedo, Andrés Escobar. "An ARIMA model for forecasting Wi-Fi data network traffic values". Ingeniería e Investigación, vol. 29, no. 2, pp. 65-69. 2009.

[11] Cesar Hernández, Víctor Gómez, Alejandro Peña, Calidad del servicio en la energía eléctrica residencial. Editorial UD, Bogotá, Colombia, pp. 134, marzo de 2013.

[12] Cesar Hernández, Hans Márquez, Discapacidad auditiva. Editorial UD, Bogotá, Colombia, pp. 112, noviembre de 2014.

[13] R. Kooji, K. Ahmed and K. Brunnstrom, "Perceived quality of channel zapping", Proc. of 5th IASTED Int. Conf. on Communication Systems, Networks, and Applications, Aug. 2006.

[14] Siebert, P.; Van Caenegem, T.N.M.; Wagner, M., "Analysis and Improvements of Zapping Times in IPTV Systems," IEEE Transactions on Broadcasting, vol. 55, no. 2, pp. 407-418, June 2009.
[15] Daniel Mauricio Chiqui, “Estudio de factibilidad de IPTV en la red IP/MPLS de etapa EP utilizando VPN/MPLS”, Universidad Politécnica Salesiana. Cuenca, 2015.
[16] C. Orbe, “Estudios de migración de sistemas de audio y video por suscripción bajo la modalidad de cable físico a IPTV con sugerencias en el ámbito regulador”. Quito: Escuela Politécnica Nacional, abril 2010.
[17] Erick Vinicio Sánchez “Implementación de IPTV a través de Enlaces de Internet de Banda Ancha” Universidad San Carlos de Guatemala, Guatemala, 2008.
[18] Luis Pedraza, Cesar Hernández, Katherine Galeano, Enrique Rodríguez de la Colina, Ingrid Páez, Ocupación espectral y modelo de radio cognitiva para la ciudad de Bogotá. Editorial UD, Bogotá, Colombia, pp. 126, noviembre de 2015.

AUTHOR PROFILE

Ricardo Alonso Ferro Bolivar Electronic Engineering, Master of Science in Information and Communications. Professional maintenance of Radio and TV in Señal Colombia. , Bogotá D.C. Colombia.

Cesar Augusto Hernandez Suarez Electronic Engineer, Master in Information Science and Communications, Candidate To 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.

Gabriel Alberto Puerta Aponte Electronic Engineering, Master of Science in Information and Communications. Professor and researcher at Universidad Minuto de Dios, Bogotá DC, Colombia.