



## **APLICAÇÃO DE MODELOS UNIVARIADOS E REDE NEURAL FEED-FORWARD NA DEMANDA TURÍSTICA INTERNACIONAL DO ESTADO DE MATO GROSSO DO SUL, BRASIL**

### **APPLYING UNIVARIATE MODELS AND FEED-FORWARD NEURAL NETWORK ON INTERNATIONAL TOURISTIC DEMAND IN MATO GROSSO DO SUL STATE, BRAZIL**

Bruno Matos Porto, Universidade Federal de Mato Grosso do Sul/CPAQ, Brasil, bmatosporto@gmail.com

Dra. Daniela Althoff Philippi, Universidade Federal de Mato Grosso do Sul/CPAQ, Brasil, daniela\_philippi@yahoo.com.br

Vanessa Aline Wagner Leite, Universidade Federal de Mato Grosso do Sul/PPGEO-CPAQ, Brasil, vanessawleite@gmail.com

#### **Resumo**

O objetivo deste trabalho foi gerar previsões e comparar a precisão preditiva no curto prazo de modelos univariados e rede neural (NNAR), na previsão da demanda de turismo do estado de Mato Grosso do Sul (MS). Foram utilizadas as séries temporais ARIMA, Holt-Winters (HW) versões aditiva e multiplicativa e NNAR com uso do software R, com base nos dados da ferramenta base de dados extrator do Instituto Brasileiro de Turismo (Embratur). Os dados são referentes às chegadas turísticas por todas as vias registradas no MS entre janeiro de 2007 a setembro de 2016. As previsões da demanda turística abrangeram o intervalo de outubro a dezembro de 2016, sendo então comparadas aos dados reais do mesmo período. As previsões dos modelos foram comparadas por meio dos critérios MAPE, MAD e MSD, onde os resultados gerados pelas medidas de erros mostraram que a (NNAR) é altamente precisa, assim superou todos os outros modelos e em segundo lugar no ranking de acuracidade destacou-se HW a versão multiplicativa. Para pesquisas futuras sugerimos análises de previsões fora da amostra de (1, 2, 4, 6, 12 e 24 meses) de um vasto número de séries temporais.

**Palavras-chave:** Previsão; Séries Temporais; Medidas de erros; Acuracidade.

#### **Abstract**

This work aimed at forecasting as well as comparing predictive accuracy in short term univariate models and neural network (NNAR), when predicting the tourism demand in the State of Mato Grosso do Sul (MS). It was used temporal series ARIMA, Holt-Winters (HW) additive and multiplicative versions as well as NNAR using R software based on the data acquired in the Brazilian Institute of Tourism (Embratur). The data refer to touristic arrivals by all registered routes in the Mato Grosso do Sul State between January 2007 and September 2016. The touristic demand forecasting covers a range from October to December 2016 then compared with actual data from the same period. The forecast models were compared by MAPE, MAD and MSD criteria in which results generated by errors measurement have shown that (NNAR) is highly accurate surpassing all other models and secondly in the accuracy ranking, multiplicative version of HW was highlighted. We suggest to future studies analyzing forecasts out of the samples (1, 2, 4, 6, 12 and 24 months) in a wide number of time series.

**Keywords:** Forecast; Time Series; Error Measurement; Accuracy.

## 1. INTRODUÇÃO

As previsões automáticas de um amplo número de séries cronológicas univariadas são frequentemente indispensáveis nos negócios (Hyndaman & Khandakar, 2008). A atividade turística necessita de um planejamento de previsão de demanda turística (CHU, 1998; SERRA, TAVARES, & SANTOS, 2005). Neste enfoque é consenso que planejamento relaciona-se com as decisões que serão tomadas e afetará o futuro (GOH & LAW, 2002).

As previsões de curto prazo são relevantes para vários objetivos de gestão como bem explicado por Gunter e Önder (2015, p. 123), “previsão de curto prazo da previsão da demanda nos próximos dois ou três meses pode ajudar o destino a ter mais flexibilidade operacional, por exemplo, em termos de número de ônibus do aeroporto para o centro da cidade”. Diante disto os serviços turísticos necessitam de um planejamento preditivo que é obtido via predição acurada da demanda turística (GOH, LAW, & MOK, 2008). Outro aspecto significativo de dados da demanda turística é a natureza da sazonalidade que deve ser identificada antes da seleção do modelo de previsão ideal que considere a sazonalidade (GOH & LAW, 2002).

O estudo é importante devido aos impactos do fenômeno turístico na economia do Brasil e sua parte considerável no desenvolvimento econômico e social através de geração de empregos diretos, indiretos e renda (Casanova, Araujo, Silva, & Rocha, 2012), tendo em vista que a indústria do turismo no Brasil, com base nos dados disponíveis pela Embratur, representa 9% do PIB, assim gerou direta e indiretamente mais de 7,5 milhões de empregos no Brasil, além disso, mantém milhões de micro e pequenas empresas de muitos segmentos da economia (EMBRATUR, 2017d).

A previsão de demanda do turismo é um assunto significativo para os praticantes e acadêmicos (SONG & LI, 2008; Kim, Wong, & George Athanasopoulos, 2011; GUNTER & ÖNDER, 2015). A predição de demanda turística desempenha um papel fundamental no planejamento turístico uma vez que os projetos de investimento no turismo que utilizam altos recursos financeiros precisam de uma previsão precisa do que irá ocorrer futuramente com a demanda (CHO, 2003). É amplamente concordado que as regiões específicas precisam de previsões precisas da procura do turismo, mas existem poucas pesquisas neste sentido devido à falta de informação estatística (CLAVERIA & TORRA, 2014).

A pesquisa de (CHU, 1998) aplicou os modelos ARIMA sazonal e não sazonal, Holt-Winters nos dados de chegada de turistas nos países da Ásia-Pacífico. Os resultados revelaram que a previsão no curto ou médio prazo de ARIMA sazonal e não sazonal foram os mais acurados dos modelos testados, jugado pelo critério MAPE, assim, testou-se a acuracidade dos modelos ARIMA sazonal e não sazonal, Holt-Winters, por meio, de MAPE e receberam um percentual abaixo de 10%.

No estudo de (GOH & LAW, 2002) aplicaram-se os modelos SARIMA e MARIMA, com intervenções nas dez séries temporais de chegadas mensais para Hong Kong de dez países de origem e, para avaliar os desempenhos das previsões, utilizaram-se os índices de desempenho MAD, RMSE, RMSPE, MSE, MAPE e Theil's U. Destacou-se o modelo preditivo MARINA, no entanto depende do mercado de origem, pois em alguns mercados o desempenho individual do modelo SARIMA supera as previsões de MARINA.

Numa revisão de cento e vinte e uma publicações sobre previsão da demanda turística e modelagem de 2000 até 2007, na escola de administração de hotelaria e turismo da universidade politécnica de Hong Kong concluiu-se que não há modelo de previsão que em

diferentes situações tenha um desempenho de previsão melhor com relação aos demais modelos na literatura especializada de previsão da demanda (SONG & LI, 2008).

Na procura do turismo de Paris compararam-se modelos univariados (ARIMA, ES e Naive 1) e multivariados (VAR clássico e bayesiano, TVP, EC-ADLM) de previsão com a finalidade de avaliar a acuracidade dos sete modelos mediante as medidas de erro RMSE e MAE. Os resultados evidenciaram que todos os modelos testados tiveram melhor desempenho preditivo do que o modelo Naive 1, em vários horizontes de predição e mercados de origem (GUNTER & ÖNDER, 2015).

Na previsão de curto, médio e longo dos dados de chegadas turísticas em países da Europa, através da aplicação dos modelos ARIMA, ETS, NN, TBATS, ARFIMA, MA, WMA, SSA-R e SSA-V. E comparação da acurácia destes modelos pela medida de discrepância RMSE, DC e DM. Constatou-se que em diferentes horizontes de previsão e mercados pesquisados, não identificou um único modelo que tivesse melhor desempenho do que todos os diversos modelos testados em todas as circunstâncias (Hassani, Silva, Antonakakis, Filis, & Gupta, 2017).

Diante dessas constatações, o objetivo desta pesquisa é realizar previsões de curto prazo especificamente num único horizonte de três meses à frente e comparar as previsões, por meio, dos critérios MAD, MAPE e MSD. Aplicado os modelos univariados (ARIMA, Holt-Winters versões aditiva e multiplicativa) e rede neural com intervenção. No contexto turístico nos dados de MS em uma única série temporal de chegadas turísticas internacionais.

## **2. BASE TEÓRICA**

A seguir apresentam-se, com base na literatura consultada, demanda turística, Holt-Winters (HW) versões aditiva e multiplicativa, ARIMA, rede neural e medida de precisão.

### **2.1 DEMANDA TURÍSTICA**

Nas últimas décadas, houve um crescimento significativo da procura do turismo no mundo (GOH & LAW, 2002; SONG & LI, 2008; Hassani, Silva, Antonakakis, Filis, & Gupta, 2017). Além disso, os serviços turísticos têm uma natureza perecível o que estabelece a necessidade de previsão precisa da demanda do turismo (GOH, LAW, & MOK, 2008; GUNTER & ÖNDER, 2015).

A predição da demanda turística de uma cidade é fundamental para a estratégia de planejamento do ponto de vista do turismo, que proporciona expansão dos serviços turísticos tanto em investimento de recursos humanos (treinamento) como estrutura física para atender as necessidades das pessoas (Casanova, Araujo, Silva, & Rocha, 2012). Neste sentido, Claveria e Torra (2014, p. 220), afirmam que “as previsões precisas do volume do turismo desempenham um papel importante no planejamento do turismo, pois permitem que os destinos prevejam necessidades de desenvolvimento de infraestrutura”.

Alguns motivos para se prever uma futura demanda turística são apontados por Gunter e Önder (2015) como: “O quarto do hotel que não é vendido, amanhã será uma perda de receita. Um destino também precisa saber sobre quantos turistas são esperados para que a quantidade de voos para o destino pode ser ajustado, novos hotéis podem ser construídos e funcionários adicionais podem ser contratados” (p. 123).

### **2.2 HOLT-WINTERS (HW)**

O modelo foi criado em 1960, por Winters este modelo de previsão é uma extensão do modelo de Holt, assim apresenta um parâmetro a mais e equação complementar para a sazonalidade

(MAKRIDAKIS, 1976). Dentre suas principais vantagens são a sua simples utilização (CHATFIELD, 1978; Chatfield & Yar, 1988; Silver, 2000); o seu bom desempenho, de modo geral e na prática (Chatfield & Yar, 1988); o baixo custo de operação (NEWBOLD & GRANGER, 1974; SERRA, TAVARES, & SANTOS, 2005) e útil para prever como o modelo complexo de Box-Jenkins (Silver, 2000).

O modelo de série temporal HW, tem muita utilidade para os decisores e planejadores estratégicos que visam um modelo econômico e relativamente preciso com a finalidade de amplo horizonte de previsão e significativo número de séries de tempos (Brandon, Jarrett, & Khumawala, 1987).

De acordo com Chatfield (1978) “Existem dois tipos de modelo sazonal: uma versão aditiva que pressupõe que os efeitos sazonais são de tamanho constante e uma versão multiplicativa que pressupõe que os efeitos sazonais são proporcionais em tamanho ao nível médio dessazonalizado local” (p. 264). O modelo Holt-Winters baseia-se em quatro equações a primeira, o componente nível (1), em seguida, componente tendência (2), então, o componente sazonal (3) e, por fim, as previsões do tempo (4). As versões são exibidas na sequência

$$L_t = \alpha \left( \frac{x_t}{S_{t-m}} \right) + (1-\alpha)(L_{t-1} + T_{t-1}) \quad (1)$$

$$T_t = \beta (L_t - L_{t-1}) + (1-\beta)T_{t-1} \quad (2)$$

$$S_t = \gamma \left( \frac{x_t}{L_t} \right) + (1-\gamma)S_{t-m} \quad (3)$$

$$F_{t+n} = (L_t + nT_t)S_{t-m+n} \quad (4)$$

Fonte: Adaptado de Chatfield (1978)

Enquanto na versão aditiva são as equações:

$$L_t = \alpha(x_t - S_{t-m}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (5)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (6)$$

$$S_t = \gamma(x_t - L_t) + (1 - \gamma)S_{t-m} \quad (7)$$

$$F_{t+n} = L_t + nT_t + S_{t-m+n} \quad (8)$$

Fonte: Adaptado de Chatfield (1978)

Onde:  $L_t$  – estimativa do grau médio dessazonalizada no período  $t$ ;  $T_t$  – motivo sazonal de estima para o tempo  $t$ ;  $S_t$  – prazo de tendência de estima para o tempo  $t$ ;  $m$  – numero de considerações em ciclo sazonal;  $F_{t+n}$  – previsões de período  $n$  em diante;  $n = 1, 2, \dots$ ;  $x_t$  –

numero observado e  $\alpha, \beta$  e  $\gamma$  - constantes de ponderação de nível, tendência e sazonalidade, nesta ordem (CHATFIELD, 1978).

### 2.3 MODELOS AUTO-REGRESSIVOS INTEGRADOS DE MÉDIAS MÓVEIS (ARIMA)

As séries cronológicas são formadas de valores observados em vários períodos ordenados e subsequentes (STEVENSON, 1981; KAZMIER, 1982). Dentre os modelos utilizados para analisar uma série de tempo univariada tem-se o ARIMA. Proposto por Box-Jenkins em 1976, este modelo usa uma combinação de três componentes que são autorregressivo (AR), filtro de integração (I) e médias móveis (MA) (BÜNDCHEN & WERNER, 2016).

O modelo ARIMA utilizado nesta pesquisa é uma versão automática chamada de auto.arima. Onde o valor de  $d$  é escolhido com base em repetidos testes de raiz de unidade KPSS, em seguida são selecionados os valores de  $p$  e  $q$  minimizando o Critério de Informação Akaike (AIC) para determinar os valores.

$$\phi(B)(1 - B^d)y_t = c + \theta(B)\varepsilon_t \quad (9)$$

Onde,  $\varepsilon_t$  é um processo de ruído branco com zero médio e variância  $\sigma^2$ ,  $B$  é o operador de mudança de turno, e  $\phi(Z)$  e  $\theta(Z)$  são polinômios da ordem  $p$  e  $q$ , respectivamente. Para garantir causalidade e invertibilidade, presume-se que  $\phi(Z)$  e  $\theta(Z)$  não têm raízes para  $|Z| < 1$  (BROCKWELL & DAVIS, 1991). Se  $c \neq 0$ , existe um polinômio implícito de ordem  $d$  na função de previsão (Hyndaman & Khandakar, 2008). O modelo de previsão automático ARIMA seleciona a ordem de  $p$ ,  $q$  e  $d$ , por meio do critério de Informação conforme o AIC:

$$AIC = -2 \log(L) + 2(p + q + P + Q + k) \quad (10)$$

Em que  $k = 1$  se  $c \neq 0$  e 0 de outra forma, e  $L$  é a probabilidade máxima do modelo fixado para dados diferidos  $(1 - B^m)^D (1 - B)^d y_t$ . A probabilidade de o modelo completo para  $y_t$  não é realmente definida e, portanto, o valor da AIC para diferentes níveis de difusão não é comparável (Hyndaman & Khandakar, 2008).

Assim, o algoritmo Hyndman-Khandakar seleciona o modelo que apresenta o menor valor AIC de: ARIMA (2,  $d$ , 2), ARIMA (0,  $d$ , 0), ARIMA (1,  $d$ , 0) e ARIMA (0,  $d$ , 1) dentre os quatros modelos (Hassani, Silva, Antonakakis, Filis, & Gupta, 2017). Para os leitores mais interessados nos detalhes de funcionamento da modelagem ARIMA basta ver a seção 8.7 em (HYNDMAN & ATHANASOPOULOS, 2013b).

### 2.4 REDE NEURAL

A rede neural de Feed-forward constitui-se em camadas: uma de entrada, uma de saída e, de modo geral uma ou mais camadas escondidas e a capacidade de aprendizagem em reconhecer padrões desta rede neural possibilita ser utilizada como um modelo de previsão aplicado nos negócios (LAW, 2000).

A procura do turismo é uma área especialmente importante na análise de dados para predição e acurácia do desempenho de previsão de redes neurais em comparação aos modelos de séries temporais (CLAVERIA & TORRA, 2014). O modelo de rede neural usado nesta pesquisa pertence ao pacote do software R, denominado forecast e o modelo de previsão presente no manual do pacote chama-se nnetar este modelo funciona de forma semelhante aos neurônios do cérebro, assim a arquitetura da rede neural é composta por camadas.

As entradas constituem a camada inferior e as saídas formam a camada superior é possível ter camadas ocultas (Hyndman & Athanasopoulos, 2013a). Atendendo as orientações de

Hyndman e Athanasopoulos (2013a, p. 9.3), que afirmam “O número de camadas ocultas e o número de nós em cada camada oculta devem ser previamente especificados.”

Assim definimos  $p = 2$ ,  $P = 2$ ,  $size = 5$   $lambda = 3$ ,  $decay = 0.28$ . Onde:  $p$  - Dimensão de incorporação para séries temporais não sazonais;  $P$  - Número de atrasos sazonais utilizados como insumos;  $size$  - Número de nós na camada oculta;  $lambda$  - Parâmetro de transformação Box-Cox;  $decay$  - parâmetro de decaimento (Hyndman, et al., 2018).

Diante da flexibilidade das redes neurais elas não têm um processo sistêmico para criação de modelos, logo a obtenção de um modelo de rede neural robusto ocorre mediante a escolha de um grande número de parâmetros que são obtidos experimentalmente, por meio, de tentativa e erro (PALMER, MONTANO, & SESÉ, 2006).

Este método foi utilizado neste estudo para selecionar a rede neural especificada aqui. Os detalhes sobre o funcionamento do modelo nnetar pode ser visto na seção 9.3, em (Hyndman & Athanasopoulos, 2013a). A figura 1 apresenta a arquitetura geral de uma rede do tipo Feed-forward.

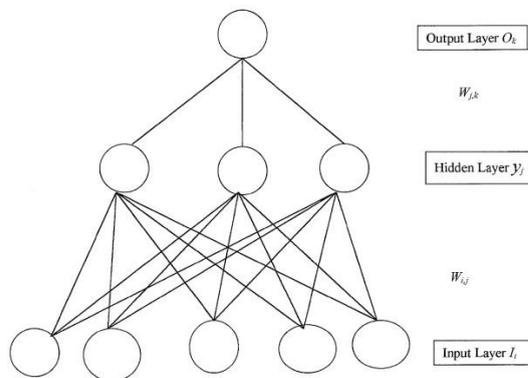


Figura 1 - Simplificação rede neural com três camadas

Fonte: Elaborado por Law (2000)

## 2.5 MEDIDA DE PRECISÃO

A partir da explicação que devesse julgar a adaptação de qualquer medida de erro utilizada pelo resultado gerado da precisão das previsões fora da amostra do modelo de previsão, assim uma medida de erro para ser selecionada precisa aliar-se a finalidade da predição e os objetivos dos decisores (Makridakis, 1993). Neste contexto numa pesquisa bibliográfica sobre previsão de demanda turística de 2005 até 2015, com um total de 129 artigos, (Bündchen, 2016), constatou que *Mean Absolute Percentage Error* (MAPE) é o critério mais usado na predição da procura do turismo.

O MAPE fornece os resultados em porcentagem que facilita a interpretação pelos decisores, bem como gera um indicativo qualitativo da precisão do modelo de previsão testado (FLORES, 1986; Makridakis, 1993). O *Mean Squared Error* (MAD) mede a acuracidade global e gera um indicativo da disseminação global em que todos os erros são dados pesos iguais (LAW, 2000). A medida é mesurada nas mesmas unidades dos dados históricos, além disso, é fácil para demonstrar aos tomadores de decisão por conta da facilidade de cálculo (FLORES, 1986). Neste sentido MAD constitui-se uma das primeiras medidas de erro, sendo singela de calcular e simples de compreender (BONETT & SEIER, 2003).

O *Mean Squared Error* (MSE) é uma medida comumente usada para ver a adaptação de um método de previsão em relação aos dados e medir o desempenho das previsões geradas, o critério também é fácil de calcular e de modo geral um bom critério para medir o desempenho de previsões da única série de tempo (Chatfield C. , 1988).

Este critério é comumente utilizado para comparar métodos de previsão, além de verificar o erro quadrático ele se diferencia de outras medidas de precisão porque demonstra qual modelo minimiza grandes erros uma vez que mostra estes erros mais do que outros critérios (THOMPSON, 1990).

$$MAD = \frac{\sum |R-D|}{n} \quad (11)$$

$$MSD = \frac{\sum (R-D)}{n-1} \quad (12)$$

$$MAPE = \frac{\sum |R-D/R|}{n} \cdot 100, (R \neq 0) \quad (13)$$

Fonte: Adaptado de (MIN, 2008)

Sendo, R = valor real de demanda; D = previsão; n = número de observações.

Medidas de erro são métodos de gerar informações sobre a precisão de certo modelo de previsão para prever os dados reais, tanto para modelos que são ajustadas aos dados quanto para previsões de períodos, fora da amostra (Makridakis, 1993).

### 3. PROCEDIMENTOS METODOLÓGICOS

Nesta seção descreve-se a área de estudo, coleta de dados e o pacote usado no ambiente de trabalho R studio.

#### 3.1 ÁREA DE ESTUDO

O Estado do Mato Grosso do Sul (Figura 2), localizado no centro-oeste brasileiro, possui cerca de 357.145,531 Km<sup>2</sup> e conta com uma população estimada em mais de 2,5 milhões de habitantes (IBGE, 2016). Situado em região transfronteiriça com Paraguai e Bolívia a Oeste, faz divisa com os Estados brasileiros Paraná, São Paulo, Minas Gerais e Goiás a Leste e o Estado de Mato Grosso ao Norte.

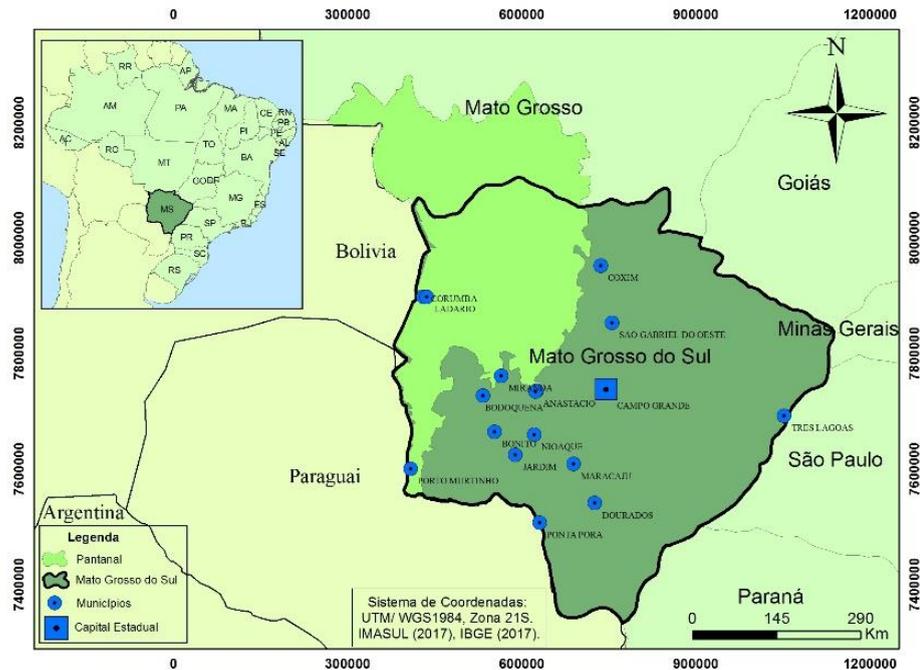


Figura 2 - Localização do Estado de Mato Grosso do Sul, Brasil

Fonte: Elaborado pelos autores através do software ARCGIS (2012)

O Estado encontra-se em parte inserido em uma importante unidade geológica como a Bacia Sedimentar do Pantanal que lhe confere potencialidades turísticas manifestadas nos municípios de Corumbá, Miranda e Aquidauana correspondendo ao Pantanal e nos municípios de Bodoquena, Bonito e Jardim pela relevância de seu ambiente cárstico (MATO GROSSO DO SUL, 2017). Localizado na zona de transição da borda do Pantanal sul-matogrossense. Essa unidade representa também um importante cenário hidrográfico, além da riqueza em sua biodiversidade e da beleza paisagística.

O Pantanal sendo a maior planície alagável do mundo e patrimônio da humanidade segundo a UNESCO (MATO GROSSO DO SUL, 2017), abrange entre suas atividades turísticas o ecoturismo, buscando realizar suas atividades de maneira sustentável preservando e conservando as belezas naturais. O turismo cultural relacionadas a festas e eventos culturais do estado, o turismo de pesca esportiva em virtude da riqueza fluvial do rio Paraguai e seus afluentes. Suas atividades contam com pousadas pantaneiras que oferecem pacotes para contemplação da fauna e flora, passeios de chalanas pelos rios da bacia hidrográfica do rio Paraguai, observação de aves típicas da região, Safaris e Voos panorâmicos (ROTTA, LUNA, & WEIS, 2006).

Adjacente ao Pantanal, devido ao seu relevo cárstico e rochas carbonáticas, a região de Bodoquena, Bonito e Jardim possuem peculiaridades turísticas através de seus “rios de águas cristalinas que abrigam exuberante fauna ictiológica; cachoeiras, fauna terrestre, flora; grutas e ainda pelo contraste da planície com a Serra da Bodoquena” Barbosa e Zamboni (2000, p. 06). Apresentando “inúmeras cavernas, dolinas, ressurgências e sumidouros, dentre outras feições, além de promover uma extrema limpidez dos cursos d’água que nela têm suas nascentes” Mato Grosso do Sul (2010, p. 15).

### 3.2 COLETA DE DADOS

Os dados foram coletados a partir da ferramenta base de dados extrator que de acordo com o Embratur (2017a) é: “O extrator de dados sobre a chegada de turistas internacionais não residentes ao Brasil é uma ferramenta eletrônica interativa que representa importante incremento na disseminação de informações sobre o setor turismo relativas Brasil, unidades da federação, regiões”.

Além disso, serve para “A série disponibilizada permite o acompanhamento da evolução do fluxo receptivo internacional de turistas, e dá ao usuário a flexibilidade, rapidez, precisão e consistência de informações que a análise gerencial de informações requer” (EMBRATUR, 2017b). Um detalhe importante sobre a coleta e a ferramenta é que “O extrator de dados é uma customização do Saiku, software de código aberto para análise de dados desenvolvida pela Meteorite BI” Embratur (2017c).

Tanto a ferramenta como as informações sobre a mesma são disponibilizadas pelo Instituto Brasileiro de Turismo (Embratur) na web site da empresa, portanto mediante a ferramenta a coleta foi referentes às chegadas turísticas por todas as vias (totais) dos períodos de janeiro de 2007 a setembro de 2016, atingindo 117 períodos mensais de amostra. Os dados utilizados encontram-se no quadro 1:

MÊS/ANO	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Janeiro	6.888	6.715	19.747	21.260	5.477	7.314	6.031	8.454	7.542	12.544
Fevereiro	5.412	6.073	13.896	15.066	4.469	3.205	4.165	5.129	4.994	6.687
Março	4.583	6.252	13.527	14.540	2.890	4.314	4.129	3.909	4.400	6.011
Abril	3.406	2.922	3.821	4.167	4.045	2.993	3.144	4.302	4.533	5.003
Mai	3.716	3.243	1.162	2.952	3.072	1.897	2.290	4.085	3.697	4.262
Junho	3.080	2.675	955	3.244	2.142	4.196	2.269	6.905	3.250	4.278
Julho	5.957	3.462	2.035	2.221	2.528	4.072	3.694	4.512	4.941	6.212
Agosto	5.946	3.702	1.401	1.538	1.402	2.919	2.817	3.867	3.508	5.398
Setembro	3.542	3.070	964	1.058	1.467	2.545	1.230	3.488	1.636	4.973
Outubro	3.660	2.746	361	652	2.866	2.309	2.745	3.334	4.192	-
Novembro	3.809	3.431	269	876	3.241	1.732	2.945	3.732	4.225	-
Dezembro	5.210	5.217	257	566	5.501	6.395	6.064	10.282	9.683	-

Quadro 1 - Chegadas turísticas internacionais por todas as vias para Mato Grosso do Sul de janeiro de 2007 a setembro de 2016

Fonte: Adaptada pelos autores a partir de (EMBRATUR, 2017)

Após a coleta da amostra para avaliar a normalidade desta amostra foi aplicada a estatística Shapiro-Wilk Test (SWT) em R usando a função *shapiro.test* para detectar a distribuição normal na amostra. O Shapiro-Wilk Test (SWT) é um poderoso e competitivo teste para a normalidade (Lee, Qian, & Shao, 2014). Realizado o *shapiro.test* e analisado os resultados apresentados na tabela 1. Verifica-se que a amostra possui uma distribuição normal.

SHAPIRO-WILK	W	P-VALUE
	0.76	1.512e-12

Tabela 1 - Resultado Shapiro-Wilk Test (SWT)

Fonte: Elaborada pelos autores

### 3.3 PACOTE EM R STUDIO

Para fazer as previsões e comparar a capacidade preditiva dos modelos ARIMA, HW versões aditiva, multiplicativa e Rede Neural, foi adotado com o apoio do software estatístico R. Procedeu-se as análises de previsão e acuracidade das previsões no *R studio* que é uma extensão do ambiente R, mediante o pacote *forecast* do R (Hyndman R. , et al., 2018). Os cálculos das medidas de erro MAD, MAPE e MSD também foram executados com o pacote *forecast* no *R studio*.

As funções em R utilizadas para prever os valores no curto prazo de três meses são *auto.arima*, *forecast.HoltWinters* e *nnetar*, já para comparar a acuracidade dos modelos de previsão foi usada a função *accuracy*. Assim, contemplando o período entre janeiro de 2007 até setembro de 2016, verificou-se o comportamento das previsões com os dados reais (já disponibilizados pela própria Embratur) concernentes ao período de outubro a dezembro de 2016.

Esses recursos computacionais são de livre acesso e atualmente disponíveis para planejadores turísticos e acadêmicos. No momento em que esses dados foram coletados não estavam disponíveis pela Embratur as observações mais atuais de 2018, por essa razão determinamos o último trimestre de 2016, como o período de previsão.

## 4. RESULTADOS DA PESQUISA

A princípio procedeu-se a criação de um gráfico de série temporal mediante os dados coletados de chegadas turísticas e atendem-se as orientações de (HYNDMAN & ATHANASOPOULOS, 2013), que recomendam inicialmente mostrar os dados em gráfico, construiu-se o gráfico de série temporal representado na figura 3.

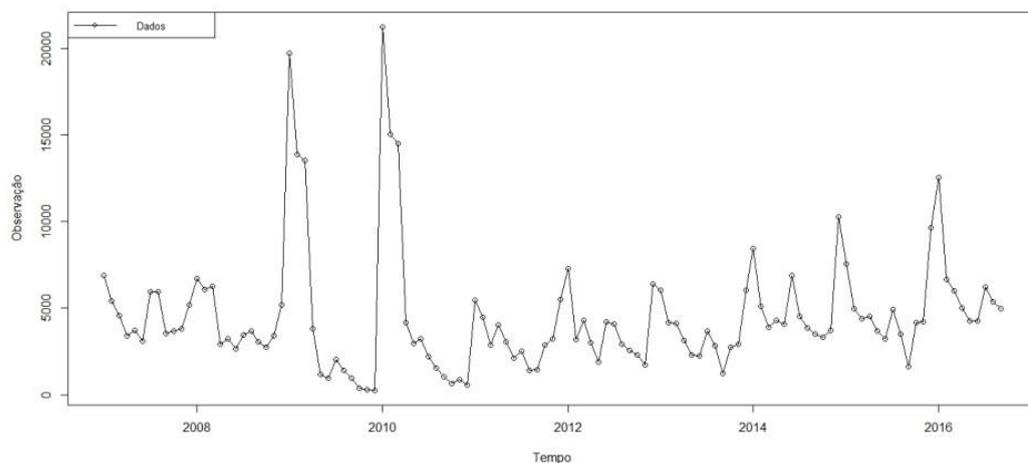


Figura 3 - Chegadas turísticas internacionais por todas as vias

Fonte: Elaborado pelos autores através do software R e com base em dados da EMBRATUR (2017)

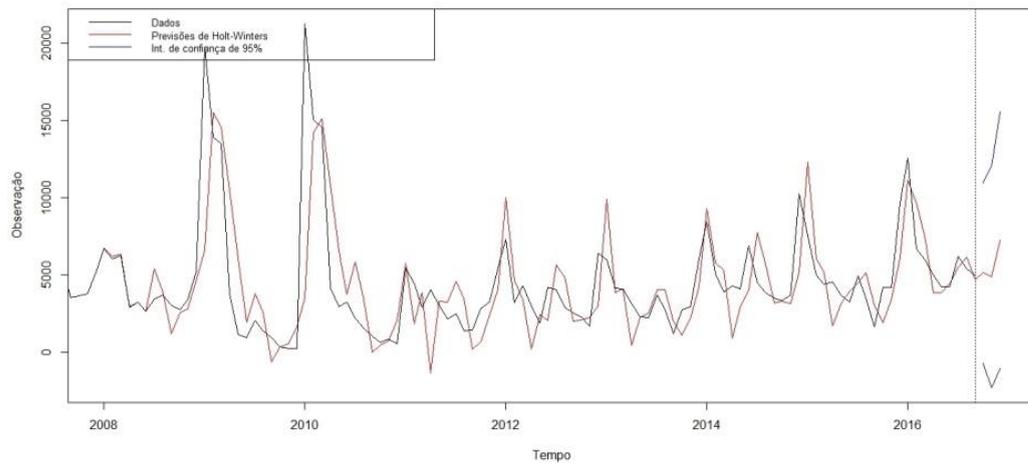


Figura 4 - Previsões de Holt-Winters versão aditiva de outubro até dezembro de 2016, com intervalo de confiança de 95%

Fonte: Elaborado pelos autores com software R (2017)

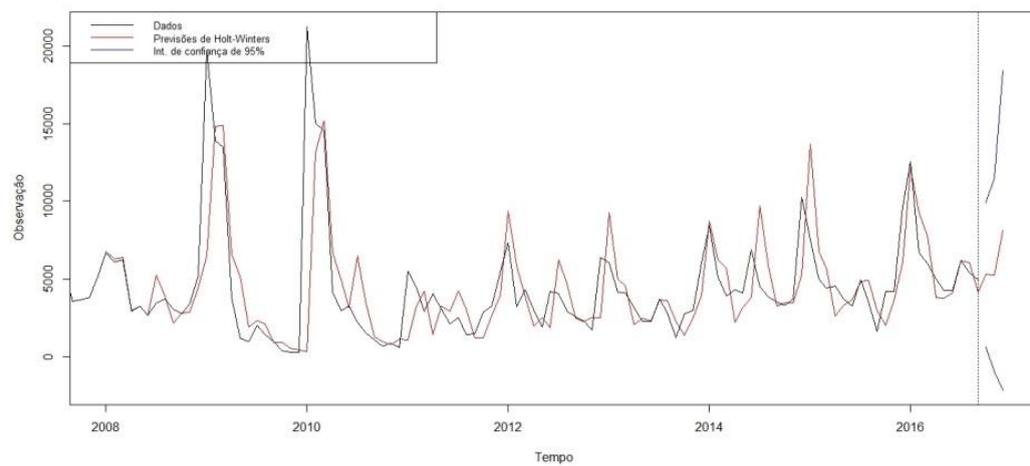


Figura 5 - Previsões de Holt-Winters versão multiplicativa de outubro até dezembro de 2016, com intervalo de confiança de 95%

Fonte: Elaborado pelos autores com software R (2017)

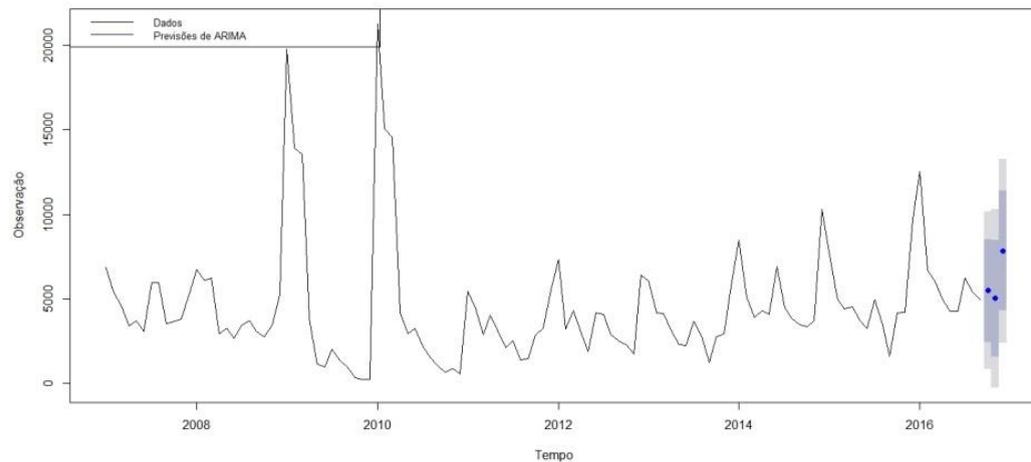


Figura 6 - Previsões de ARIMA de outubro até dezembro de 2016

Fonte: Elaborado pelos autores com software R (2017)

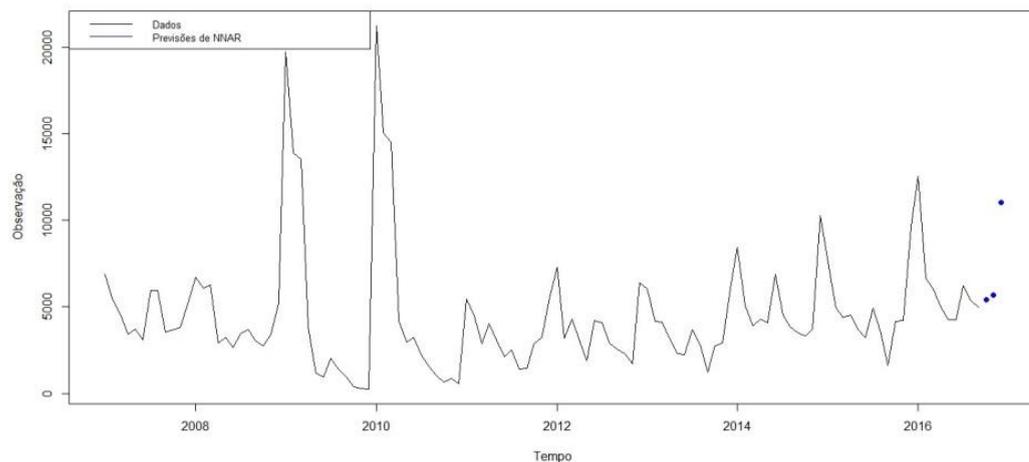


Figura 7 - Previsões de NNAR de outubro até dezembro de 2016

Fonte: Elaborado pelos autores com software R (2017)

Os parâmetros de alisamento para as equações de nível, tendência e sazonalidade de ambas as versões de Holt-Winters, o modelo ARIMA selecionado com o menor critério de informação akaike (AIC) e rede neural (NNAR) onde  $p$  é o número automaticamente selecionado para série de tempo não sazonal.  $P$  é o número de inputs atrasadas e  $size$  é o número de nós na camada escondida estes referentes às previsões são demonstradas na tabela 2.

HW ADITIVA ( $\alpha, \beta, \gamma$ )	HW MULTIPLICATIVA ( $\alpha, \beta, \gamma$ )	ARIMA	NNAR ( $p; P; size$ )
(0.7152064,0,0.3542533)	(0.703771,0,0.2775127)	(1,0,0) (1,0,0)	(2,2,5)

Tabela 2 - Parâmetros das previsões dos modelos

Fonte: Elaborada pelos autores

Os modelos descritos foram utilizados para gerar as previsões de 3 meses das chegadas turísticas. As observações e previsões de Holt-Winters versões aditiva, multiplicativa,

ARIMA e rede neural em números para os meses de outubro até dezembro de 2016, podem ser visualizadas na tabela 3.

ANO-2016					
MESES	DADOS REAIS	HW ADITIVA	HW MULTIPLICATIVA	ARIMA	REDE NEURAL (NNAR)
Outubro	4.946	5.143,43	5.286,24	5.495,44	5.406,95
Novembro	5.716	4.887,94	5.265,71	5.021,20	5.676,80
Dezembro	10.998	7.267,71	8.135,01	7.843,58	11.053,84

Tabela 3 - Comparação dos valores previstos com dados reais

Fonte: Elaborada pelos autores com base em EMBRATUR (2017)

As medidas de erro MAPE, MAD e MSD apresentam a precisão dos modelos em comparação com os dados reais. Uma característica óbvia na tabela 4, é que o modelo NNAR superou as duas versões de HW e ARIMA, assim apresentou um MAPE inferior a 10% que é considerado pela classificação de (LEWIS, 1982), como previsão excelente.

MODELO	MAPE	MAD	MSD
HW aditiva	17.46538%	1453,63890	4879908,48
HW multiplicativa	13.59625%	991,01214	2838412,03
ARIMA	17.31526%	1099,92570	3578332,30
Rede neural (NNAR)	3.50438%	-159.1936	72376,55

Tabela 4 - Medidas de erro

Fonte: Elaborada pelos autores

O segundo melhor desempenho individual fica com HW versão multiplicativa isto com base nos resultados dos critérios MAPE, MAD e MSD como pode ser observada na tabela 4. De acordo com as diretrizes de (LEWIS, 1982), os resultados de MAPE entre 10% e 20% são interpretados como previsão precisa, diante disso, HW ambas as versões e ARIMA apresentaram boas previsões.

## 5. CONSIDERAÇÕES FINAIS

Quanto aos resultados obtidos na pesquisa, mostrou-se que a rede neural (NNAR) fornece previsões, mais precisas da demanda turística internacional de Mato Grosso do Sul, do que ARIMA e Holt-Winters (versões aditiva e multiplicativa). Os resultados obtidos estão consistentes com o que foi constatado por (CHO, 2003).

Além disso, o MAPE de 3,50% deste estudo alcançou os resultados de MAPE (3,44%) obtidos pela previsão complexa da rede MPL do fenômeno turístico na pesquisa de (PALMER, MONTANO, & SESÉ, 2006). Isto indica que é viável a utilização do modelo NNAR para prever a demanda turística de Mato Grosso do Sul. A pesquisa limitou-se em um único horizonte de previsão (3) meses fora da amostra e uma única série temporal.

Para superar essa limitação sugere-se a estudos futuros executar a previsão de vários horizontes de previsão (1, 2, 4, 6, 12 e 24 meses à frente) e para essas predições utilizar um amplo número de séries temporais (no mínimo dez) para constatar o nível de generalização

dos resultados verificados, isso também é recomendável por conta dos resultados de (GOH & LAW, 2002; SONG & LI, 2008; Hassani, Silva, Antonakakis, Filis, & Gupta, 2017).

## REFERÊNCIAS

- ARCGIS, ESRI. (2012). *10.1. Redlands*(Software). California: ESRI.
- Barbosa, M. A., & Zamboni, R. A. (2000). *Formação de um 'cluster' em torno do turismo de natureza sustentável em Bonito-MS*. Acesso em 14 de Agosto de 2017, disponível em <https://goo.gl/9JFCiM>.
- BONETT, D. G., & SEIER, E. (2003). Statistical inference for a ratio of dispersions using paired samples. *Journal of Educational and Behavioral Statistics*, 28(1), 21-30.
- Brandon, C., Jarrett, J. E., & Khumawala, S. B. (1987). A comparative study of the forecasting accuracy of Holt-Winters and economic indicator models of earnings per share for financial decision making. *Managerial Finance*, 13(2), 10-15.
- BROCKWELL, P. J., & DAVIS, R. A. (1991). Time series: theory and methods. *Springer-Verlag*(2).
- Bündchen, C. (2016). Previsão de demanda turística e a acurácia das previsões frente à realização de megaeventos. *Escola de Engenharia, Programa de Pós-Graduação em Engenharia de Produção*. Porto Alegre, Brasil: Universidade Federal do Rio Grande do Sul.
- BÜNDCHEN, C., & WERNER, L. (2016). Comparação da Acurácia de Previsões de Demanda Turística em Sedes Olímpicas. *Revista Turismo em Análise*, 27(1), 85-107.
- Casanova, S., Araujo, V. M., Silva, W. V., & Rocha, D. T. (2012). Previsão da demanda turística da cidade de foz do iguaçu: uma aplicação com os modelos arima. *Turismo: Visão e Ação*, 14(3).
- CHATFIELD, C. (1978). The holt-winters forecasting procedure. *Applied Statistics*, 264-279.
- Chatfield, C. (1988). Apples, Oranges and Mean Square Error. *International Journal of Forecasting*, 4(4), 515-518.
- Chatfield, C., & Yar, M. (1988). Holt-Winters forecasting: some practical issues. *The Statistician*, 129-140.
- CHO, V. (2003). A comparison of three different approaches to tourist arrival forecasting. *Tourism management*, 24( 3), 323-330.
- CHU, F.-L. (1998). Forecasting tourism demand in Asian-Pacific countries. *Annals of Tourism Research*, 25(3), 597-615.
- CLAVERIA, O., & TORRA, S. (2014). Forecasting tourism demand to Catalonia: Neural networks vs. time series models. *Economic Modelling*, 36, 220-228.
- EMBRATUR. (2017). *Acesse o EXTRATOR DE DADOS*. Acesso em 29 de Junho de 2017, disponível em <https://goo.gl/TNukIX>.
- EMBRATUR. (2017a). *O QUE É*. Acesso em 29 de Junho de 2017, disponível em <https://goo.gl/SqyAao>.
- EMBRATUR. (2017b). *PARA QUE*. Acesso em 29 de Junho de 2017, disponível em <https://goo.gl/4WrFcu>.
- EMBRATUR. (2017c). *TERMOS E CONDIÇÕES DE USO*. Acesso em 29 de Junho de 2017, disponível em <https://goo.gl/k6Wgx5>.
- EMBRATUR. (2017d). *Turismo já ajudou países a superar crises e pode fazer o mesmo no Brasil*. Acesso em 22 de Junho de 2017, disponível em <https://goo.gl/QwCPQy>.
- FLORES, B. E. (1986). A pragmatic view of accuracy measurement in forecasting. *Omega*, 14(2), 93-98.

- GOH, C., & LAW, R. (2002). Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism management*, 23(5), 499-510.
- GOH, C., LAW, R., & MOK, H. (2008). Analyzing and forecasting tourism demand: A rough sets approach. *Journal of Travel Research*, 46(3), 327-338.
- GUNTER, U., & ÖNDER, I. (2015). Forecasting international city tourism demand for Paris: Accuracy of uni- and multivariate models employing monthly data. *Tourism Management*, 46, 123-135.
- Hassani, H., Silva, E. S., Antonakakis, N., Filis, G., & Gupta, R. (2017). Forecasting accuracy evaluation of tourist arrivals. *Annals of Tourism Research*, 63, 112-127.
- Hyndaman, & Khandakar. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 27(3), 2-22.
- HYNDMAN, R. J., & ATHANASOPOULOS, G. (2013). *Forecasting: principles and practice*. (OTexts, Ed.) Acesso em 01 de Julho de 2017, disponível em <https://www.otexts.org/fpp/1/6>.
- Hyndman, R. J., & Athanasopoulos, G. (2013a). *Forecasting: principles and practice*. (OTexts, Ed.) Acesso em 30 de Junho de 2017, disponível em <https://www.otexts.org/fpp/9/3>.
- HYNDMAN, R. J., & ATHANASOPOULOS, G. (2013b). *Forecasting: principles and practice*. (OTexts, Ed.) Acesso em 20 de Julho de 2017, disponível em <https://www.otexts.org/fpp/8/7>.
- Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., . . . Zhou, Z. (2018). *forecast: Forecasting Functions for Time Series and Linear Models*. (CRAN, Ed.) Acesso em 01 de Julho de 2017, disponível em <https://cran.r-project.org/web/packages/forecast/forecast.pdf>.
- IBGE. (2016). *Instituto Brasileiro de Geografia e Estatística, Estados*. Acesso em 23 de Novembro de 2017, disponível em <https://goo.gl/arjBQb>.
- KAZMIER, L. J. (1982). *Estatística aplicada à economia e administração* (1ª ed.). São Paulo: McGRAW-Hill do Brasil.
- Kim, J. H., Wong, K., & George Athanasopoulos, S. L. (2011). Beyond point forecasting: evaluation of alternative prediction intervals for tourist arrivals. *International Journal of Forecasting*, 27(3), 887-901.
- LAW, R. (2000). Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. *Tourism Management*, 21(4), 331-340.
- Lee, R., Qian, M., & Shao, Y. (2014). On rotational robustness of Shapiro-Wilk type tests for multivariate normality. *Open Journal of Statistics*, 4, 964-969.
- LEWIS, C. D. (1982). Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting. *Butterworth-Heinemann*.
- MAKRIDAKIS, S. (1976). A survey of time series. *International Statistical Review/Revue Internationale de Statistique*, 29-70.
- Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting*, 9(4), 527-529.
- Mato Grosso do Sul. (2010). *BODOQUENA-PANTANAL, GEOPARK. Dossiê de candidatura à Rede Global de Geoparks Nacionais sob auspício da Organização das Nações Unidas para Educação, Ciências e Cultura*. Acesso em 14 de Agosto de 2017, disponível em <https://goo.gl/7sSUrg>.
- MATO GROSSO DO SUL. (2017). *Fundação de Turismo de MS-FUNDTUR: Indicadores Turísticos de MS*. Acesso em 14 de Agosto de 2017, disponível em <https://goo.gl/h8icqB>.

- MIN, J. C. (2008). Forecasting Japanese tourism demand in Taiwan using an intervention analysis *International Journal of Culture, Tourism and Hospitality Research*, 2(3), 197-216.
- NEWBOLD, P., & GRANGER, C. W. (1974). Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society. Series A (General)*, 131-165.
- PALMER, A., MONTANO, J. J., & SESÉ, A. (2006). Designing an artificial neural network for forecasting tourism time series. *Tourism Management*, 27(5), 781-790.
- ROTTA, M. A., LUNA, H. S., & WEIS, W. A. (2006). *Ecoturismo no Pantanal. Corumbá: Embrapa Pantanal*. (Embrapa, Ed.) Acesso em 14 de Agosto de 2017, disponível em <https://goo.gl/Eh6YRe>.
- SERRA, C. M., TAVARES, H. R., & SANTOS, J. C. (29 Outubro a 01 de Novembro de 2005). Aplicação de séries temporais na análise de demanda turística no Estado do Pará usando os modelos de Holt-Winters. *XXV Encontro Nac. de Eng. de Produção*, pp. 3077-3083.
- Silver, M. (2000). *Estatística para administração*. São Paulo: Atlas.
- SONG, H., & LI, G. (2008). Tourism demand modelling and forecasting—A review of recent research. *Tourism management*, 29( 2), 203-220.
- STEVENSON, W. J. (1981). *Estatística aplicada à administração* (1ª ed.). São Paulo, Brasil: Harper & Row do Brasil.
- THOMPSON, P. A. (1990). An MSE statistic for comparing forecast accuracy across series. *International Journal of Forecasting*, 6(2), 219-227.