ARIMA ANALYSIS FOR GLIOMA GROWTH AFTER RADIOTHERAPY TREATMENT

Tamara Cristina Onias

Eliane Da Silva Christo

tamara_onias@id.uff.br

elianechristo@id.uff.br

Programa de Pós-Graduação em Engenharia de Produção - UFF EEIMVR

Address, Zip-Code, State, Province, Country

Av. Dos Trabalhadores n. 420 - Vila Santa Cecilia - Volta Redonda, 27255125, RJ, Brazil

Vanessa da Silva Garcia

vanessa.sgarcia@gmail.com

Programa de Pós-Graduação em Engenharia de Produção - UFF EEIMVR

Av. Dos Trabalhadores n. 420 - Vila Santa Cecilia - Volta Redonda, 27255125, RJ, Brazil

Gustavo Benitez Alvarez

benitez.gustavo@gmail.com

Programa de Pós-Graduação em Engenharia de Produção - UFF EEIMVR

Av. Dos Trabalhadores n. 420 - Vila Santa Cecilia - Volta Redonda, 27255125, RJ, Brazil

Abstract. *Cancer, generic word for disorderly proliferation of destructive cells, is still a mystery illness for the scientific community. This article will approach, in general way, glioma, brain tumors that arise from glial glands, responsible for protecting brain neurons and other regions of the nervous system. In several existing treatment methods, the most effective used is the irradiation therapy. Its main objective is the destruction of residual cells after removal of the tumor by surgery, such as to delay or prevent tumor recurrence. Many studies have been done in the field of computer modeling in order to equalize the tumor growth rate after treatment with radiation. The objective of this article is to describe an appropriate ARIMA model to analyze the forecast of glioma growth after irradiation therapies and compare its efficiency with others models.*

Keywords: *Time Series, ARIMA, Glioma, Radiotherapy.*

1 INTRODUCTION

According to INCA (*Brazilian National Cancer Institute*), glioma is the most frequent histological type of Central Nervous System cancer in Brazil, representing between 40% and 60% of all cases and being more common in the adult age group (INCA, 2016). Glioma are tumors that can arise in the brain or vertebral column. They receive this name owing to the fact that proliferation arises from the glial glands, which protect brain neurons and other regions of the nervous system. According to *World Health Organization*, glioma are classified by degree, higher the degree, higher the mitosis rate and the level of aggressiveness, which makes the patient's survival time less than one year.

There are several treatment methods for glioma such as surgery, chemotherapy, radiation therapy, hormone therapy and palliative care. Radiation therapy is the most common choice of treatment for glioma, being the most effective among them, a fact scientifically proven. The main purpose of radiotherapy is the destruction of the residual cells after or not removal of the tumor by surgery, in such a way as to delay or prevent recurrence of the tumor.

Continuous models were developed to simulate tumor evolution after these therapies, making it possible to predict the evolution of the tumor radius for each treatment dosage, and to estimate the optimal dose. It is possible to highlight the continuous model proposed by K. R. Swanson (2008), which consists of a second-order partial differential equation of the diffusive reactive type.

Several studies already exist in the academic environment foresee the evolution of the tumor radius for each dose of treatment, estimating the optimal dose. These forecasts can be stipulated through time series that aim to predict future values or study the structure of the series and its relation to other series. There are several methods for structuring these series, such as Exponential Cushion as approached by Jesus et al (2016). In addition, the Box & Jenkins methods that are based on the definition of an integrated auto regression model and moving averages (ARIMA) that represent the stochastic process generating the time series.

This paper will aim to find the ARIMA model more suitable for the prediction of glioma growth that minimizes the possible errors of the series and compare its effectiveness in relation to other methods already used in the bibliography. The computational software for analysis is Minitab. It is expected as a result of this paper, to verify, model and simulate, with a certain level of confidence, the survival rate of glioma progression after radiotherapy procedures from analyzes of the time series.

2 LITERATURE REVIEW

Briefly, it is presented the continuous model that describes the effects of the radiotherapy will be represented by an equation, proposed from a data survey, which can be experimental or simulated. In 2008, Kristin Swanson, a scientist specializing in oncology mathematics, developed a partial second-order differential equation of the diffusive reactive type that describes this evolution, equation (1). This model was solved by the finite difference method, resulting in a temporal data domain generated by the Crank-Nicolson method by Silva et Al (2014). The objective of this article is not study this model, but use its numeric results to analyze the growth of tumor with time series.

Proceedings of the XXXVIII Iberian Latin-American Congress on Computational Methods in Engineering P.O. Faria, R.H. Lopez, L.F.F. Miguel, W.J.S. Gomes, M. Noronha (Editors), ABMEC, Florianópolis, SC, Brazil, November 5-8, 2017.

$$\frac{\partial c}{\partial t} = \nabla.(D\nabla c) + \rho c - R\big(\alpha, d(x,t)\big)c \text{ em } \Omega \quad (1)$$

$$c(x,0) = c_0;\ \hat{n} \cdot \nabla c = 0 \text{ em } \partial\Omega \quad (2)$$

The partial derivative refers to the rate of concentration of tumor cells at the point x of the brain (Ω) at time t. D is the coefficient that represents the speed of the cell to move in its neighborhood in mm² / day, ρ s the rate of tumor mass growth in units / day. $d(x,t)$ is the dose to be fractionated in the patient in the position x at time t. The argument ρ represents the tax of proliferation in units/day that describes the tumor masse growth by the exponential law (JESUS, 2016). Finally, $c(x,0) = c_0$ corresponds to the initial condition of the patient at the time of the first examination. $\hat{n} \cdot \nabla c = 0 \text{ em } \partial\Omega$ is the contour condition, which ensures that the equation represents only the brain domain (SWANSON et Al, 2008).

The continuous model described in equation (1) e (2) has origin in 90s, after studies that proves that gliomas do not do metastases for out of the brain and that tumor proliferation follow exponential law for its growth (RESTKY, 2009).

To quantify the glioma tumor, dominants parameters (D and ρ) are simulated for each patient using two magnetic resonance images. As well, it is possible to describe the behavior of tumor growth in response to radiotherapy through the parameter $R\big(\alpha, d(x,t)\big)$. This parameter use the linear quadratic model, presented by equation (3), to calculate effects of the dosage fractionation by adjustable parameters α and β. Adjustable parameters verify the efficacy of radiation in relation to probability of survival of cells.

$$S\big(\alpha, d(x,t)\big) = e^{-\alpha BED}, BED = N\left(d + \frac{d^2}{\alpha/\beta}\right) \quad (3)$$

The term S (α, d(x, t)) represents the survival probability, BED is the biological effective dosage of radiation, N the number of dosage fractions. Adjustable parameters α and β are related to the cell damage: a linear (α), related to a simple DNA tape break and a quadratic (β) related to a double DNA tape break. The ratio α/β represents the sensibility of the tissue to the fractionation dosage. The term referring to the effects of radiotherapy is represent by equation (4).

$$R(\alpha, d(x,t)) = \begin{cases} 0 & \text{for } t \notin \text{therapy} \\ 1 - S\big(\alpha, d(x,t)\big) & \text{for } t \in \text{therapy} \end{cases} \quad (4)$$

Doses are defined by fractionation according to the number of days of treatment, called DOT. Doses are measured in fractions of Grays (Gy) that represents the quantity of radiation energy absorbed per masse unit that means, a Joule of radiation absorbed by a kilogram of material (J/kg). For the data of this article, the protocol used is according to the University of Washington Medical Center that advises a dose limit of 61.2 Gy + 5% per treatment for the patient (SWANSON et Al, 2008).

Focusing on individual treatment planning for the patient, Silva et Al (2014) resolves the equation (1) numerically through finite difference method, discretizing the temporal domain via Crank-Nicolson, which allows the conduction of predictions across the set of observations of time-correlated tumor cell concentrations in response to radiotherapy. Therefore, making possible the monitoring and control of the tumor, estimating the efficacy of the treatment. For the planning of the treatment, estimates of tumor behavior are required. Therefore, Jesus et Al

Proceedings of the XXXVIII Iberian Latin-American Congress on Computational Methods in Engineering P.O. Faria, R.H. Lopez, L.F.F. Miguel, W.J.S. Gomes, M. Noronha (Editors), ABMEC, Florianópolis, SC, Brazil, November 5-8, 2017.

(2016) used time series, more precisely the exponential smoothing method of Holt to predict the tumor growth. In order to compare the obtained results, the Holt method with additive error (A, A, N) and multiplicative error (M, N, A) was used. For all methodologies, five types of dose fractionation were simulated, concluding that the DOT = 5 days are more effective, reducing the tumor radius size and maintaining its constant growth, close to zero, for a longer period of time than those observed for other dosages.

The proposed work aims to use another method of time series to predict the glioma growth, considering that the DOT=5 is the best option of treatment. The Box & Jenkins methods or ARIMA was choose to predict the glioma growth in this work.

3 PROPOSED WORK

The Box - Jenkins model (ARIMA) has the ability to manipulate, in principle, time series of any nature, which is why it was approached in this article. Auto regression models (AR) and Moving Average (ARMA) models predict time series values by combining past real values.

The ARIMA method is based on the definition of an integrated auto-regressive model and moving averages (ARIMA) that represent the stochastic process generating the time series, based on an applicable ARMA model in the description of stationary time series (which do not vary in relation to time), extending this concept to non-stationary time series (MUELLER, 1996).

The number of times the original series needs to be differentiated before generating a stationary series is called the "order of homogeneity", denoted by d. These processes are described by the auto-regressive models, moving averages of orders p, d and q and ARIMA (p, d, q), which can be considered a seasonal fraction (s1, s2, s3) (MUELLER, 1996).

The difference between the classical regression and the time series models is that in ARMA (or ARIMA) time series models, independence between observations cannot be assumed. Quite the opposite, the self-regressive and moving-averages models will model the degree of self-correlation between deviations and lagged observations (MUELLER, 1996). ARIMA models are adjusted to the original series. In ARMA models, the series is a differentiated series. If Yt is the differentiated series, the ARMA models (p, q) are described by:

$$Yt = \alpha + \phi 1Yt\text{-}1 + \ldots + \phi pYt\text{-}p + et - \theta 1et - 1 - \ldots - \theta qet - q \quad (5)$$

The variables p and q represent the number of parameters related to the generation pattern and α is a constant term that is related to the mean of the stochastic process (MUELLER, 1996). The sum of the parameters $\phi 1$, $\phi 2$, ..., ϕp must be less than 1 to characterize a stationary process. The parameters $\theta 1$, $\theta 2$, ..., θq represent the weights of the moving average being positive or negative. Et equals to the error of random events that are not characterized by the model.

For identifying ARIMA models, there are some tools those can be used: ACF and PACF plots.

As first step, the time series is differenced until it is stationarized, the next step is to find an appropriated ARIMA model to determine how AR or MA terms correct autocorrelations that remains in the differenced series. Looking at the autocorrelation function (ACF) and partial

Proceedings of the XXXVIII Iberian Latin-American Congress on Computational Methods in Engineering P.O. Faria, R.H. Lopez, L.F.F. Miguel, W.J.S. Gomes, M. Noronha (Editors), ABMEC, Florianópolis, SC, Brazil, November 5-8, 2017.

autocorrelation (PACF) plots of the differenced series, it is possible to identify the numbers of AR and/or MA terms that are necessary. The ACF plot is a bar graph of the coefficients of correlation between a time series and delays of itself. The PACF plot is a plot of the partial correlation coefficients between the series and delays of itself.

The Partial correlation between two variables is the amount of correlation between them, which is not explained by their mutual correlations with a specified set of other variables. A partial autocorrelation is the amount of correlation between a variable and a lag of itself that is not explained by correlations at all lower-order-lags (NAU, 2017).

Time series forecasting method needs to be tested to evaluate its efficiency. In this work it will be used AIC and BIC, those allow know the model that best fits the series of simulated data analyzed. The objective of those tests is to minimize AIC and BIC values for a functional model. For lowers values, better quality in the data fit for the model. Anyway, AIC and BIC are used for different meanings (NAU, 2017).

The structure of the model was made through an interactive cycle, making the choice of the model based on the data collected. The construction of the model was distributed in five steps:

1. Data preparation: In this step, a series of temporal data were collected.
2. Model choice: In this step, the model was chosen through the analysis of the behavior of the autocorrelation functions (ACF) and partial autocorrelation (PACF).
3. Estimation of parameters: In this step the estimation of parameters (p, d, q) of the model was made, that generated the time series Yt, the moving averages (parameter q), the autoregressive stationary variable (p parameter) and the integration variable Non-stationary (parameter d).
4. Adequacy of the model: In this stage, the model verification was performed through the AIC (Akaike Information Criteria) and BIC (Bayesian Information Criteria) tests.
5. Use of the model for forecasting: In this last step, the model was used for forecasting by making the correlogram of the series predicted with the original series.

For the methodology of this article, a 6th stage was implemented in order to compare the efficiency of the model defined with others in the literature. Figure 1 represents the flowchart used for the methodology of this article.

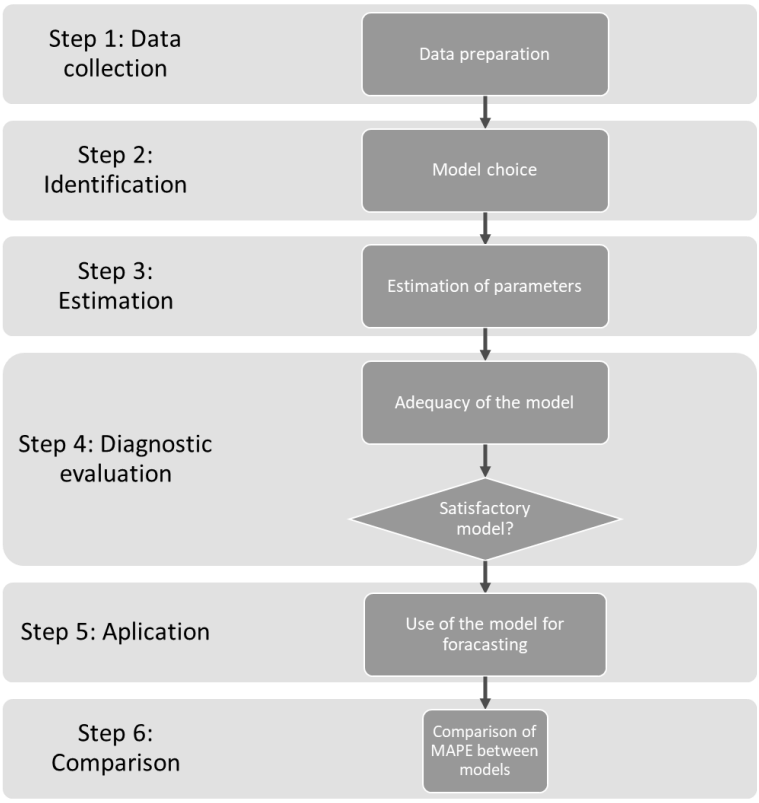

Figure 1. Flowchart of the methodology using ARIMA

It is worth noting that predictions using ARIMA models will be effective for a short period and the best predictions will be those with a minimum mean square error (SAE).

4 RESULTS

4.1 Data Collection

For this work, it was used the results of the simulations made by Silva e Al (2014) through the reactive partial equation of the diffusive type, equation (1), which calculates the growth rate of tumor cells in response to radiotherapy using the model of Swanson. These data were established by the concentration of glioma cells for each position in the patient's brain as described in section II.

For the approximation of the radius, values were calculated from the initial concentration (C_0) of glioma cells, using the finite difference method (SILVA, 2014), where the radius can be defined by:

$$R = L_c \times H_x \tag{6}$$

R is the radius in cm, the immediate upper observation at 61.26% of , and the value of L divided by the number of observations in the period of 12 hours. This approximation was performed for DOT = 5 days.

As result of the simulation, there are a series of tumor radius in function of the time in days, present by the graph of figure 2.

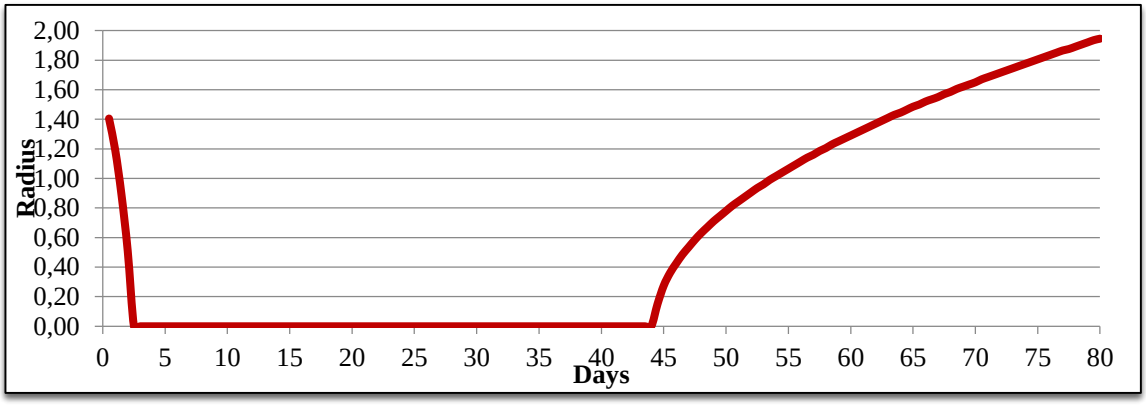

Figure 2. Glioma radius evolution after radiotherapy – Adaptation (SILVA, 2014)

4.2 Application of ARIMA for radiotherapy

In this phase, it was verified the existence of the necessity of a transformation in the original series with the aim of stabilizing its variance. Resulting in a more symmetrical and near-normal distribution.

The first requirement for ARIMA modeling is that the time series of data to be modeled is stationary or can be transformed into one. It can be defined that a time series is stationary if it has a constant mean and has no tendency over time.

Considering the series of radius per day, the appropriate form of the model was analyzed calculating the autocorrelation function (ACF) and the partial autocorrelation function (PACF).

It was possible to observe that for the first interval the partial autocorrelation is high, it corresponds to the initial phase, before treatment, of the glioma. However, in the course of the evolution of the graph, the PACF tends to zero.

These results were compared with the theoretical patterns of ACF and PACF, due to the impossibility of resembling them, it is concluded that there is a need to differentiate the data and relaunch the autocorrelation methods (WILLIAN, 2016).

After the first differentiation (d=1) and launching the methods, it is obtained the graphs of Figure 3 and Figure 4.

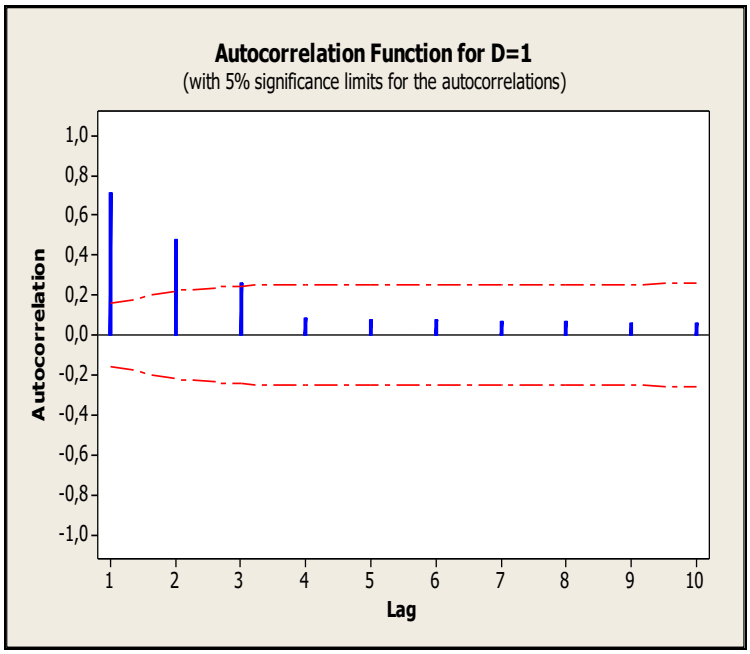

Figure 3. Autocorrelation Function for D=1 (ACF)

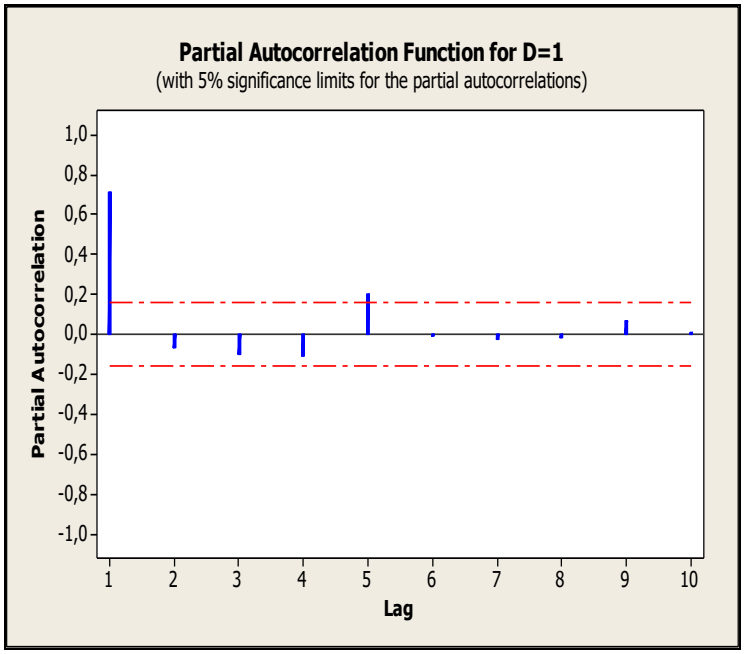

Figure 4.Partical Autocorrelation Function for D=1 (ACF)

Analyzing figures (3) and (4) and comparing them with the theoretical standards, it is possible to realize that the model analyzed resembles the ARIMA (1, 1, 2), (2, 1, 2) models. Analysis done according to the number of peaks out of range for each graph.

After identifying the possible models, the software Minitab ® software was used to implement the ARIMA procedure. Resulting in the figures (5), (6), (7) and (8).

- **ARIMA (1,1,2)**

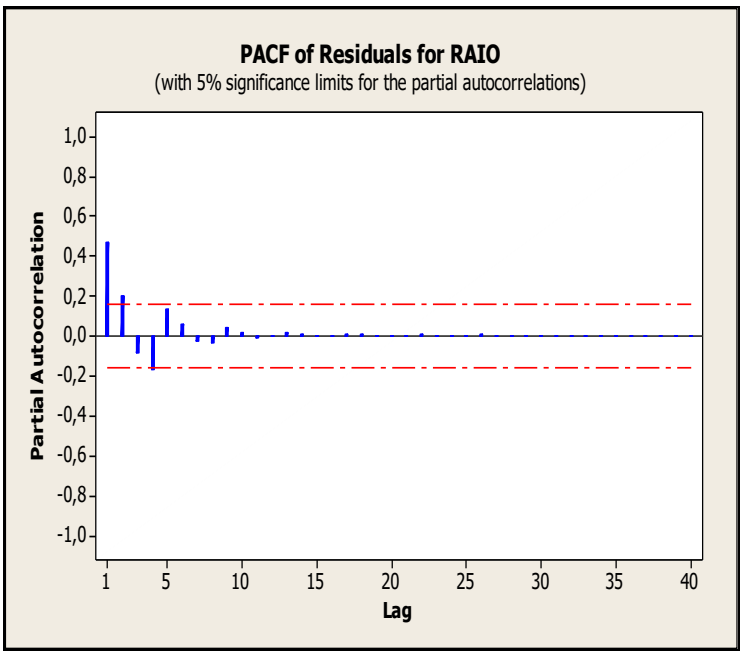

Figure 5. PACF of Residuals for model ARIMA (1,1,2)

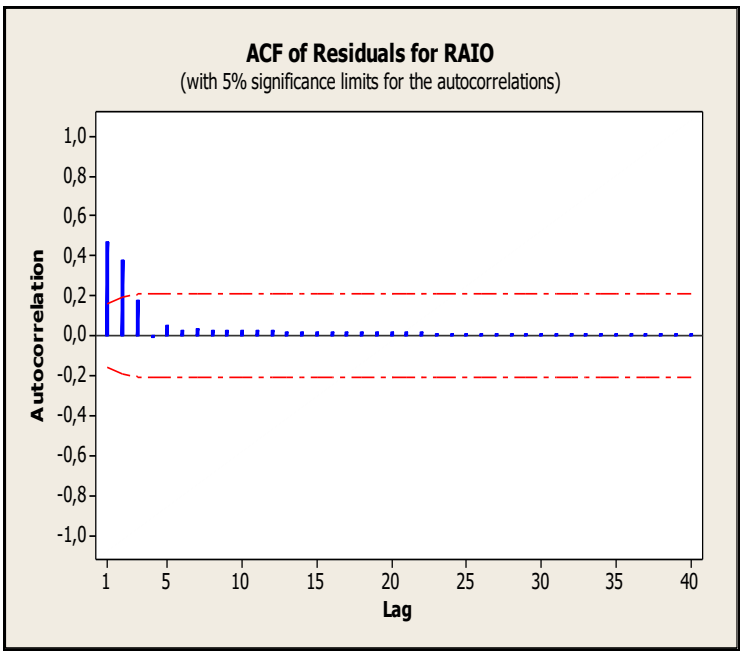

Figure 6. ACF of Residuals for model ARIMA (1,1,2)

- **ARIMA (2,1,2)**

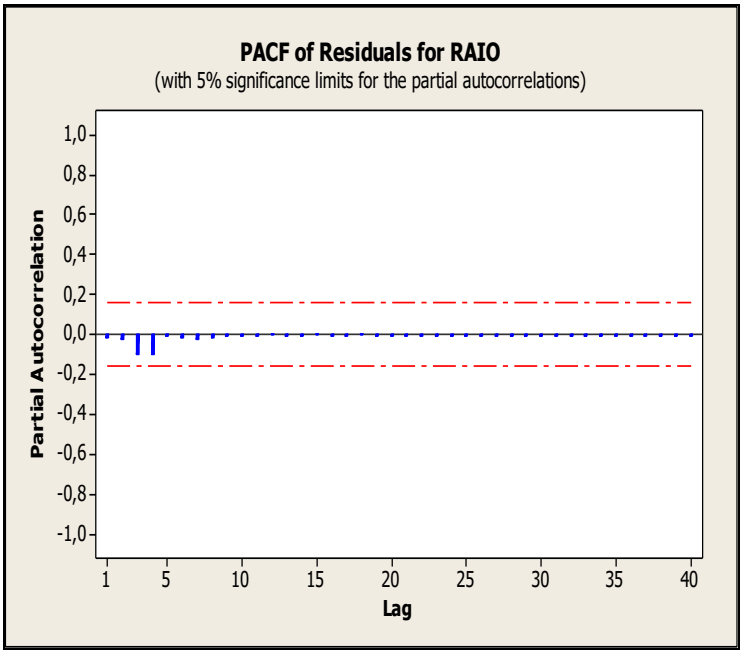

Figure 7. PACF of Residuals for model ARIMA (2,1,2)

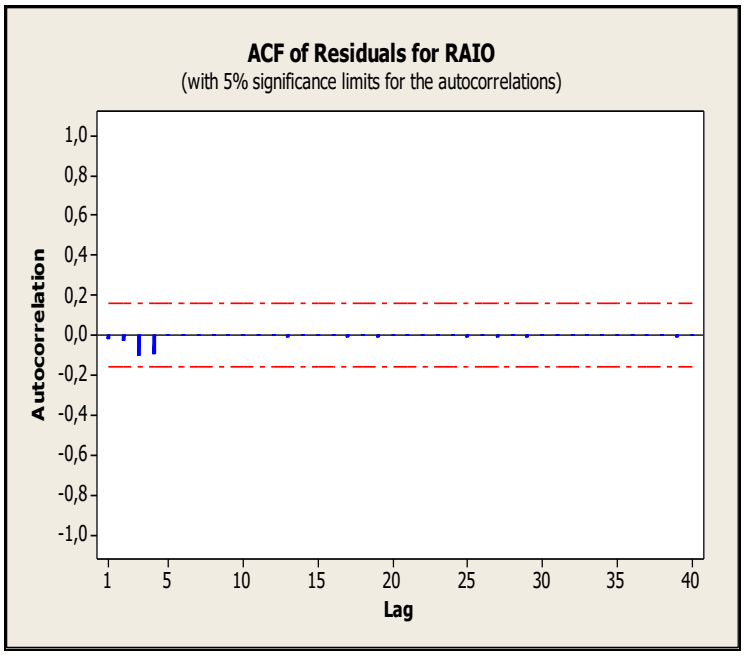

Figure 8. ACF of Residuals for model ARIMA (2,1,2)

Only analyzing figures is not possible to decide exactly which best model is suitable for the series, therefore it was analyzed the tests AIC (Akaike Information Criterion) and BIC (Bayesian Information Criteria), defined by:

$$AIC = n \ln(\bar{\sigma}^2) + 2(k+1) \quad (7)$$

$$BIC = n \ln(\bar{\sigma}^2) + (k+1) * \log n \quad (8)$$

The values of table 1 were found, for n = number of samples and k the number of variables of the model.

Table 1 AIC x BIC for Radiotherapy

	AIC	BIC
ARIMA (1,1,2)	-1756	-1750
ARIMA (2,1,2)	-1820	-1814

AIC and BIC increase as SQE increases. As the smaller AIC and BIC values are preferable, the best model is the ARIMA (2,1,2). Resulting in the parameters of Figure 4 for the ARIMA model.

```
Final Estimates of Parameters
Type       Coef  SE Coef      T      P
AR   1   1,6550   0,0984  16,82  0,000
AR   2  -0,6572   0,0927  -7,09  0,000
MA   1   0,9277   0,0937   9,90  0,000
MA   2   0,0083   0,0736   0,11  0,911

Differencing: 1 regular difference
Number of observations:  Original series 160, after differencing 159
Residuals:    SS =  0,277370 (backforecasts excluded)
              MS =  0,001789  DF = 155

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag             12     24     36     48
Chi-Square     3,1    3,2    3,2    3,2
DF               8     20     32     44
P-Value      0,925  1,000  1,000  1,000
```

Figure 9. Parameter of ARIMA's model (2,1,2)

4.3 MAPE for ARIMA x HOLTs for Radiotherapy

Estimating the model and choosing the one that best suited the series, predictions were made for the growth of the glioma and the error was calculated by:

$$MAPE = \frac{1}{n} \sum ABS\left(\frac{\text{Real} - \text{Previsto}}{\text{Real}} \right) \quad (9)$$

MAPE = 0,923024469

Comparing the MAPE values for the Holt Method and for the ARIMA, it was verified that the Holt Method has the highest value (2.339927656) (JESUS, 2016). Thus the prediction for the ARIMA method is the most effective.

CONCLUSION

In this article, simulated data generated by continuous models describing the growth behavior of glioma in response to radiotherapy were analyzed.

Regarding the modeling of ARIMA, it was verified the good adjustment of the methodology to the simulation of glioma growth. It was possible to find a model that fit the data (2, 1, 2), guaranteeing a prediction with error of 0,923024469, which is smaller than the error found in the literature using the exponential smoothing method of Holt.

These results are important, especially for researchers who wish to continue studies through continuous models for glioma growth. These results show that parameters linked to the characteristics of each therapy should be incorporated into the models in order to be able to perform a better survival analysis. Studies such as these may be essential for health experts to make decisions about which therapies are necessary or ideal for a given patient and predict their evolution after treatment.

Acknowledgements

Acknowledgments to all authors involved, to postgraduate programs PPGEP and MCCT of the Federal University Fluminense.

REFERENCES

INCA, “Estimativa 2016: incidência de câncer no Brasil / Instituto Nacional de Câncer José Alencar Gomes da Silva – Rio de Janeiro: INCA, 2015. Available: <http://www.inca.gov.br/estimativa/2016/estimativa-2016-v11.pdf> Access: Mars, 4, 2016.

JESUS J.C., E. D. S. Christo, V. D. S. Garcia, and G. B. Alvarez, “Time Series Analysis for Modeling of Glioma Growth in Response to Radiotherapy.” IEEE Latin America Transactions, March 2016, Vol.14 (3), pp.1532-1537

RESTKY, M. W., D. E. Swartzendruber, R. H. Wardwell, and P. D. Bame. Is gompertizian or exponential kinetics a valid description of individual human cancer growth? Medical Hypotheses, pages 90-110, 2009.

MUELLER, A. Uma aplicação de redes neurais artificiais na previsão do mercado acionário. 1996. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-graduação em Engenharia de Produção, UFSC, Florianópolis.

NAU, R. Statistical forecasting. Available in: https://people.duke.edu/~rnau/411home.htm. 2017. Access: Mars, 4, 2016.

SILVA, J. J. d. "Modelagem Computacional Aplicada ao Tratamento do Câncer via Medicina Nuclear," Mestrado, Pós-Graduação Em Modelagem Computacional Em Ciência E Tecnologia Universidade Federal Fluminense, 2014

SWANSON K. R., R. Rockne, E. C. A. Jr., and J. K. Rockhill. A mathematical model for brain tumor response to radiation therapy. Jornal of Mathematical Biology, 580: 560-580, 2008

WILLAMYS Rangel Nunes Sousa, M. Sousa Couto, A. Felix Castro, and M. Pereira Santos Silva, "Evaluation of desertification processes in ouricuri-pe through trend estimates of times

series", Latin AmericaTransactions, IEEE Latin America Transactions, Feb. 2013, Vol.11(1), pp.602-606

WILLIAN Jacobs, A. Mendonça Souza and R. Ruviaro Zanini, "Combination of Box-Jenkins and MLP/RNA models for forecasting", IEEE Latin America Transactions (Volume: 14, Issue: 4, April 2016)