FORECASTING OF BUFFALO MILK IN A BRAZILIAN DIARY USING THE ARIMA MODEL

Lara Moura Silva Saúde*, Gustavo Teodoro Gabriel and Pedro Paulo Balestrassi

ABSTRACT

Brazil presents a great diversity in herd rearing. These herds are divided into nine categories that are of great importance, both for domestic consumption and for export. Among these nine groups, buffaloes are the type of herd that has the least amount of animals in breeding stock. Despite presenting a small number of heads, the consumption of milk and dairy products in this group has been growing in recent years. In view of this situation, the objective of this article is to use the ARIMA time series model to model and forecast the milk demand of a buffalo dairy for one year and thus align product demand with production demand and inventory. For the analysis and construction of the model, a time series from a Brazilian dairy was used. These data represent the demand for raw materials from January 2011 to December 2016. The model was based on historical dairy data only. Initially, it is observed that the historical series shows trend and seasonality. After the data modelling, the forecast model reached was the ARIMA model (2,1,4) $(0,1,1)_{12}$, through Minitab software.

Keywords: *Bubalus bubalis*, buffaloes, time series, ARIMA, milk demand, dairy

INTRODUCTION

The dairy industry in Brazil is of great importance within the livestock sector. According to research conducted by Brazilian Institute of Geography and Statistics (IBGE) in 2016, milk production in the second quarter of 2016 corresponds to more than 5 million litters of milk. Among this total milk production, there is milk production of buffaloes.

According to the Ministry of Agriculture, Livestock and Supply (2016), the herd of buffalo is estimated at around 1.15 million. According to the Brazilian Association of Buffalo Breeders (2016), this amount represents 2.0% of the cattle herd. In addition, it states that 30.0% of this herd is intended for milk production. The IBGE (2016) shows the evolution of the number of buffaloes from 1981 to 2015 in Brazil.

The buffaloes represent the smallest number of herds in Brazil, being behind chickens, poultries, cows, pigs, quails, sheep, goats and horses, respectively, according to Figure 2. Although in this position, consumption of milks and derivatives has been growing in Brazil. Buffalo population presents an important role in the demand of meat and milk (Wanapat and Kang, 2013). In addition, the demand for buffalo milk

Federal University of Itajubá, Itajubá-Minas Gerais, Brazil, *E-mail: laramss@gmail.com

grew about 301.0% as opposed to cow's milk that grew around 53.9% in the same period (Jorge *et al*., 2011). Bastianetto (2009); Gregory (2014), affirm that in Brazil the production of buffalo milk has been driven by the appreciation of milk and an increasing demand of the consumer market.

Although the world buffalo population has been slightly increasing, their vital role are accountable for the demand of meat and milk.

In this way, it is necessary to predict the demand for buffalo milk for production and market share. Thus, the present work aims to predict the supply of milk to a dairy milk of buffalo using the ARIMA method.

LITERATURE REVIEW

Time series are among the most important issues that analysts face in various areas of research, from finance and economics, to operations management, policy, social issues, productivity, among other applications. A time series is a chronological or time-oriented sequence of observations about a variable of interest (Montgomery *et al*., 2015). The time series forecast aims to predict future events and serve as a basis during decision making.

Autoregressive integrated moving average models (ARIMA)

There are statistical characteristics that classify the time series into two types: stationary and non-stationary. The stationarity of a time series can be determined by removing records from different points in the process and observing the general behaviour of the time series. If the series presents similarity, it can be affirmed that the series is stationary, otherwise it is non-stationary (Montgomery *et al*., 2015). A detailed analysis with observation of graphical patterns and statistical tests of unit root allows the researcher to confirm the stationarity of the series. The most commonly used unit root test is Dickey-Fuller (Makridakis *et al*., 1998).

It is assumed that the time series are stationary, so the first step is to transform the series into stationary if it is not, as the case of the study of this work. The most commonly used transformation is to differentiate the series successively until obtaining a stationary series (Morretin and Toloi, 2014). The first difference is given by:

$$W_t=Z_t-Z_{t-1}=(1-B)Z_t=\Delta Z_t \quad (1)$$

The number of differences needed to make the series stationary is called the integration order. Since W_t is a difference of Z_t, then Z_t is an integral of W_t, in this way, it can be said that Z_t follows an integrated autoregressive model of moving averages, or ARIMA model, of order (p, d, q), where p and q are the orders of $\phi(B)$ and $\theta(B)$, respectively.

$$\phi(B)\Delta^d \; Z_t = \theta(B) \; a_t \quad (2)$$

Since all the roots of are outside the unit circle, we can write Equation 2 as $\varphi(B)Z_t = \theta(B)a_t$ which is a non-stationary autoregressive operator of order p + d, with d roots equal to one and the remaining p outside the unit circle (Morettin and Toloi, 2014).

$$\varphi(B)Z_t = \theta(B)a_t \quad (3)$$

Therefore, the usual form of the ARIMA model, for the forecast calculation is given by:

$$Z_t=\varphi_1 Z_{t-1}+\varphi_2 Z_{t-2}+\ldots+\varphi_{p+d} Z_{t-p-d}+a_t-\theta_1 a_{t-1}-\ldots-\theta_q a_{t-q} \quad (4)$$

Which $\varphi(B)=1-\varphi_1 B-\varphi_2 B^2-\ldots-\varphi_{p+d} B^{p+d}$

The temporal series can present strong periodic patterns, this characteristic is referred to as seasonal behaviour (Montgomery *et al*., 2015). The ARIMA seasonal models are known as SARIMA (p, d, q) (P, D, Q), where p and P indicate non-seasonal and seasonal AR respectively, q and Q indicate non-seasonal and seasonal MA respectively, Integrated component of the non-seasonal component called regular difference and D means the integrated order of the seasonal component called the seasonal difference (Makridaki *at al*.,1998). SARIMA (p, d, q) (P, D, Q) can be expressed by:

$$\phi(B)\Phi(B^{12})(1-B^{12})^D(1-B)^d Z_t=\theta(B)\Theta(B^{12})\alpha_t \quad (5)$$

Which $\theta(B)=1-\theta_1 B-\ldots-\theta_q B^q$;

$\phi(B)=1-\phi_1 B-\ldots-\phi_p B^p$;

$\Phi(B^s)=1-\Phi_1 B^s-\ldots-\Phi_P B^{sP}$ This is the seasonal autoregressive operator of order P and season s;

$(1-B^{12})^D$ This is the operator seasonal difference with D indicating the number of seasonal differences;

$\Theta(B^s)=1-\Theta_1 B^s-\ldots-\Theta_Q B^{sQ}$ This is the seasonal moving average of order Q and season s.

Model evaluation parameters

The use of ARIMA allows obtaining several models that represent the time series. Therefore, we must choose the best model to perform the forecast. There are several types of measures that serve to compare models, scale-dependent measures, error-based measures, measures based on relative errors and relative measures (Hyndman; Koehler, 2006). The most used measures based on relative errors are: Mean Absolute Percentage Error (MAPE), Median Absolute Percentage Error (MdAPE), Root Mean Square Percentage Error (RMSPE) and Root Median Square Percentage Error (RMdSPE), calculated according to Equations 6, 7, 8 and 9.

$$MAPE = Mean(|p_t|) \quad (6)$$

$$MdAPE = Median(|p_t|) \quad (7)$$

$$RMSPE = \sqrt{Mean(p_t^2)} \quad (8)$$

$$RMdSPE = \sqrt{Median(p_t^2)} \quad (9)$$

Where $p_t = \frac{100e_t}{Y_t}$ and e_t is the model's error

According to Hyndman and Koehler (2006), if all data is positive and much larger than zero, MAPE should be used for simplicity.

MATERIALS AND METHODS

Stages of the Box-Jenkins methodology

Box *et al.* (2016) states that there are three steps to build an ARIMA model:

1. Identification: it aims to identify, among the existing versions of models, which best describes the behavior of the series. The behaviour of autocorrelation functions (ACF) and partial autocorrelation functions (PACF) allow the estimation of the parameters of the model.

2. Estimation: it estimates the parameters Φ and φ of the auto-regressive component, the parameters θ and Θ of the moving-averages component and the variance of εt.

3. Verification: it is evaluated if the estimated model is adequate to describe the behavior of the data, evaluating waste and the error of the model.

If the model is not suitable, repeat the cycle by starting the identification phase. Only when a satisfactory model is obtained that it is possible to predict the time series.

The Figure 3 shows in more detail all steps according Box *et al.* (2016). We used the software Minitab to support our study.

CASE STUDY: FORECASTING OF BUFFALO MILK DEMAND

Identification of ARIMA model

Through data collection of the last 6 years, since January 2011, it was possible to obtain the demand curve for buffalo milk (raw material for manufacturing the final product), according to Figure 4. The data were available by the company by monthly demand for buffalo milk and in liters.

According to Bruselli and Carvalho (2002), the production of buffalo milk has a seasonality, its production is more concentrated in autumn and winter, and the breeding season is spring and summer, justifying the seasonality shown in Figure 4. Moreover, the day length influences in breeding activity, thus buffaloes calving during unfavorable season may effect he ovarian activity until the following favorable season (Terzano *et al.*, 2013). This justifies the seasonality of buffalo milk. It is also possible to observe trend, which confirms the previous information of growth of buffalo milk production.

Analysing Figure 5 and Figure 6, we can infer that the time series shows tendency, it is not stationary, besides showing the seasonality. The ACF values, in sinusoidal form with slow decay to zero, confirm that it is a non-stationary seasonal series. As well as the PACF graph, which decreases to zero, oscillating between positive and negative values, reaffirming the seasonality of the series.

To confirm the stationarity of the series, the Dickey-Fuller test was performed. The P-value of the test resulted in a value greater than 0.05, thus accepting the null hypothesis of existing unit root and consequently of the series being non-stationary.

Moreover, a periodogram was performed to determine the period of seasonality presented in the series. Figure 7 shows that the period of the series is 12 months.

Estimation of ARIMA model

Based on the graphs of ACF (Figure 5), PACF (Figure 6) and peridogram (Figure 7), the model parameters can be estimated using Minitab software. The model that best fits the original data series is the SARIMA $(2,1,4)(0,1,1)_{12}$. The model SARIMA $(2,1,4)(0,1,1)_{12}$ can be described in the Equation 10:

$$(1-\phi_1 B-\phi_2 B^2)(1-B)(1-B^{12})z_t=(1-\theta_1 B-\theta_2 B^2-\theta_3 B^3-\theta_4 B^4)(1-\Theta B^{12})a_t \quad (10)$$

The coefficients found can be seen in Table 1. In order to have a good mathematical model, it is ideal that the P-value of all coefficients is less than 0.05. Thus, Equation 11 represents the adjusted model of monthly demand of buffalo milk.

Verification ARIMA model

The verification of the model was based on analysis of residuals. Figure 9 confirms that the residues are normal (normality P-value = 0.210), homoscedastic and random (correlation p-value = 0.592). In this sense, it can be inferred that the noises behave as white noises and are not self-correlated.

In addition to the graphs presented, which show that the residues are white, it can be observed that ACF and PACF of residuals present random behaviors. Figures 9 and 10 show the ACF and

PAC of the residuals respectively.

For the model chosen, the model MAPE error was approximately 6.675%, which represents a small error.

Another way of analyzing the model is through the modified Box-Pierce test, also known as Ljung-Box, which does not reject for any k-phase the hypothesis of uncorrelated errors. Thus, all coefficients must have p-values greater than 0.05. Table 2 presents the results of the Ljung-Box test.

Forecasting using ARIMA model

Equation 10 allows predicting the milk demand in the first three months of 2017. The historical series presents data from January 2011 to December 2016. Thus, January, February and March were foreseen. Figure 11 presents the historical series with the prediction and 95.0% confidence interval.

The data from these predicted months are compared with actual milk demand values and shown in Table 3. The forecast results also have a lower and higher confidence limit, called confidence limits.

The predicted values approach the real values, being within the 95.0% confidence interval. The MAPE error for the forecast of these 3 months is 5.787%, considered a small error. Table 3 shows the difference in total milk demand and the forecast for the first three months of 2017. It is observed that the predicted and actual values approach. Nevertheless, a hypothesis test was performed to test the differences between the actual values and those forecasted. The P-value of 0.994 accepts, with 95.0% confidence, that there is no difference between actual and forecasted demand.

Thus, the forecast for the year 2017 can be determined using SARIMA $(2,1,4)(0,1,1)_{12}$. Figure 12 presents the forecast for the year 2017 with the 95.0% confidence interval. The demand forecast for buffalo milk for 2017 grew in relation to 2016. This is confirmed by performing the hypothesis test comparing the 2017 forecast with the 2016 demand, the P-value of 0.003 affirms with 95.0% confidence that demand will grow in 2017.

CONCLUSIONS

The modeling of the historical data of time series allowed observing that the buffalo milk demand in Brazil has been increasing in the last years and presents seasonality. Knowledge of these data made a better prediction possible, considering calving periods that decreases production and production periods that increases significantly the production.

We obtained the forecasting for the first three months of 2017 to confirm that the forecast represented reality through a hypothesis test. The prediction for the three-month period made possible an error of prediction MAPE = 5.787%. The model is SARIMA $(2,1,4)(0,1,1)_{12}$ and represents an equation where the residues behave normally and are controlled.

Moreover, we forecasted the demand for 2017 that showed an increase in the demand, compared to 2016. The hypothesis test allowed us to affirm that the demand for 2017 will increase in relation to the year 2016, not considering specific events that may modify this result. This forecast provided, not only for the company under study, but for buffalo milk producers the expected demand for 2017. Thus, the company and producers of buffalo milk in Brazil can expect the increase of the market and to plan in advance the scenarios that will face. It is recommended that the database be fed and Box-Jenkins steps retake for more accurate forecasts for future periods. In addition, ARIMA

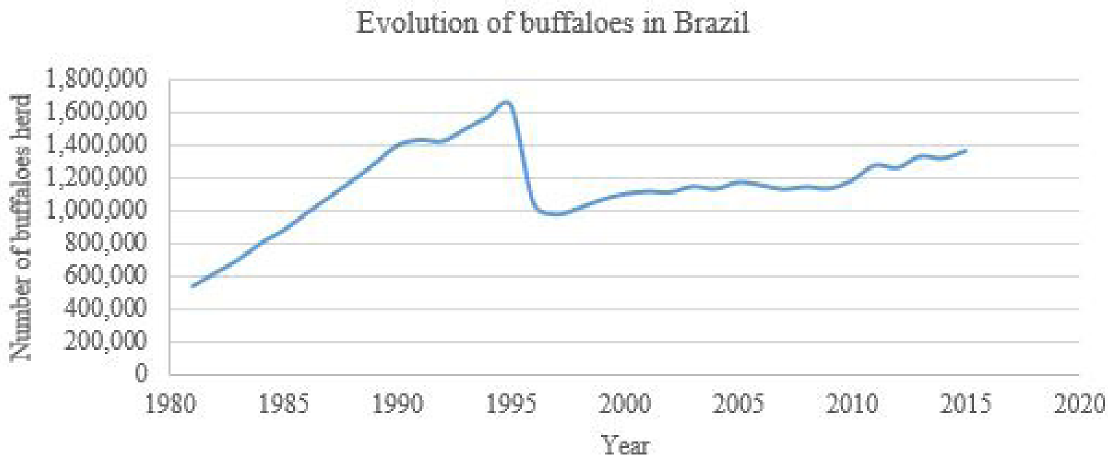

Figure 1. Evolution of the buffaloes herd (1981-2015), Source: IBGE (2016).

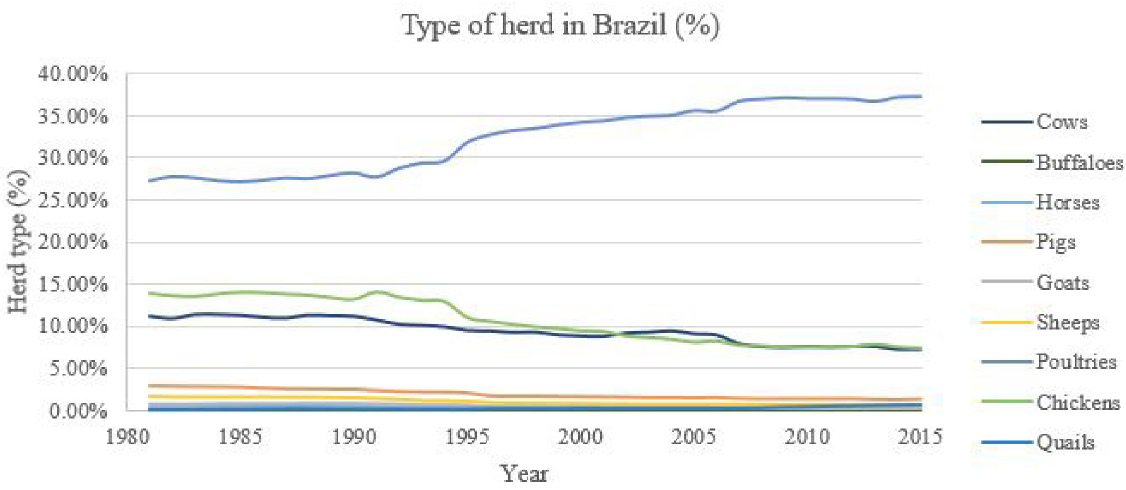

Figure 2. Percentage of livestock in Brazil (1981-2015), Source: IBGE (2016).

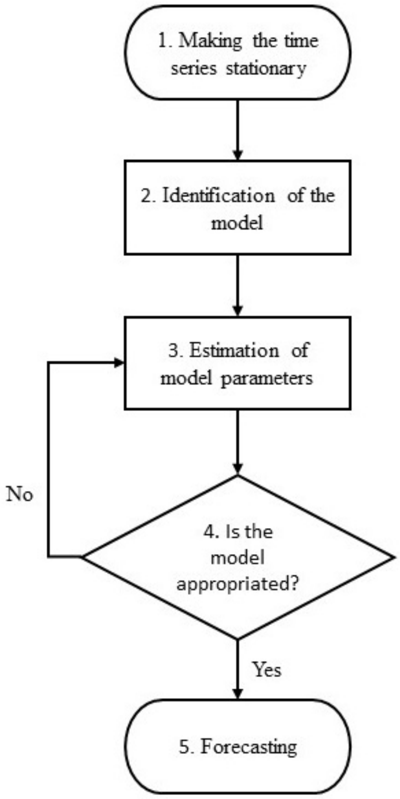

Figure 3. Flowchart of ARIMA models of Box & Jenkins.

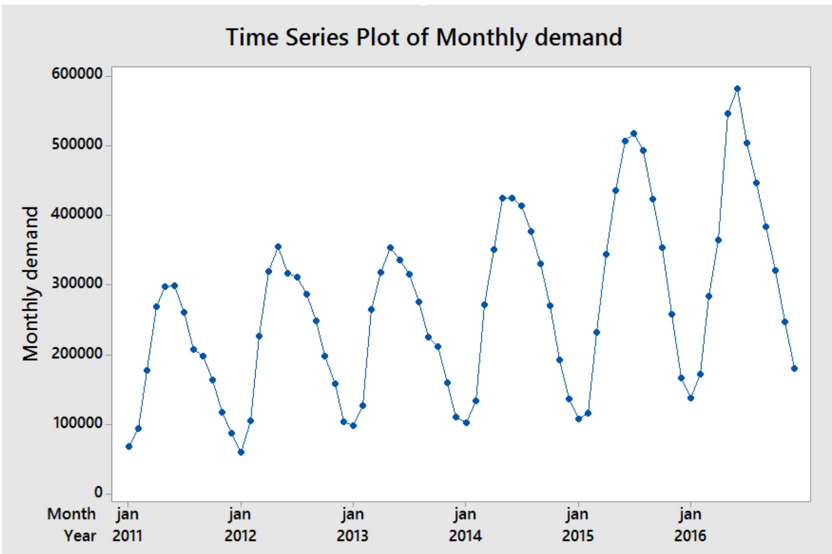

Figura 4. Time series of demand monthly of buffalo milk in litres.

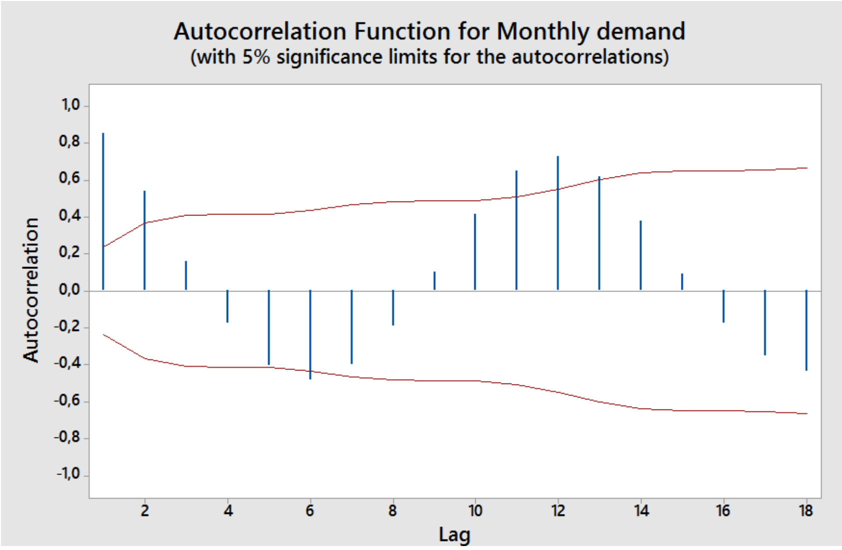

Figure 5. ACF of demand monthly of buffalo milk.

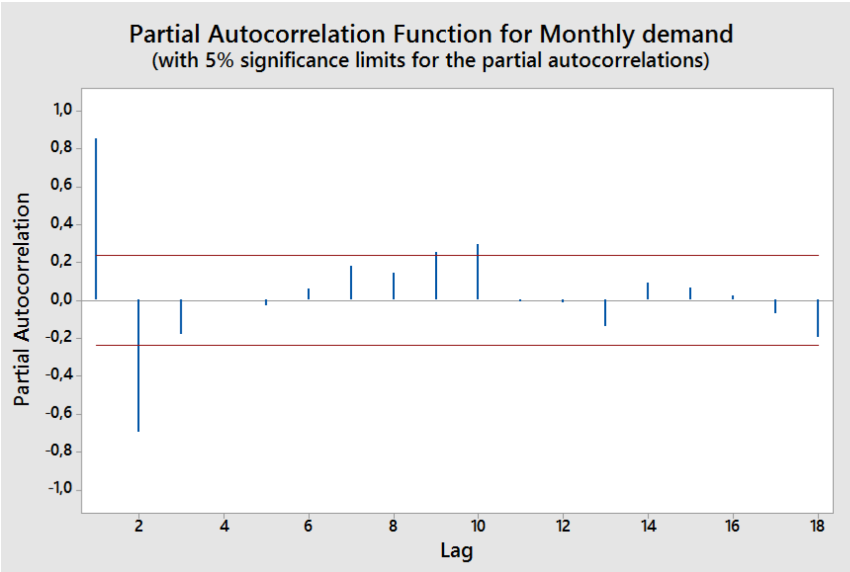

Figure 6. PACF of demand monthly of buffalo milk.

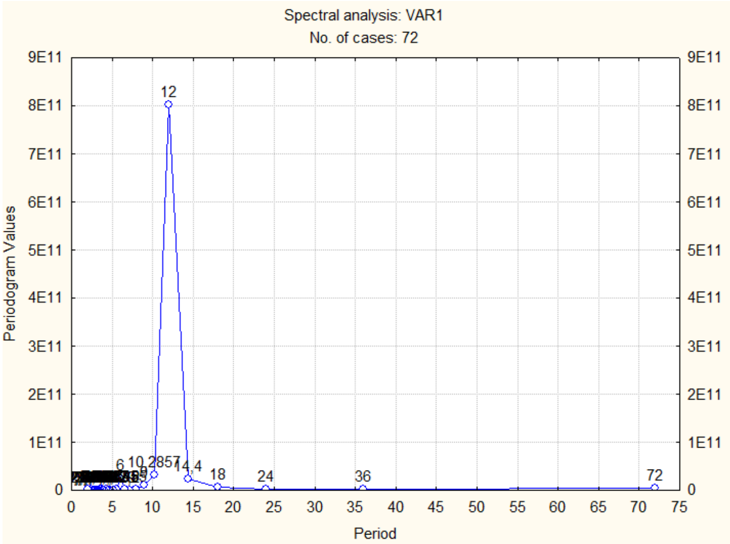

Figure 7. Periodogram.

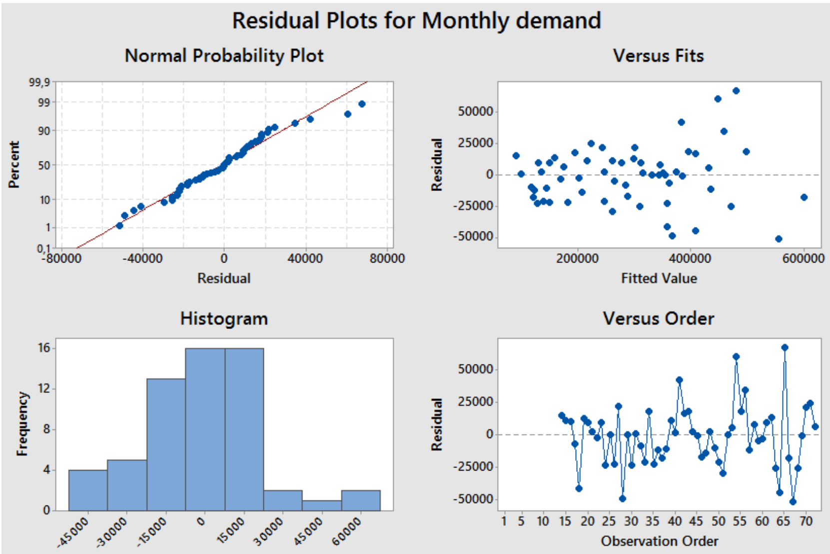

Figure 8. Residuals plots.

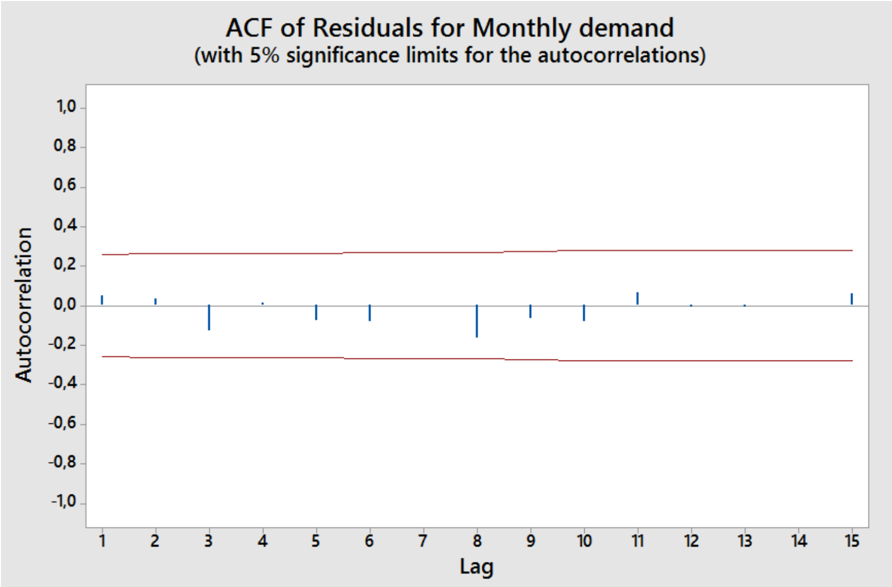

Figure 9. ACF of residuals.

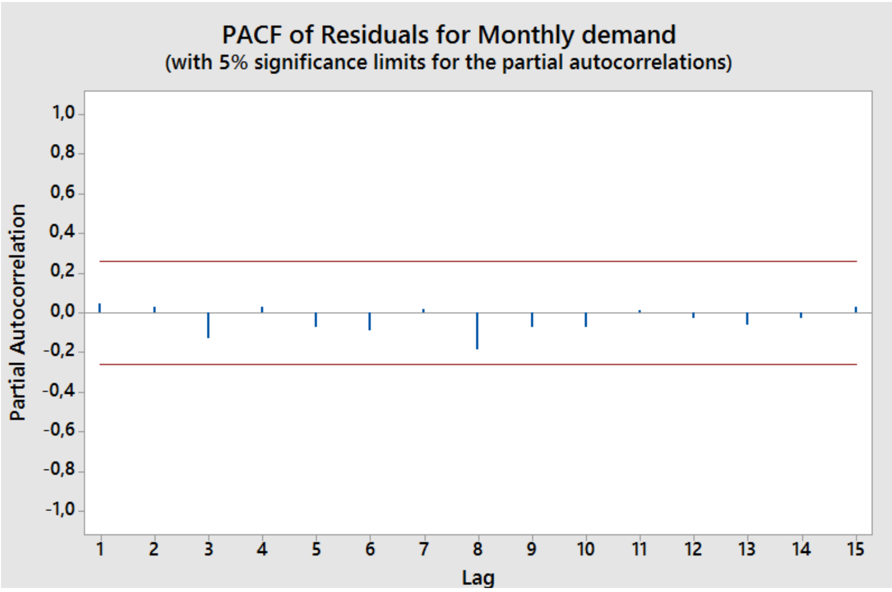

Figure 10. PACF of residuals.

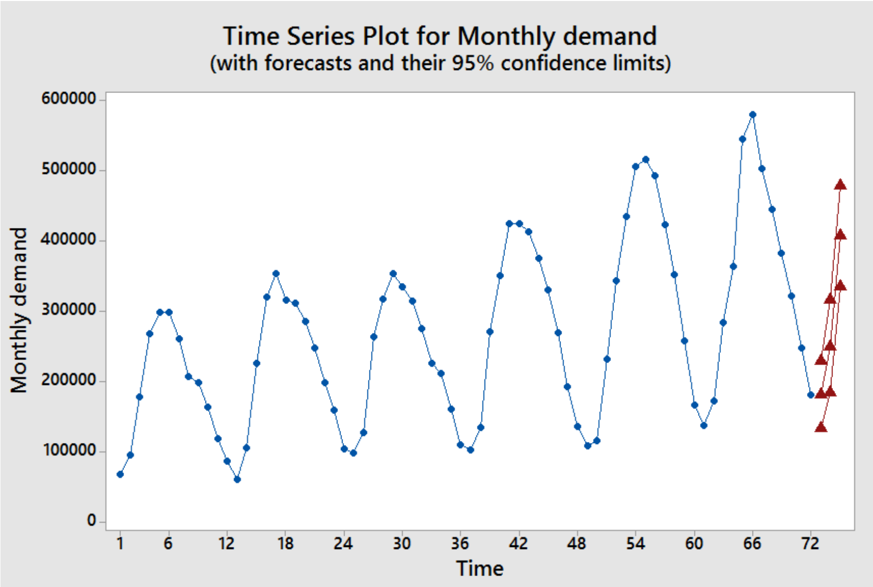

Figure 11. Forecasting three-months period for demand monthly of buffalo milk in litres.

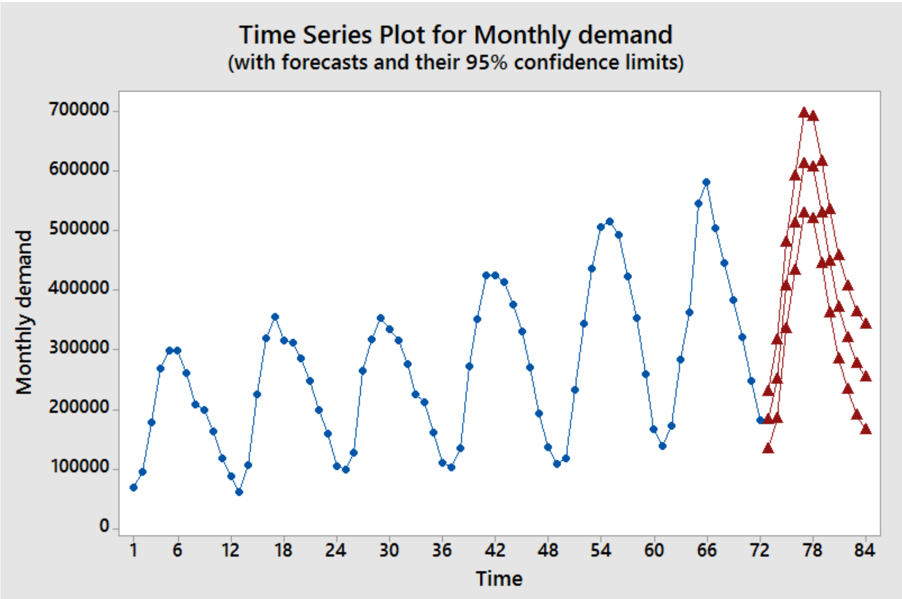

Figure 12. Previsão da demanda de leite mensal para o ano de 2017.

Table 1. Estimation of parameters for ARIMA model.

Type	Coef	SE	Coef	T	P
AR	1	1.6324	0.0581	28.10	0.000
AR	2	-0.9962	0.053	-18.78	0.000
MA	1	1.6911	0.0628	26.92	0.000
MA	2	-0.7687	0.1565	-4.91	0.000
MA	3	-0.5301	0.2293	-2.31	0.025
MA	4	0.5074	0.1492	3.4	0.001
SMA	12	0.6337	0.1788	3.54	0.001

Table 2. Modified Box-Pierce Test (Ljung-Box).

Lag	12	24	36	48
Chi-square	5.2	9.0	18.1	30.7
DF	5	17	29	41
P-value	0.396	0.940	0.941	0.881

Table 3. Forecasting three months period for demand monthly of buffalo milk in litres.

Period	Actual	Forecast	Lower	Upper
Jan/2017	175733	181018	133193	228843
Feb/2017	230256	250097	184416	315777
Mar/2017	432059	407277	335251	479302

model is suitable for predicting the demand of buffalo milk for the company under study, since it can reliably forecast and represent reality.

For future work, other prediction methods, such as neural networks or hybrid methods, can be performed and compared. It is also possible to carry out the same study for other companies to verify if the model identified in this study suits the demand for buffalo milk in other cities, companies and purposes.

REFERENCES

Associação Brasileira dos Criadores de Búfalos, ABCB. 2017. Available on http://www.bufalo.com.br/abcb.html. Accessed February 25, 2017.

Bastianetto, E. 2009. Criação de búfalos no Brasil: Situação e perspectiva. *Revista Brasileira de Reprodução Animal*, **6**: 98-103.

Box, G.E.P., G.M. Jenkins, G.C. Reinsel and G.M. Ljung. 2016. *Time Series Analysis: Forecasting and Control*, 5th ed. John Wiley and Sons, Hoboken, Ney Jersey, USA. 712p.

Delurgio, S.A. 1998. *Forecasting Principles and Applications*, 1st ed. McGraw-Hill, Singapore. 802 p.

Gregorz, L., R.S. Rossi, J.P.G. Mendes, N. Neuwirt, E.C. Marques, P.A. Melville and B.M. Monteiro. 2014. Ocorrência dos principais agentes bacterianos e parasitários em fezes diarreicas de bezerros búfalos nos estados de São Paulo e Paraná. *Arq. Inst. Biol.*, **81**(2): 180-185. DOI: 10.1590/1808-1657001122012.

Gujarati, D.N. 2000. *Econometria B*ásica. Makron Books, São Paulo, Brazil. 846p.

Jorge, A.M., A.G. Couto, G.A. Crudeli and E.M. Patiño. 2011. *Produção de B*úfalas de *Leite*. FEPAF, Botucatu, Brazil. 181p.

Instituto Brasileiro de Geografia e Estatística, IBGE. 2016. Avaliable on http://www.ibge.gov.br/home. Accessed in February 25, 2016.

Makridakis, S., S.C. Wheelwright and R.J. Hzdman. 1998. *Forecasting Methods and Applications*, 3th ed. John Wiley and Sons, New Jersey, USA. 642p.

Montgomery, D.C., C.L. Jennings and M. Kulahci. 2008. *Introduction to Time Series Analysis and Forecasting*, 1st ed. John Wiley and Sons, New Jersey, USA. 469p.

Morretin, P.A. and C.M.C. Toloi. 2014. *Previsão de S*éries *Temporais*, 2nd ed. Atual Editora, São Paulo, São Paulo, Brazil.

Terzano, M.G., M. Maschio, O. Barbato and V.L. Barile. 2013. Pre-ovulatory follicle size, LH peak values and pregnancy induced by three synchronization treatments in buffalo cows, during non-breeding season. *Buffalo Bull*., **32**(Special Issue 2): 513-516.

Wanapat, M. and S. Kang. 2013. World buffalo production: Challenges in meat and milk production, and mitigation of methane emission. *Buffalo Bull*., **32**(1): 1-21.