Modeling and analysis of a time series of equine infectious anemia cases in the state of Tocantins, Brazil, between 2007 and 2019

Modelagem e análise da série temporal dos casos de anemia infecciosa equina em equídeos no estado do Tocantins, Brasil, entre 2007 e 2019

Alessandro José Ferreira dos Santos[1*]; Jardel Martins Ferreira[2]; Francisco Baptista[3]; Marco Augusto Giannoccaro da Silva[4]; Bruna Alexandrino[5]; Ana Paula Coelho Ribeiro[5]; Raydleno Mateus Tavares[6]; Katyane de Sousa Almeida[5]

Highlights:
Equine infectious anemia (EIA) is endemic and non-seasonal in the state of Tocantins.
The ARIMA model (2,1,1) predicted the occurrence of 276 cases of EIA in Tocantins.
The ARIMA model (2,1,1) had good predictive ability for this time series.
Predictive models allow the planning of control actions by the Official Veterinary Service.

Abstract

Equine infectious anemia (EIA) is a viral infectious disease that affects *Equidae* and is clinically characterized by intermittent fever, anemia, depression, emaciation, and edema. To evaluate disease dynamics in the state of Tocantins, Brazil, a time series of EIA cases in the period 2007–2019 was analyzed to describe the pattern of occurrence and define the autoregressive integrated by moving average (ARIMA) model best suited to make predictions of cases of this disease for the period 2020–2021. The modeling and statistical analysis of the time series were performed using R software. The ARIMA model (2,1,1) was evaluated by Holdout cross-validation, in which data from the periods 2007–2017 and 2018–2019 were used as training and test data, respectively. The analyses showed that EIA was endemic and non-seasonal in Tocantins. The ARIMA model (2,1,1) showed good predictive capacity adjusted for this time series. However, the prediction of 276 cases of EIA in Tocantins for the period 2020–2021 may vary depending on the demand for diagnostic tests for *Equidae* transportation and herd sanitation in farms considered infection foci. The ARIMA model helps predict the number of EIA cases in Tocantins and improves planning for disease control by the Official Veterinary Service.
Key words: ARIMA model. Equine infectious anemia. Forecasting. Time series. Tocantins

[1] Pesquisador, M.e, Núcleo de Estudos Avançados em Geoprocessamento e Estatística, Agência de Defesa Agropecuária do Estado do Tocantins, Adapec/TO, Araguaína, TO, Brasil. E-mail: dr.alessandro.vet@gmail.com

[2] Discente de Mestrado do Programa de Pós-Graduação em Sanidade Animal e Saúde Pública nos Trópicos, Universidade Federal do Tocantins, UFT, Araguaína, TO, Brasil. E-mail: jardelmf@gmail.com

[3] Prof. Dr., Departamento de Epidemiologia Veterinária, Higiene e Saúde Pública, UFT, Araguaína, TO, Brasil. E-mail: baptista@uft.edu.br

[4] Prof. Dr., Programa de Pós-graduação em Sanidade Animal e Saúde Pública nos Trópicos, UFT, Araguaína, TO, Brasil. E-mail: marcogiannoccaro@uft.edu.br

[5] Prof[as] Dr[as], Programa de Pós-graduação em Sanidade Animal e Saúde Pública nos Trópicos, UFT, Araguaína, TO, Brasil. E-mail: bralexandrino@mail.uft.edu.br; apcribeiro@hotmail.com; katyanesalmeida@uft.edu.br

[6] Médico Veterinário, Programa Estadual de Sanidade dos Equídeos, Agência de Defesa Agropecuária do Estado do Tocantins, Adapec/TO, Palmas, TO, Brasil. E-mail: raydleno85@gmail.com

* Author for correspondence

Resumo

A anemia infecciosa equina (AIE), doença infecciosa viral que acomete os equídeos, é caracterizada clinicamente por causar febre intermitente, anemia, depressão, emaciação e edema. Com o objetivo de elucidar a dinâmica dessa doença no estado do Tocantins, foi realizada a análise da série temporal dos casos de AIE em equídeos entre 2007 e 2019 para descrever o padrão de sua ocorrência, além de definir o modelo autorregressivo integrado por média móvel (*Autoregressive Integrated by Moving Average* - ARIMA) mais adequado para se realizar previsões dos casos dessa doença para os anos de 2020 e 2021. A modelagem e análise estatística da série temporal em estudo foi realizada por meio do *software* R. O modelo preditivo ARIMA(2,1,1) foi avaliado por meio da validação cruzada utilizando a técnica de *Holdout*, em que os dados de 2007 a 2017 foram utilizados como treino e os dados de 2018 e 2019 foram utilizados como teste. As análises mostraram que a AIE é endêmica no estado do Tocantins e sem padrão de sazonalidade. O modelo ARIMA(2,1,1) apresentou boa capacidade preditiva ajustada para a série temporal em estudo. Porém, a previsão aproximada de 276 casos de AIE em equídeos para os anos de 2020 e 2021 no estado do Tocantins pode variar em decorrência da demanda por exames dessa doença para o trânsito dos equídeos, bem como do saneamento de propriedades consideradas foco. A modelagem ARIMA pode ser utilizada na previsão dos casos de AIE em equídeos no estado do Tocantins o que permite melhorar o planejamento para a execução das ações de controle dessa doença por parte do Serviço Veterinário Oficial.

Palavras-chave: Anemia infecciosa equina. Modelagem ARIMA. Previsão. Série temporal. Tocantins.

Introduction

Equine infectious anemia (EIA) is an infectious disease that affects *Equidae* and is caused by viruses of the family *Retroviridae*, subfamily *Orthoretrovirinae*, genus *Lentivirus*. Clinical manifestations include intermittent fever, anemia, depression, emaciation, and edema. Hematophagous insects, especially *Tabanidae* family, transmit EIA; however, its transmission may also occur iatrogenicly (Van Der Kolk & Kroeze, 2013).

Given its transmissibility, incurability, and difficulty to control, EIA is a significant obstacle to horse farming in Brazil, causing economic losses to owners and breeders. For this reason, euthanasia of infected animals is mandatory and is carried out under the supervision of the Official Veterinary Service (Serviço Veterinário Oficial - SVO), although the isolation of seropositive animals is allowed in high-risk regions (Almeida et al., 2006).

In view of the social, cultural, and economic relevance of horse farming in the state of Tocantins, knowledge about the dynamics of EIA is crucial to identify the pattern of disease occurrence and provide data to the SVO to guide the implementation of sanitary measures for disease control. The objective of this study is to describe the pattern of occurrence of EIA in Tocantins and define an autoregressive integrated by moving average (ARIMA) model to predict the number of cases for the period 2020–2021 using time series 2007–2019.

Materials and Methods

Study site

The study was carried out in the state of Tocantins (longitude, 46–51° W, latitude, 5–13° S), in the north of Brazil. This state contains 139 municipalities and a total area of 277,620.914 km². The climate is semi-humid, divided into two zones, the Tropical Equatorial Zone in the north of the state and Tropical Central Brazil in the south, and is characterized by a rainy season in the summer (October to April) and a dry season in the winter (May to September). Temperatures vary from 25 to 40 °C, and the annual rainfall is above 1,500 mm (Nascimento, 2011).

Data collection

To analyze the pattern of disease occurrence, monthly data for the periods 2007–2010 and 2014–2017 were extracted from the interactive panel of the database of the National Zoosanitary Information System (Sistema Nacional de Informação Zoossanitária–ZIS) of the Ministry of Agriculture, Livestock, and Supply (Ministério da Agricultura, Pecuária e Abastecimento [MAPA], 2019). Data for the periods 2011–2013 and 2018–2020 were provided by the State Equine Health Program (Programa Estadual de Sanidade dos Equídeos–PESE) of the Agricultural and Livestock Defense Agency of the State of Tocantins (Agência de Defesa Agropecuária do Estado do Tocantins–Adapec/TO). A case means a sick or infected animal with a confirmed diagnosis (Brasil, 2019). Data were consolidated using spreadsheet Microsoft Excel® version 2013 and saved in "file.CSV" (separated by commas) format.

Modeling and statistical analysis

The modeling of the time series and statistical analyses were performed using R software version 3.6.0. A time series contains three components: trends, seasonality, and error (Morettin & Toloi, 2006):

$$Z_t = T_t + S_t + \alpha_t$$

where Z_t is the data for period t, T_t is the trend during period t, S_t is the seasonal component in period t, and α_t is the random component.

To determine the best predictive model, the time series (Z_t) was decomposed using the "decompose( )" function from R software, and its three components were analyzed individually. Trend (T_t) was analyzed by linear regression using the "lm( )" function from R software. The need to remove trends was evaluated using the "ndiffs ( )" function from R software. The presence of structural breaks in the time series was evaluated by the statistical method "OLS-MOSUM" using functions "Efp( )" and "sctest( )" from "strucchange" statistical package (Zeileis, Leisch, Hornik, & Kleiber, 2002).

The existence of seasonal component (S_t) in the time series was assessed using the functions "qs( )", "sesas( )", and "series( )" from "seasonal" statistical package of software "X-13 ARIMA- SEATS" (Sax and Eddelbuettel, 2018). The correction of seasonal adjustment for the variables "rainy period" (October to April) and "large equestrian events" (May to September) was performed using the "as. Date( )" function from R software and the function "Genhol( )" from "seasonal" package of software "X-13 ARIMA-SEATS" (Sax and Eddelbuettel, 2018) to assess the influence of these variables on the seasonality of EIA in Tocantins.

The ARIMA model was defined based on the Box, Jenkins, Reinsel and Ljung (2016) methodology for stationary time series, as described below:

Specification: analysis of the general class of ARIMA structures (p,d,q)(P,D,Q), where "p" is the order of the non-seasonal autoregressive polynomial, "P" is the order of the seasonal autoregressive polynomial, "d" is the order of non-seasonal differentiation, "D" is the order of seasonal differentiation, "q" is the order of the non-seasonal moving average polynomial, and "Q" is the order of the seasonal moving average polynomial.

Identification: p and q values were defined based on the autocorrelation function (ACF) and partial autocorrelation function (PACF). An ARIMA process (p,d,q) is an autoregressive and moving average model differenced "d" times until it becomes stationary.

Estimation: the model parameters were tested for statistical significance using the "t_test( )" function from "BETS" package (Ferreira, Speranza, & Costa, 2018).

Diagnosis: the analysis of residuals and Ljung-Box tests were carried out using the functions "tsdiag( )", "Box.test( )", and "shapiro.test( )"

3147

from R software to determine whether the model was suitable for making predictions.

Prediction: the number of EIA cases was predicted using the function "forecast( )" from the "forecast" statistical package (Hyndman et al., 2019).

Holdout cross-validation was performed using the "window( )" function from R software to determine the predictive accuracy and generalization capacity of the ARIMA model, in which data from the periods 2007–2017 and 2018–2019 were used as training and test data, respectively, considering Theil's U statistic.

Results

Between 2007 and 2019, there were 3,754 EIA cases in Tocantins, and linear regression analysis showed a slight decreasing trend ($y = -1.2402x + 2,521.1330$, $p < 0.05$, $R^2 = 0.053$). There were no structural breaks in the time series since the null hypothesis (H_0) of empirical fluctuation was not rejected ($S = 0.21711$, $p > 0.05$).

Seasonality was determined for the complete time series (2007–2019) and for the latest eight years because the series had more than 96 observations. The analysis of the original time series (qsori) demonstrated that EIA was not seasonal in Tocantins ($p > 0.05$). The correction of automatic seasonal adjustment (qsorievadj) demonstrated that the rainy season and large equestrian events did not significantly affect the pattern of occurrence of EIA ($p > 0.05$); however, average occurrence of cases of this disease was higher in February and March.

The results of the Zivot-Andrews unit root test or stationarity test showed that the original time series needed to be differenced to become stationary. Therefore, the non-seasonal differentiation order (d) was defined as equal to one.

The ACF and PACF correlograms provide an indication of the specification and identification of the ARIMA model to be used (Figure 1). Both correlograms showed an exponential decrease in significant parameters in the first two lags of the autoregressive (p) and moving average (q) polynomials of the differenced time series, respectively (Figures 1A and B). Therefore, the model (2,1,1) was estimated and confirmed using significance and residual analyses (Figures 1C and D) since all lags were within the confidence interval.
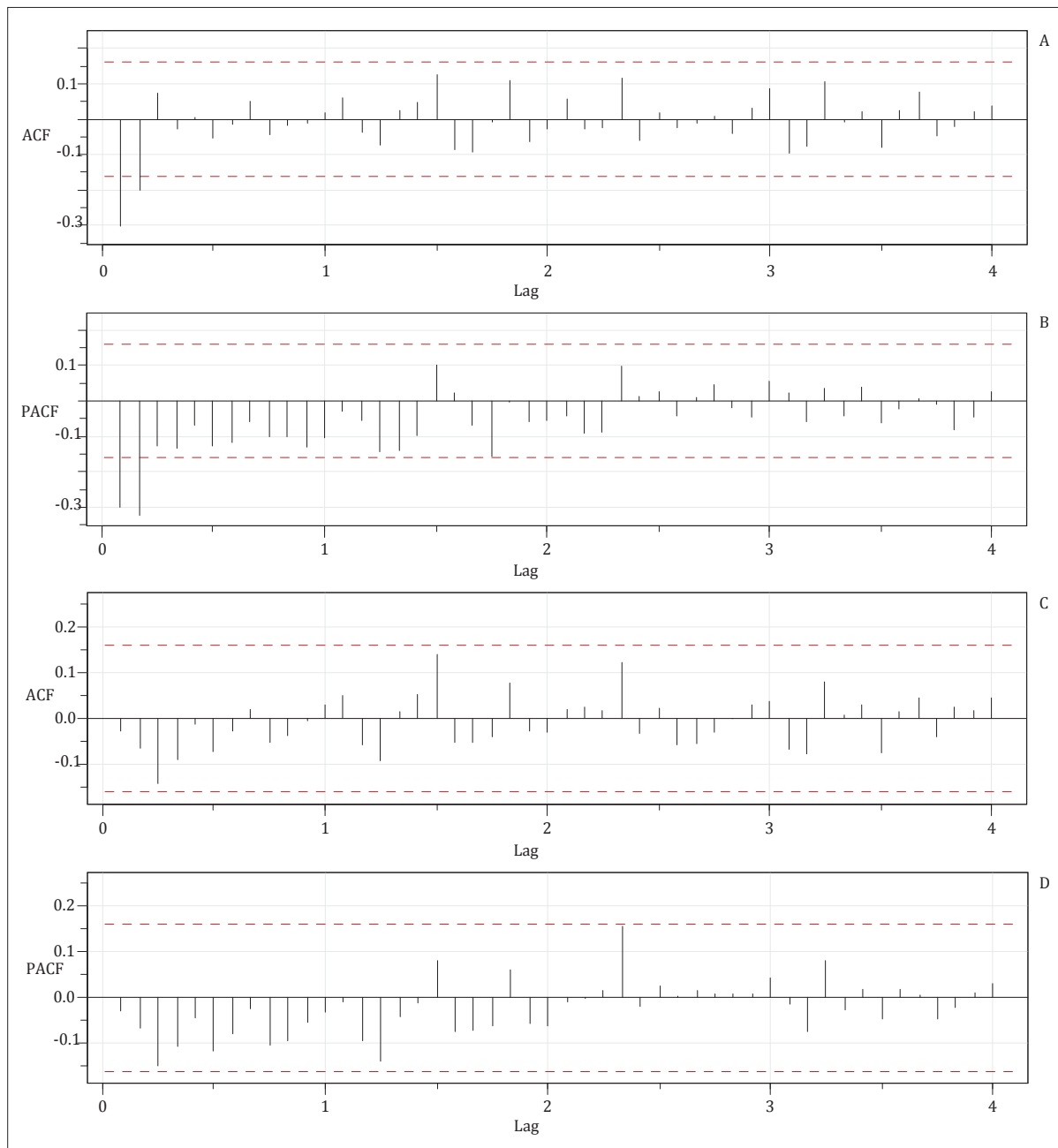
Figure 1. A- Correlogram of autocorrelation function (ACF); B - Correlogram of partial autocorrelation function (PACF); C - Correlogram of residuals of the ACF of the ARIMA model (2,1,1); Correlogram of residuals of the PACF of the ARIMA model (2,1,1).

A diagnosis was performed to confirm the characteristics of model residuals (2,1,1), including the absence of linear autocorrelation and normality. Data distribution was non-normal according to the Shapiro-Wilk normality test (W = 0.90716, p < 0.05) because of a discrepant variation around the zero average. There was no linear autocorrelation in residuals because significant lags were absent, i.e., the linear section of the time series was captured by the ARIMA model (2,1,1). Furthermore, the

results of the Ljung-Box test (χ^2 = 16.412, degrees of freedom = 23, p > 0.05) showed that it was not possible to reject the null hypothesis (H_0) of the absence of serial dependency for different lags.

Theil's U statistics was 0.69 when training and test data were used to assess the predictive accuracy of the model (2,1,1). For the period 2020–2021, this model predicted the occurrence of approximately 276 cases of EIA in Tocantins (Table 1). Furthermore, all cases reported between January and July 2020 were within the 95% confidence interval of the predictions in the respective months, demonstrating that the model had a good short-term predictive capacity (Table 1).

Table 1
Prediction of equine infectious anemia cases in *Equidae* in the state of Tocantins, Brazil, for the period 2020-2021 by ARIMA prediction model (2,1,1)

Month	2020			2021		
	Predicted*	Reported	95% CI	Predicted*	Reported	95% CI
Jan	14	7	1.30 – 41.30	11	N/A	0.59 – 40.48
Feb	13	8	0.90 – 42.05	11	N/A	0.56 – 40.38
Mar	12	10	0.83 – 41.53	11	N/A	0.54 – 40.27
Apr	12	4	0.81 – 41.39	11	N/A	0.52 – 40.17
May	12	14	0.79 – 41.31	11	N/A	0.50 – 40.06
Jun	12	4	0.76 – 41.21	11	N/A	0.48 – 40.00
Jul	12	6	0.73 – 41.11	11	N/A	0.46 – 39.84
Aug	12	N/A	0.71 – 40.01	11	N/A	0.44 – 39.74
Sep	12	N/A	0.68 – 40.90	11	N/A	0.43 – 39.63
Oct	12	N/A	0.66 – 40.80	10	N/A	0.41 – 39.52
Nov	12	N/A	0.63 – 40.70	10	N/A	0.39 – 39.41
Dec	12	N/A	0.61 – 40.59	10	N/A	0.38 – 39.30
Total	147		-	129		-

Score 1: (*) Approximate figures.
Score 2: N/A: Non-available record.

Discussion

The number of EIA cases in Tocantins showed a decreasing trend in the study period; however, this decrease was not sufficient to produce structural breaks in the data series, which allowed considering EIA as endemic in this state, corroborating the study by Costa (2018), who performed a temporal analysis of EIA cases in Brazil between 2005 and 2016 and found that this disease was endemic in the country and the number of cases tended to stabilize. Between 2007 and 2011 and from the second semester of 2016 until the end of 2019, the number of EIA cases decreased in Tocantins. Nonetheless, whether official sanitary measures influence this reduction is unknown because the diagnosis of EIA is associated with *Equidae* transport.

Moreover, the results showed that EIA was not seasonal in Tocantins, corroborating the study by Baptista et al. (2016) using an EIA time series in the state of Rio de Janeiro between 2007 and 2011. In addition, the rainy season (October to April) and large equestrian events (May to September) did not significantly affect the pattern of disease occurrence in the study period. Notwithstanding non-seasonal, the data showed that February and March presented the highest averages of occurrence of EIA in

Equidae in the Tocantins. Ferreira-Keppler, Rafael, & Guerrero (2010) demonstrated that the abundance and richness of *Tabanus* spp. were higher in forest areas and clearings in Central Amazon in drier and warmer months (July to December), and the occurrence of this vector peaked in October. When comparing this finding with the months with the highest average occurrence of EIA in this time series (February and March), the need to better understand vector cycle and correlate these results with the pattern of occurrence of EIA in the state was underscored.

Morettin & Toloi (2006) have shown that a time series is any set of observations ordered in time and that the Box & Jenkins methodology consists of fitting ARIMA models to a dataset with a stationary characteristic, which develops in time randomly around a constant average. For this reason, the time series under study was differenced to become stationary, which allowed the developing an ARIMA model suitable for making predictions based on historical data.

The critical phase in the elaboration of a model is identification, and it is possible that several researchers identify different models for the same time series (Morettin & Toloi, 2006). However, predictions using ARIMA models are more accurate than other prediction methods because of the small number of parameters analyzed. These authors have shown that the Box & Jenkins technique requires experience and technical knowledge beyond the use of software. Therefore, the ARIMA model (2,1,1) identified for the time series dataset under study was well specified and met the criterion of the absence of residual autocorrelation and to be used in predictions of EIA cases, as demonstrated in the diagnostic stage.

Ehlers (2007) have shown that the predictive capacity of an ARIMA model can be determined by comparing its prediction errors with those of the random walk model using Theil's U statistics, which has a good predictive ability when the values are less than one. This characteristic was observed in the ARIMA model (2,1,1) adjusted for the time series under study. However, the prediction of 276 cases of EIA in Tocantins for the period 2020–2021 may vary depending on the demand for diagnostic tests for *Equidae* transportation and herd sanitation in farms considered infection foci.

Helfenstein (1996) have shown that time series forecasts are important to assist health services because, once they demonstrate the expected frequency of diseases, they help improve the planning of the distribution of resources to prevent and combat diseases. Morettin & Toloi (2006) point out that predictions are used to provide information for decision-making. Therefore, previous knowledge about the dynamics of EIA in Tocantins may allow implementing and improving risk mitigation strategies. Among these are sanitary inspections during transportation of equines to ensure compliance with sanitary regulations and during large equestrian events, as *vaquejadas*, lasso events, horseback riding, and other equestrian sports are traditionally hold in the state.

Nogueira et al. (2017) found that systematic testing could reduce the number of EIA-positive *Equidae* in farms 41-fold. In turn, Baptista et al. (2016) have shown that the absence of veterinarians working in horse farms may increase the risk of EIA, limiting disease control. Therefore, the scope of equestrian activities in Tocantins warrants the development of guidelines for improving sanitary inspections in horse farms, including the requirement of negative results for all housed animals and veterinary technical assistance.

Given that EIA is endemic in Tocantins based on the analysis of this time series, sanitary measures to control EIA in draft horses from periurban areas is fundamental, given that Chaves et al. (2014) found that one of the risk factors for EIA transmission is the permanence of *Equidae* in agglomerations for more than eight hours. This is a factor that draws attention since many of these draft horses

also participate in large equestrian events such as horseback riding, which take place across state of Tocantins at different times of the year. Almeida et al. (2006) pointed out that the effective control of EIA in draft horses avoids compromising control measures for animals of high commercial value because the presence of infection foci increases the risk for euthanasia.

The development and implementation of health education programs for owners, breeders and/or farm workers is part of control actions based on the knowledge of the epidemiology of EIA, as pointed out in this study. Moreover, since euthanasia is used to eliminate the source of infection, Almeida et al. (2006) suggest creating a fund with the participation of the private sector to reimburse breeders and owners who would feel encouraged to sanitize their *Equidae* herds.

Conclusion

EIA is endemic and non-seasonal in Tocantins. ARIMA modeling technique proved to be adequate to perform forecasts from the studied time series, proving to be a useful tool in decision making since it allows the improvement of actions to control this disease by the Official Veterinary Service. Notwithstanding, additional studies are necessary to determine the prevalence of this disease, risk factors for transmission, as well as determining ecology of vectors such as seasonality and abundance of the tabanid species identified in the different habitats of the state of Tocantins (Brazil).

Acknowledgments

References

Almeida, V. M. A., Gonçalves, V. S. P., Martins, M. F., Haddad, J. P. A., Dias, R. A., Leite, R. C., & Reis J. K. P. (2006). Anemia infecciosa equina: prevalência em equídeos de serviço em Minas Gerais. *Arquivo Brasileiro Medicina Veterinária Zootecnia*, *58*(2), 141-148. doi: 10.1590/S0102-09352006000200001

Baptista, D. Q., Bruhn, F. R. P., Rocha, C. M. B. M., Torres, F. C., Machado, E. D., Sáfadi T., & Pereira S. M. (2016). Temporal series analyses in equine infecciosa anemia cases in the State of Rio de Janeiro, Brazil, 2007 to 2011. *Revista Brasileira de Medicina Veterinária*, *38*(4), 431-438.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time Series Analysis: forecasting and control* (5th ed.). New Jersey: John Wiley & Sons Inc.

Chaves, N. P., Bezerra, D. C., Santos, H. P., Pereira, H. M., Guerra, P. C., & Silva, A. L. A. (2014). Ocorrência e fatores de risco associados à identificação da anemia infecciosa equina em equídeos de tração. *Ciência Animal Brasileira, 15(*3), 301-306. doi: 10.1590/1809-6891v15i318395

Costa, A. M. P. S. (2018). *Análise temporal da ocorrência da anemia infecciosa equina no Brasil no período de 2005 a 2016*. Dissertação de mestrado, Universidade Estadual Paulista, São Paulo, SP, Brasil.

Ehlers, R. S. (2007). *Análise de séries temporais*. Cidade: Curitiba. Departamento de Estatística, Universidade Federal do Paraná. Retrieved from http://www.each.usp.br/rvicente/AnaliseDeSeries Temporais.pdf

Ferreira, P. C., Speranza, T., & Costa, J. (2018). *Brazilian economic time series (BETS), R package, version 0.4.9*. Retrieved from http://127.0.0.1:12355/library/BETS/DESCRIPTION

Ferreira-Keppler, R., Rafael, J. A., & Guerrero, J. C. H. (2010). Sazonalidade e uso de ambientes por espécies de *Tabanidae* (diptera) na Amazônia Central, Brasil. *Neotropical Entomology*, *39*(4), 645-65. doi: 10.1590/S1519-566X2010000400028

Helfenstein, U. (1996). Box-Jenkis modelling in medical research. *Statistical Methods in Medical Research, 5*(1), 3-22. doi: 10.1177/096228029600500102.

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M.,… Yasmeen, F. (2019). *Forecast: forecasting functions for time series and linear models*. R package, version 8.5. Retrieved from http://pkg.robjhyndman.com/forecast

Ministério da Agricultura, Pecuária e Abastecimento. (2019). Coordenação de informação e epidemiologia em saúde animal. Ministério da Agricultura, Pecuária e Abastecimento. Retrieved from http:// indicadores. agricultura.gov.br/saudeanimal/index.htm

Morettin, P. A., & Toloi C. M. C. (2006). *Análise de series temporais* (2a ed. rev. e ampl.). São Paulo: Edgard Blücher.

Nascimento, J. B. (2011). *Tocantins: história e geografia* (7a ed.). Goiânia: Bandeirante.

Nogueira, M. F., Oliveira, J. M., Santos, C. J. S., Petzold, H. V., Aguiar, D. M., Juliano, R. S.,... Abreu, U. G. P. (2017). Equine infectious anaemia in equids of Southern Pantanal, Brazil: seroprevalence and evaluation of the adoption of a control programme. *Pesquisa Veterinária Brasileira*, *37*(3), 227-233.

Sax, C., & Eddelbuettel, D. (2018). Seasonal adjustment by X-13ARIMA-SEATS in R. *Journal of Statistical Software*, *87*(11), 1-17. doi: 10.18637/jss.v087.i11

Van Der Kolk, J. H., & Kroeze, E. J. B. V. (2013). *Infectious diseases of the horse: diagnosis, pathology, management, and public health*. London: Mason Publishing.

Zeileis, A., Leisch, F., Hornik, K., & Kleiber, C. (2002). Strucchange: an R package for testing for structural change in linear regression models. *Journal of Statistical Software*, *7*(2), 1-38. doi: 10.18637/jss. v007. i02