ISSN: 2237-8642 (Eletrônica)

ESTIMATIVAS DAS MODELAGENS POR REDES NEURAIS TIPO GMDH E BOX-JENKINS PARA AS SÉRIES DE TEMPERATURAS MÁXIMAS E MÍNIMAS DO MUNICÍPIO DE ARIQUEMES (RO)

CARVALHO, Roberto Luís da Silva – roberto.carvalho@ifro.edu.br
Instituto Federal de Rondônia / IFRO

DELGADO, Angel Ramon Sanchez – asanchez@ufrrj.br
Universidade Federal Rural do Rio de Janeiro / UFRRJ

RESUMO: A busca por metodologias que possam aprimorar o processo de predição de variáveis climáticas, para uma determinada região, contribui para o desenvolvimento em diversas áreas do conhecimento, seja na identificação de eventos extremos climáticos ou na previsão diária para processos agrícolas de irrigação. Neste sentido, este estudo tem como objetivo apresentar uma abordagem computacional para estimar as séries temporais associadas às temperaturas máximas e mínimas do ar, no município de Ariquemes (Rondônia). Especificamente, buscou-se modelar o comportamento das séries temporais utilizando redes neurais artificiais do tipo GMDH (Método de Grupo de Manipulação de Dados) e compará-los com a modelagem Box-Jenkins. Foram utilizados os dados climatológicos disponíveis no Instituto Nacional de Meteorologia - INMET, obtidos na estação Meteorológica Automática de Ariquemes/RO, no período de janeiro de 2011 até janeiro de 2014. Em geral, a modelagem por Redes Neurais do tipo GMDH para estas séries apresentou resultados similares aos modelos Box-Jenkins, constituindo-se, assim, numa opção para previsão de séries temporais climáticas. Modelos do tipo GMDH com quatro entradas e três camadas se mostraram mais eficientes, ou seja, com menor erro quadrático médio. A função de transferência de base radial foi mais eficiente.

PALAVRAS-CHAVE: modelagem, séries temporais, redes neurais

ESTIMATES OF THE MODELS BY NEURAL NETWORKS TYPE GMDH AND BOX-JENKINS FOR THE MAXIMUM AND MINIMUM TEMPERATURE SERIES OF THE ARIQUEMES (RO) MUNICIPALITY

ABSTRACT: The search for methodologies that can improve the process of prediction of climate variables for a given region, contributes to the development in several areas of knowledge, either in the identification of extreme weather events or daily forecasting for agricultural irrigation processer. In this sense, this study aims to present a computational approach to estimate the time series associated with maximum and minimum air temperatures, in the municipality of Ariquemes (Rondônia). Specifically, we attempted to model the behavior of the time series using artificial neural networks of the GMDH-type (Group Manipulation Data Handling) and to compare them with Box-Jenkins model. The climatological data available at the National Institute of Meteorology (INMET), obtained at the Ariquemes / RO Automatic Weather Station, were used from January 2011 to January resented similar ict climatic time series. GMDH models with four inputs and three layers were more efficient, that is, with the lowest mean square error. The radial base transfer function was more efficient.

provided by Biblioteca Digital de Teses e Dissertações da UFRR (Universidade Federal do Pará)
View metadata, citation and similar papers at core.ac.uk
brought to you by CORE

KEYWORDS: modeling, time series, neural networks.

ESTIMATIVAS DE LOS MODELADOS POR REDES NEURALES TIPO GMDH Y BOX-JENKINS PARA LAS SERIES DE TEMPERATURAS MÁXIMAS Y MÍNIMAS DEL MUNICIPIO DE ARIQUEMES (RO)

RESUMEN: La búsqueda de metodologías que puedan mejorar el proceso de predicción de variables climáticas para una región determinada, contribuye al desarrollo en varias áreas de conocimiento, ya sea en la identificación de eventos climáticos extremos o en el pronóstico diario de los procesos de riego agrícola. En este sentido, este estudio tiene

como objetivo presentar un abordaje computacional para estimar las series temporales asociadas a las temperaturas máximas y mínimas del aire, en el municipio de Ariquemes (Rondônia). En concreto, se buscó modelar el comportamiento de las series temporales utilizando redes neuronales artificiales del tipo GMDH (Método de Grupo de Manipulación de Datos) y compararlas con el modelado Box-Jenkins. Se utilizaron los datos climatológicos disponibles en el Instituto Nacional de Meteorología - INMET, obtenidos en la estación Meteorológica Automática de Ariquemes / RO, en el período de enero de 2011 hasta enero de 2014. En general, se el modelado de la red neural GMDH para estas series presentó resultados similares a los modelos Box-Jenkins, constituyendo así una opción para previsión de series temporales climáticas. Los modelos GMDH con cuatro entradas y tres capas fueron más eficientes, es decir, menor error cuadrático medio. La función de transferencia de base radial fue más eficiente.

PALABRAS CLAVE: modelado, series temporales, redes neuronales

1. INTRODUÇÃO

O estudo e o conhecimento do comportamento das variáveis meteorológicas são extremamente importantes para o entendimento do clima de uma determinada região. Sendo assim, a previsão de informações climáticas pode auxiliar na tomada de decisões e na maximização da produção agrícola, bem como prevenir dos riscos devido às drásticas mudanças climáticas. Para tanto, algumas técnicas matemáticas e estatísticas, como interpolações e modelos lineares ou não lineares, vêm sendo comumente utilizadas na estimação de informações meteorológicas (FERRAZ, 2014; COUTINHO, SILVA e DELGADO, 2016; DANTAS et al., 2016; TURCO e CHAVES, 2018). Com isso, há uma procura por metodologias que possam aprimorar o processo de predição para uma determinada região.

A não existência de um método padrão para predizer tais informações de forma exata, impulsiona a busca de técnicas que possam realizar predições consideravelmente próximas dos valores reais. Isto possibilita o emprego de técnicas de inteligência computacionais conhecidas como Redes Neurais Artificiais (RNA), que são "metaheurísticas" baseadas no funcionamento do cérebro humano, através de um processamento massivamente paralelo e distribuído, capaz de aprender através de exemplos e de generalizar a informação aprendida (HAKYKIN, 2001; BARRETO, 2002; HEATON, 2010; BRAGA, CARVALHO e LUDERMIR, 2016).

Ferreira et al. (2011) descrevem que na rede existe um conjunto de n conexões de entrada, na qual cada entrada x está ligada a um neurônio com um peso associado w. A seguir, soma-se os sinais de entrada com seus pesos; utiliza-se uma função de ativação para gerar uma saída, a qual é enviada ao próximo nó do sistema.

Objetiva-se na rede neural comparar o valor de entrada no sistema com o real valor de saída que é desejado ou que se tem como objetivo a estimar. Com isso, ocorre o treinamento do sistema em um modo de aprendizado supervisionado, em que os valores real e de saída são comparados. Sendo assim, avaliam-se os valores residuais e, caso não sejam satisfatórios, ocorre um novo ajuste no modelo visando diminuir os valores residuais. O teorema de Kolmogorov-Nielsen descreve que teoricamente basta somente uma camada oculta (intermediária) na rede. Pois, dada uma função contínua arbitrária, existe sempre para essa função uma implementação exata com uma rede neural de

três camadas, sendo a camada de entrada um vetor de dimensão n, a camada oculta composta por (2n+1) neurônios e a camada de saída com m neurônios representando as m componentes do vetor y, justificando assim a aplicação das redes neurais para modelar fenômenos lineares e não lineares (MURTGATH, 1994; KOVACS, 2002).

No contexto das séries temporais, Acock e Pachepsky (2000) utilizaram o método GMDH (Group Metod Data Handling), método de grupo de manipulação de dados, para preencher falhas de dados em estações meteorológicas das variáveis radiação solar diária, temperaturas máximas e mínimas e velocidade do vento em Stoneville – Misissippi, durante o período de 1982 a 1992. Neste sentido, visto que alguns pesquisadores utilizaram o método GMDH para a previsão e estimação de séries climáticas, optou-se na aplicação desta técnica para estimar as temperaturas máximas e mínimas de Ariquemes visando melhorar as práticas de estimação e predição de séries temporais.

O método GMDH foi desenvolvido pelo matemático russo Ivakhnenko (1966), baseado no método multicamadas Rosenblatt Peceptron (FARLOW 1981). Este tem como objetivo construir um sistema de modelagem ótima baseada somente nos dados, considerando a relação de entrada e saída do sistema. Segundo Ivakhnenko e Ivakhnenko (2000, pág. 188), o GMDH "foi desenvolvido para sistemas complexos de modelagem, previsão, identificação e aproximação de sistemas multifatoriais, diagnósticos, reconhecimento de padrões e agrupamentos de amostra de dados". A RNA GMDH é baseada na utilização do método heurístico de auto-organização, através de uma estrutura de múltiplas camadas "feedforward" (KONDO e UENO, 2012).

Diversas são as aplicações do GMDH, por exemplo Samsudin, Saad e Shabri (2009) combinaram o GMDH com algoritmos genéticos (GA) para melhorar a previsão de séries temporais. Os autores utilizaram os resultados de 4 métodos de previsões: regressão linear, regressão quadrática, suavização exponencial e ARIMA (Autorregressivo Integrado de Médias Móveis), como valores de entrada para a construção dos polinômios do GMDH. A série de dados utilizada foi a série anual da taxa de mortalidade de câncer da Pensilvânia. Dentre os resultados, os autores sugerem que as combinações das previsões com os métodos melhoram índices das previsões.

Shabri e Sansundin (2014) utilizaram em seu estudo a modelagem Box-Jenkins para determinar as variáveis de entrada da modelagem GMDH, sendo assim constituída de uma modelagem híbrida. Os autores consideram que as composições dos modelos melhoram os resultados das previsões.

Dag e Yozgaligil (2016), além da implementação do método GMDH e RGMDH (método revisado) para o software R (R CORE TEAM, 2016), verificaram melhores índices para esses métodos quando comparados com ARIMA e suavização exponencial para previsão de séries temporais das taxas de mortalidade por câncer na Pensilvânia de 1930 a 2000.

Neste sentido, este estudo tem como objetivo geral apresentar uma abordagem computacional para estimar as séries temporais associadas às temperaturas máximas e mínimas do ar, no município de Ariquemes (Rondônia). Especificamente, trata-se de modelar o comportamento das séries temporais utilizando uma RNA tipo GMDH e compará-los com a modelagem Box-Jenkins.

2. MATERIAL E MÉTODOS

2.1 LOCALIZAÇÃO E DADOS

Ariquemes é um município ao noroeste do estado de Rondônia, localizado a uma latitude 09º 54' 48" e a uma longitude 63º 02' 27" oeste, estando a uma altitude de 142 metros. Sua população estimada é de 107.345 habitantes (IBGE, 2017) e possui uma área de 4.427 km^2. O clima de Ariquemes inicialmente era identificado na classificação de KÖPPEN como tipo Aw – Clima tropical chuvoso (DUBREUIL et al., 2018). No entanto, a partir de 2014, Alvares et al. (2014) atualizaram essa classificação para Am (clima quente de moção), que se aplica a quase todo o estado de Rondônia, sendo este do tipo equatorial. Este é predominantemente quente e úmido, pois consiste basicamente em muito calor e umidade intercalados com um período de seca de 3 meses, um período de chuva de 7 meses e 2 meses de transição (CARVALHO et al., 2016).

Para a caracterização das séries temporais das temperaturas do ar (máximas e mínimas) foram utilizados os dados disponíveis no Instituto Nacional de Meteorologia - INMET, obtidos na estação Meteorológica Automática de Ariquemes/RO, no período de janeiro de 2011 até janeiro de 2014. A torre de estação meteorológica está situada nas coordenadas geográficas 09º 56' S e 62º 57' W, em 140 metros acima do nível do mar.

2.2 PROCEDIMENTO DE ANÁLISE DE DADOS

Para a análise de dados foi utilizado o software R versão 3.3.1 (R CORE TEAM, 2016) através do pacote redes neurais tipo GMDH (DAG e YOZGATLIGIL, 2016). A programação em R foi editada no software Tinn-R GUI/Editor for R Environment versão 5.01.02.00 (FARIA et al., 2016).

No procedimento de análise de dados, inicialmente, foi realizada uma análise exploratória de dados para identificar os valores médios observados, bem como os valores de máximo e mínimo para o período, identificando possíveis erros ou valores extremos.

A seguir, os modelos GMDH foram ajustados de acordo com a quantidade de variáveis de entrada (três e quatro *inputs*), número de camadas intermediárias (um, dois e três *layers*) e as funções de transferência (polinomial, sigmoide, RBF e tangente). Neste estudo, os dados observados nos dias anteriores foram utilizados como variáveis de entrada X_1, X_2, X_3 e X_4, ou seja, informações correspondentes a $n-1$, $n-2$, $n-3$, e $n-4$. Tomou-se como melhor modelo ajustado aquele com menor erro quadrático médio (EQM) dos valores preditos. Para o ajuste da rede neural o conjunto de dados foi separado em um conjunto de aprendizagem (70% dos dados) e um conjunto de testes (30%).

Na sequência, as equações foram identificadas pelo método redes neurais tipo GMDH e foram plotados os valores preditos, observados, previstos e intervalos de confiança para as séries.

2.3 MODELOS GMDH E BOX-JENKINS

O GMDH é uma técnica de modelagem que aprende as relações entre as variáveis. O método seleciona o melhor modelo na qual explica a relação entre a resposta e as séries temporais com atrasos, via funções de transferência, utilizando uma estrutura de múltiplas camadas "*feedforward*", aquela que se propaga para frente (KONDO e UENO, 2012).

No desenvolvimento do método GMDH (DAG e YOZGATLIGIL, 2016) assume-se uma série temporal com t unidades de tempo e p entradas. Com isso, modelos são construídos para os dados com atrasos de tempos (*lags*), no qual os números de observações são representados nas colunas da tabela por t - p e p é o número de entradas, das séries temporais atrasadas. A variável z é colocada como a variável resposta (substituindo os valores de y, que melhor estimam) e as demais variáveis são levadas ao modelo com retardos da série temporal x_i, com $i = 1, 2, \dots, p$.

As funções de transferência utilizadas no estudo seguem as sugeridas por Kondo e Ueno (2012): sigmoide, base radial, polinomial e função tangente (Tabela 1).

Tabela 1 – Funções de transferências utilizadas no pacote GMDH

Função de Transferência	Equação
Polinomial	$z = y$
Sigmoide	$z = \frac{1}{(1 + e^{-y})}$
Função de Base Radial (RBF)	$z = e^{-y^2}$
Tangente	$z = \tan(y)$

Fonte: Dag e Yozgaligil (2016)

Inicialmente, utiliza-se o algoritmo GMDH (IVAKHNENKO, 1966) para construir os polinômios Kolmogorov-Gabor (Equação 1) (IVAKHNENKO e IVAKHNENKO, 1995):

$$y = a + \sum_{i=1}^{m} b_i x_i + \sum_{i=1}^{m}\sum_{j=1}^{m} c_{ij} x_i x_j + \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} d_{ijkl} x_i x_j x_k + \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{l=1}^{m} e_{ijkl} x_i x_j x_k x_l + \cdots \quad (1)$$

onde m é o número de variáveis e $a, b, c, d, \dots$ são os coeficientes das variáveis no polinômio (chamados de pesos). No qual y é a variável resposta e x_i e x_j são as séries temporais atrasadas para serem regredidas (*lags*). Geralmente os termos são calculados até os termos quadráticos, ficando da seguinte forma (Equação 2):

$$y = a + \sum_{i=1}^{m} b_i x_i + \sum_{i=1}^{m}\sum_{j=1}^{m} c_{ij} x_i x_j \quad (2)$$

O método GMDH considera todos os pares de combinações de p das séries temporais atrasadas. Com isso, cada combinação entra em cada neurônio. Usando estas duas entradas, um modelo é construído para estimar a saída desejada. Na Equação 3 são considerados m = 2 ($\boldsymbol{x_i}$ e $\boldsymbol{x_j}$), sendo necessário seis coeficientes em cada modelo quando estimados:

$$y = G(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \quad (3)$$

com isso são construídos $\binom{m}{2}$ equações parciais, onde $\binom{m}{2} = m(m-1)/2$.

O método GMDH é um sistema de camadas na qual existem neurônios. O número de neurônios em uma camada (*layer*) é definido pelo número de entradas (*inputs*) de variáveis, dado pela combinação de todos os pares tomados dois a dois, ou seja, $h = \binom{p}{2}$.

De acordo com um critério externo, p neurônios são selecionados e h - p neurônios são eliminados da rede. Neste estudo, foi utilizado o mesmo critério externo dos autores do pacote, o erro médio quadrático da predição (EQMP). As saídas obtidas de neurônios selecionados tornam-se as entradas para a próxima camada. Este processo continua até a última camada. Na última camada, apenas um neurônio é selecionado. A saída obtida da última camada é o valor previsto para as séries temporais. Sendo assim, neste algoritmo existem 6 coeficientes para serem estimados em cada modelo e são estimados via RLSE – mínimos quadrados regularizados.

O esquema apresentado por Dag e Yozgatligil (2016) exemplifica uma rede com quatro variáveis de entrada e duas ou mais camadas intermediárias (Figura 1).

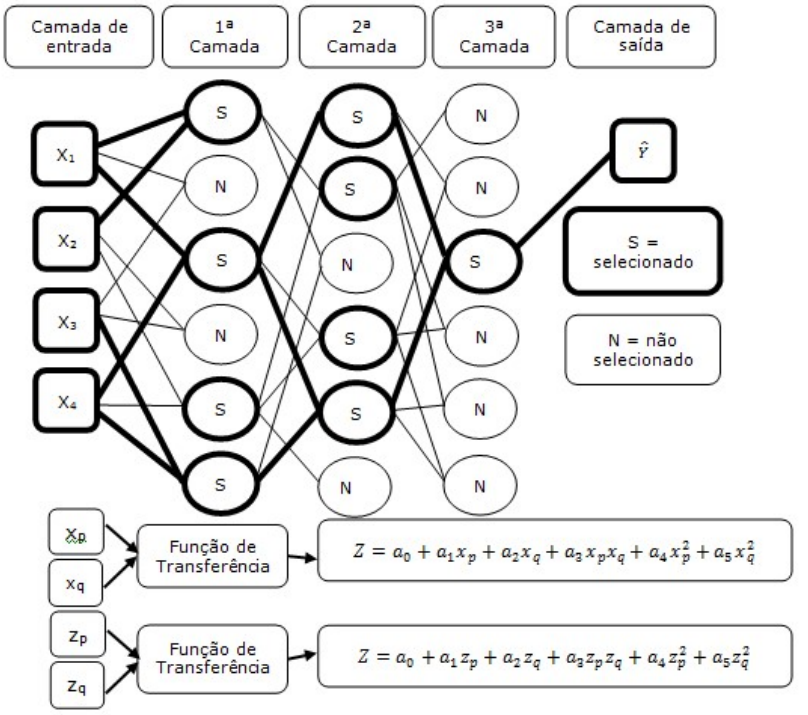

Figura 1 - Esquema dos polinômios parcias do GMDH (DAG e YOZGATLIGIL, 2016).

No pacote GMDH do software R proposto por Dag e Yozgatligil (2016) estão disponíveis os seguintes valores de parâmetros regularizadores $\lambda = \{0;\ 0{,}01;\ 0{,}02;\ 0{,}04;\ 0{,}08;\ 0{,}16;\ 0{,}32; 0{,}64;\ 1{,}28;\ 2{,}56;\ 5{,}12;\ 10{,}24\}$. Tikhonov (1963) afirmou que para $\lambda = 0$ o método de mínimos quadrados regularizados converge para o método de mínimos quadrados. Neste caso, para o ajuste dos modelos

do estudo, onde se testou variadas entradas (*inputs*) e camadas (*layers*) se optou em considerar $\lambda = 0$, ou seja, os modelos foram ajustados pelo método usual de mínimos quadrados.

Os modelos Box-Jenkins do estudo serão ajustados com base na técnica proposta por Box e Jenkins (1976). Em cada instante de tempo t, existe um conjunto de valores que a série pode assumir e aos quais estão associadas possibilidades de ocorrência. Ela é utilizada para a previsão de séries temporais cujo processo estocástico não é estacionário. Logo, a série original passará por algumas diferenciações a fim de torná-la estacionária. O número necessário de diferença para tornar uma série estacionária é denominado ordem de integração (d). Os modelos Box-Jenkins são classificados em estacionários (média móvel - MA, autorregressivo – AR e média móvel autorregressivo – ARMA) e não estacionários (média móvel, autorregressivo e integrado - ARIMA). No entanto, se o comportamento da série de dados se repetir em um certo período os modelos são ajustados considerando a autocorrelação sazonal (sazonal autorregressivo - SAR(P), sazonal de média móvel - SMA(Q), sazonal autorregressivo de média móvel - SARMA (P,Q)S e sazonal autorregressivo integrado de média móvel - SARIMA (P,D,Q)S) (MORETTIN e TOLOI, 2006; BOX, JENKINS e REINSEL 2008). A estrutura geral SARIMA (p, d, q) x (P, D, Q)s é expressa por:

$$\phi_P(B)\Phi_P(B^s)\nabla^d\nabla_s^D Z_t = \theta_q(\mathrm{B})\theta_\mathrm{Q}(B^s)\epsilon_t. \quad (4)$$

Em que:

B - Representa um operador de defasagem;

B^s - Representa um operador sazonal de defasagem;

$\phi_p(B)$ - Representa o componente autorregressivo de ordem p;

$\theta_q(\mathrm{B})$ - Representa o componente de média móvel de ordem q;

$\Phi_P(B^s)$ - Representa o componente sazonal autorregressivo de ordem P;

$\Theta_Q(B^s)$ - Representa o componente sazonal média móvel de ordem Q;

$\epsilon_t = \phi_P(B)Z_t$ ruído branco.

d - Representa o número de diferenças.

$\nabla^d = (1-B)^d$– Representa diferença ordinária.

$\nabla_s^D = (1-B^s)^D$- Representa diferença sazonal.

Este operador diferença é definido como: $Z_t - Z_{t-1} = Z_t - BZ_t = (1-B)Z_t = \nabla Z_t$

Ressalta-se que os componentes da modelagem SARIMA serão analisados e, caso não sejam significativos, novos modelos serão ajustados retirando seus componentes, ou seja, poderão ser retiradas dos modelos as parcelas autorregressiva, média móvel, autorregressivo sazonal, média móvel sazonal. Sendo assim, busca-se o ajuste de modelos SARIMA, mas devido a parcimônia do modelo, bem como as características de ajuste, pode-se optar por modelos de ARIMA, ARMA, entre outros.

No modelo ARMA (autorregressivo de médias móveis), a série temporal Z_t é uma função de seus valores históricos e pelos termos aleatórios correntes e passados, seu modelo é representado na Equação 5:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots \phi_q Z_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots .. - \theta_q \epsilon_{t-q} \quad (5)$$

No modelo ARIMA (autorregressivo integrado de médias móveis) se a série temporal analisada não for estacionária em nível, devem-se aplicar diferenças até que a condição de estacionariedade se torne válida, no qual (d) é número de diferenças necessárias e (I) a ordem de integração, com isso seu modelo pode ser descrito como na Equação 6:

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots \phi_q W_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots .. - \theta_q \epsilon_{t-q} \quad (6)$$

em que $W_t = \Delta^d Z_t$. Na estrutura ARIMA (p, d, q) significa que esta possui p parâmetros auto-regressivos, d diferenciações a partir da série original e q parâmetros de médias móveis.

2.4 MEDIDAS DE QUALIDADE DE AJUSTE DOS MODELOS

Para se verificar os melhores ajustes dos modelos foram utilizados os critérios: erro quadrático médio e o critério de informação de Akaike – AIC (AKAIKE, 1974). O AIC é utilizado para comparar dois modelos e é expresso por (Equação 7):

$$AIC = -2\ln(L) + 2K \quad (7)$$

em que L é a função de verossimilhança e K é o número de parâmetros estimados.

O erro quadrático médio - EQM é definido por (Equação 8):

$$EQM = \frac{1}{n}\sum_1^n \varepsilon_i^2 \quad (8)$$

em que n é o número de prognósticos, com $\varepsilon_i^2 = (E_i - O_i)^2$, sendo E_i valor estimado e O_i valor observado no período i (MONTGOMERY e RUNGER, 2009; WILKS, 2006).

Por fim, comparou-se os valores previstos com os valores observados no período 01 a 05 de janeiro de 2014. A escolha do período de cinco dias para previsão se deu por causa da limitação do modelo GMDH que prevê cinco passos à frente somente. Para esta análise foi considerado o erro relativo percentual (Equação 9):

$$\boldsymbol{ERP} = \frac{x_p - x}{x} \times \mathbf{100} \quad (9)$$

onde $\boldsymbol{x_p}$ é o valor previsto e $\boldsymbol{x}$ é o valor observado no dia (BEVINGTON e ROBINSON, 2003).

3. RESULTADOS E DISCUSSÃO

As estatísticas descritivas das séries climáticas para o período de janeiro de 2011 a dezembro de 2013 são apresentadas na Tabela 2.

Tabela 2 – Estatísticas descritivas das séries temporais climáticas do município de Ariquemes (RO) no período de janeiro de 2011 a dezembro de 2013.

Séries Climáticas	Média	Mediana	Desvio padrão	Mínimo	Máximo
Temperatura máxima (°C)	32,05	32,10	2,52	23,90	38,10
Temperatura Mínima (°C)	23,75	23,90	1,51	15,30	28,10

A seguir são apresentados os modelos ajustados para as variáveis temperaturas máximas e mínimas do ar (ºC). Para se identificar o melhor ajuste buscou–se avaliar o erro quadrático médio de cada modelo em função do número de entradas (três ou quatro), quantidade de camadas intermediárias da rede neural tipo GMDH (um a três) e os tipos de funções de transferência (polinomial, sigmoide, função de base radial - RBF e tangente).

3.1 AJUSTE DO MODELO GMDH PARA TEMPERATURA MÁXIMA DIÁRIA (ºC)

Na Tabela 3 são apresentados os erros quadráticos médios dos modelos ajustados para a série de temperatura máxima diária do ar (%).

Tabela 3 – Valores dos erros quadráticos médios de acordo com o número de entrada, camadas intermediárias e funções de transferência para os modelos ajustados de temperatura máxima.

Modelo	Entradas	Camadas intermediárias	Funções de transferência			
			Polinomial	Sigmoide	RBF	Tangente
Um	Quatro	Três	3,3220	3,3417	3,3140*	3,3239
Dois	Quatro	Duas	3,3260	3,3446	3,3208	3,3278
Três	Quatro	Uma	3,4250	3,4426	3,4155	3,4337
Quatro	Três	Três	3,3710	3,3878	3,3644	3,3733
Cinco	Três	Duas	3,3593	3,3857	3,3550	3,3611
Seis	Três	Uma	3,4230	3,4409	3,4136	3,4312

* - menor erro quadrático médio observado

De acordo com os dados da Tabela 3, o modelo com quatro entradas, três camadas intermediárias e a função de transferência RBF apresentou menor erro quadrático médio (EQM = 3,3140). Observa-se ainda, que nos modelos ajustados com 4 entradas o número de camadas foi fundamental para a redução de EQM, quanto maior o número de camadas menor foi o EQM. No entanto, com três entradas esse comportamento não foi observado.

Os coeficientes do modelo GMDH com menor EQM (3,3140) estão descritos na Equação 10:

$$y = G(z_3, z_4) = 0{,}556 - 0{,}929z_3 + 2{,}064z_4 - 0{,}463z_3z_4 + 0{,}132z_3^2 - 1{,}216z_4^2 \quad (10)$$

Na Figura 2 são apresentados os valores observados, ajustados e previstos para a temperatura máxima do ar (ºC) no período de janeiro de 2011 a janeiro de 2014. Na Figura 3, os dados apresentados correspondem ao período de 01 dezembro de 2013 a 05 janeiro de 2014, somente.

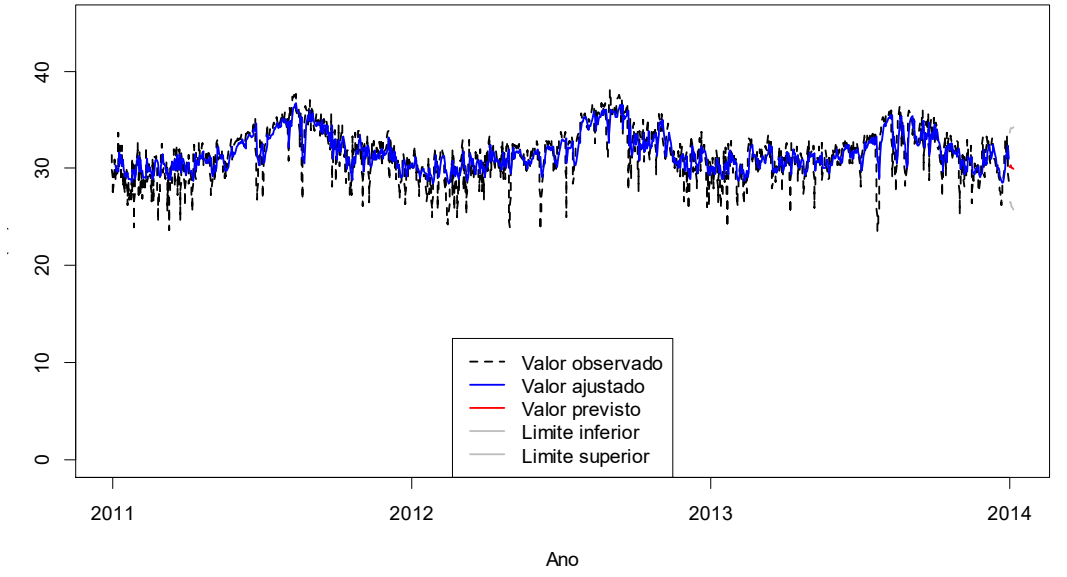

Figura 2 - Valores ajustados, observados, previstos e intervalo de confiança para a série de temperatura máxima diária pelo método GMDH, com quatro valores de entrada, três camadas intermediárias e função de transferência RBF.

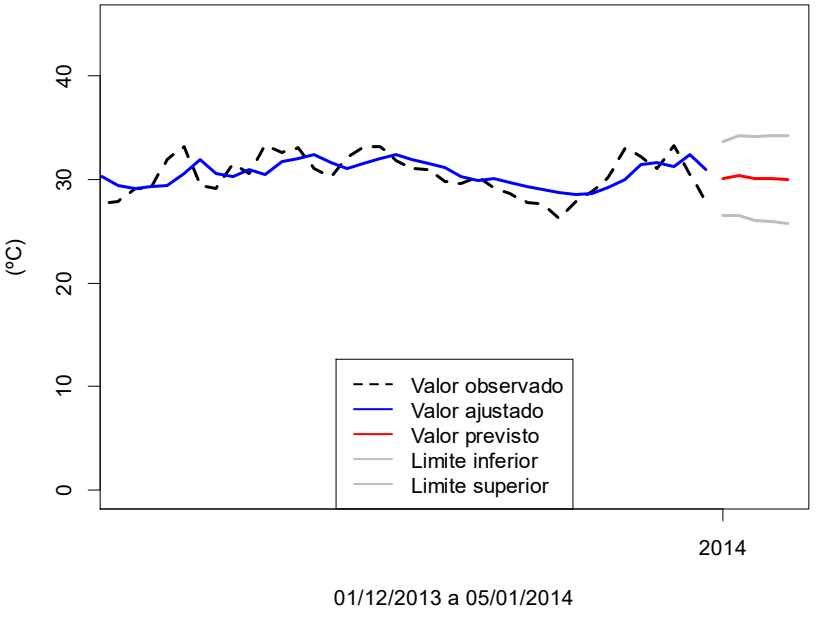

Figura 3 - Valores ajustados, observados, previstos e intervalo de confiança para a série de temperatura máxima diária pelo método GMDH, com quatro valores de entrada, três camadas intermediárias e função de transferência RBF, no período de 01/12/2013 a 05/01/2014.

3.2 AJUSTE DO MODELO GMDH PARA TEMPERATURA MÍNIMA DIÁRIA (ºC).

Na Tabela 4 são apresentados os erros quadráticos médios dos modelos ajustados para a série de temperatura mínima diária do ar (ºC).

Tabela 4 – Valores dos erros quadráticos médios de acordo com o número de entrada, camadas intermediárias e funções de transferência para os modelos ajustados de temperatura mínima.

Modelo	Entradas	Camadas intermediárias	Funções de transferência			
			Polinomial	Sigmoide	RBF	Tangente
Um	Quatro	Três	0,9615	0,9512	0,9462*	0,9679
Dois	Quatro	Duas	0,9778	0,9630	0,9628	0,9772
Três	Quatro	Uma	0,9897	1,0025	0,9881	0,9864
Quatro	Três	Três	0,9761	0,9524	0,9527	0,9748
Cinco	Três	Duas	0,9769	0,9621	0,9619	0,9763
Seis	Três	Uma	0,9889	1,0016	0,9872	0,9855

* - menor erro quadrático médio observado

De acordo com a Tabela 4, o modelo com quatro entradas, três camadas intermediárias e a função de transferência Base Radial (RBF) apresentou menor erro quadrático médio (EQM = 0,9462). Neste caso, nos modelos ajustados com três ou quatro entradas o número de camadas foi fundamental para a redução de EQM.

Os coeficientes do melhor modelo ajustado estão descritos na Equação 11:

$$y = G(z_1, z_4) = 1{,}131 - 7{,}033z_1 + 6{,}772z_4 - 0{,}160z_1z_4 + 3{,}949z_1^2 - 4{,}492z_4^2 \quad (11)$$

Na Figura 4 são apresentados os valores observados, ajustados e previstos para a temperatura mínima do ar (ºC) no período de janeiro de 2011 a janeiro de 2014, enquanto na Figura 5 o período representado refere-se a 01 dezembro de 2013 a 05 de janeiro de 2014.

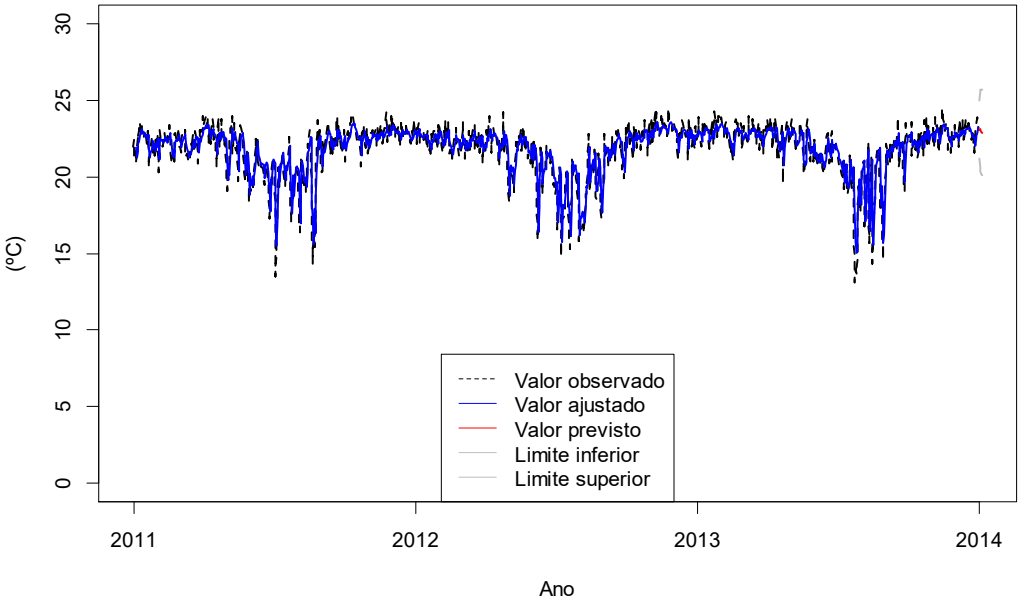

Figura 4 - Valores ajustados, observados, previstos e intervalo de confiança para a série de temperatura mínima diária pelo método GMDH, com quatro entradas, três camadas intermediárias e função de transferência RBF.

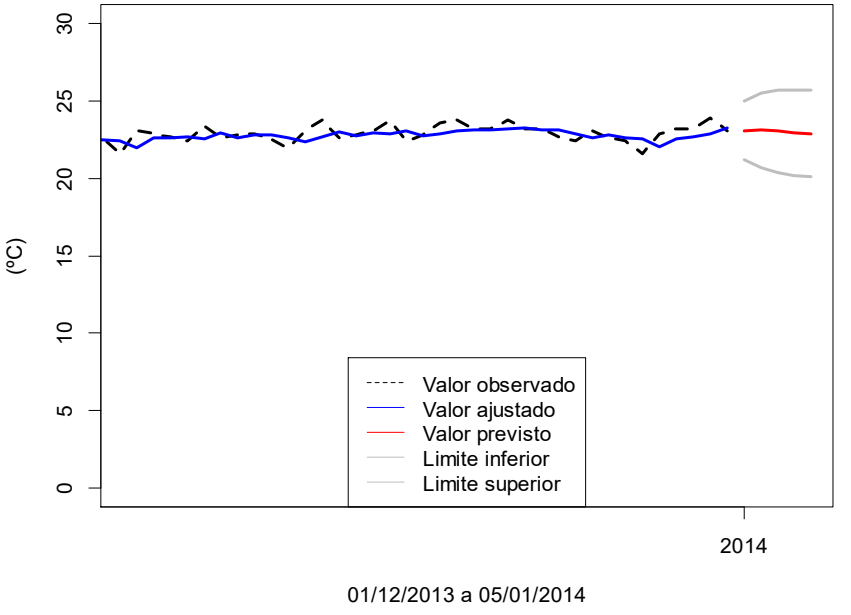

Figura 5 - Valores ajustados, observados, previstos e intervalo de confiança para a série de temperatura mínima diária pelo método GMDH, com quatro entradas, três camadas intermediárias e função de transferência RBF, no período de 01/12/2013 a 05/01/2014.

3.3 AJUSTE DO MODELO ARMA PARA TEMPERATURA MÁXIMA DO AR (ºC).

A seguir, na Tabela 5, são apresentados os indicadores de ajuste do modelo.

Tabela 5 - Histórico de construção do modelo ARMA para a variável temperatura máxima (ºC)

MODELO	AIC (Critério de informação de Akaike)	EQM (Erro Quadrático Médio)
(1,0,0)	4588,09	3,828
(2,0,0)	4517,73	3,583
(3,0,0)	4505,77	3,538
(3,0,1)	4446,95	3,342

Verificou-se que o melhor modelo ajustado foi o ARMA (3, 0, 1), que apresentou um erro quadrado médio de 3,342 e menor valor no AIC (4446,95). Os coeficientes do modelo foram significativos (Tabela 6).

Tabela 6 - Parâmetros estimados do modelo ARMA (3, 0, 1) para a série temperatura máxima do ar.

Parâmetro	Estimativa	s	t	p
Autorregressivo, *Lag* 1	1,308	0,034	38,483	0,000
Autorregressivo, *Lag* 2	-0,214	0,049	-4,326	0,000
Autorregressivo, *Lag* 3	-0,100	0,032	-3,125	0,002
Média móvel, *Lag* 1	-0,922	0,015	-61,443	0,000
Intercepto	31,300	0,681	45,972	0,000

O modelo encontrado para a série de temperatura máxima ARMA (3, 0, 1) pode ser descrito (Equação 12) como:

$$(1 - 1{,}3080\, B + 0{,}214\, B^2 + 0{,}100\, B^3)Z_t = (1 + 0{,}922B)\epsilon_t + 31{,}300 \qquad (12)$$

Na Figura 6 são apresentados os valores ajustados e observados para a série de temperatura máxima do ar, para o período de janeiro de 2011 a dezembro de 2013, e os valores previstos para os primeiros 5 dias de janeiro de 2014 (Figura 7).

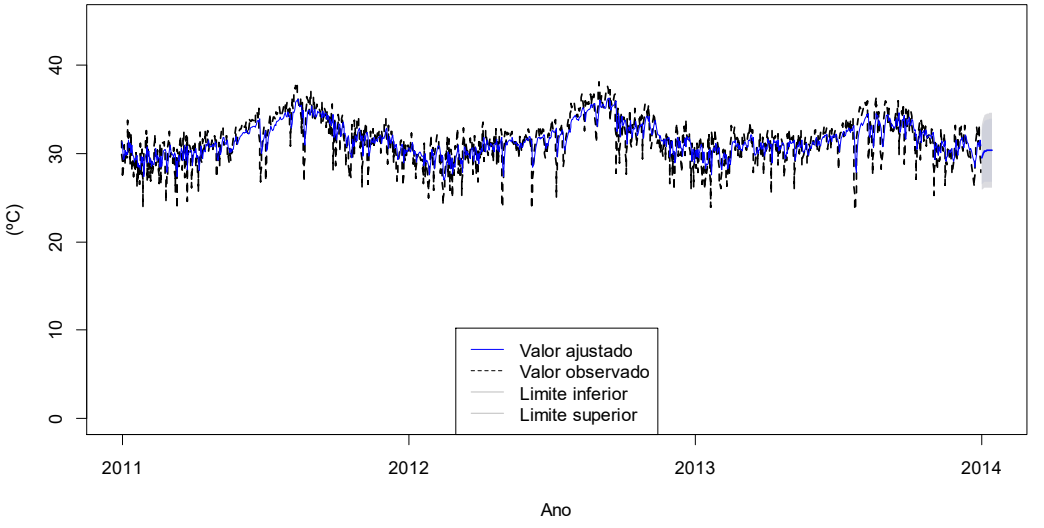

Figura 6 - Valores ajustados, observados, previstos e intervalo de confiança para a série de temperatura máxima do ar pelo método ARMA (3, 0, 1).

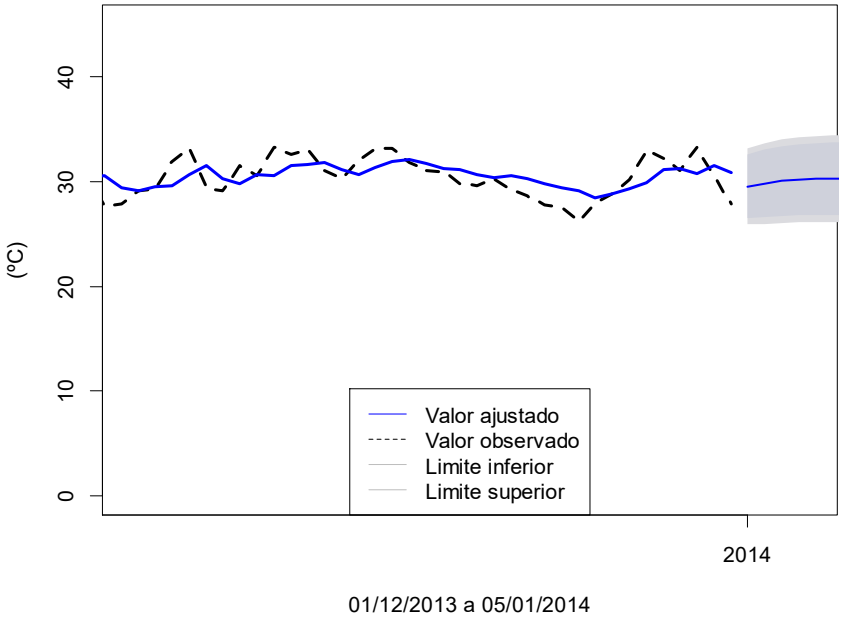

Figura 7 - Valores ajustados, observados, previstos e intervalo de confiança para a série de temperatura máxima do ar pelo método ARMA (3, 0, 1), no período de 01/12/2013 a 05/01/2014.

3.4 AJUSTE DO MODELO ARIMA PARA TEMPERATURA MÍNIMA DO AR

A seguir, na Tabela 7, serão apresentados os indicadores de ajuste do modelo.

Tabela 7 - Histórico de construção do modelo ARIMA para a variável temperatura mínima do ar.

MODELO	AIC (Critério de informação de Akaike)	EQM (Erro Quadrático Médio)
(1,0,0)	3165,2	1,045
(2,0,0)	3157	1,035
(3,0,0)	3153,79	1,030
(1,1,1)	3095,49	0,983

Verificou-se que o melhor modelo ajustado foi o ARIMA (1, 1, 1), que apresentou um erro quadrado médio de 0,983 e menor valor no AIC (3095,49). Os coeficientes do modelo foram significativos, sendo essas informações descritas na Tabela 8.

Tabela 8 - Parâmetros estimados do modelo ARIMA (1, 1, 1) para série temperatura mínima do ar.

Parâmetro	Estimativa	s	t	p
Autorregressivo, *Lag* 1	0,584	0,031	19,072	0,000
Média móvel, *Lag* 1	-0,921	0,013	-71,313	0,000

O modelo encontrado para a série de temperatura mínima do ar ARIMA (1, 1, 1) pode ser equacionado (Equação 13) como:

$$(1 - 0{,}584\,B)\nabla Z_t = (1 + 0{,}921B)\epsilon_t \qquad (13)$$

Na Figura 8 são apresentados os valores ajustados e observados para a série de temperatura do mínima do ar, para o período de janeiro de 2011 a dezembro de 2013, e os valores previstos para os primeiros 5 dias de janeiro de 2014 (Figura 9).

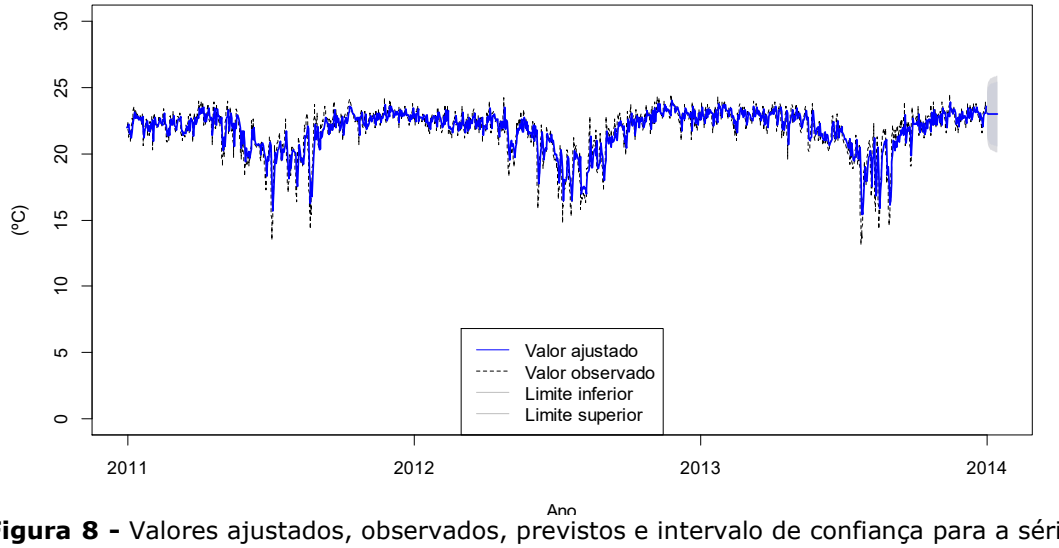

Figura 8 - Valores ajustados, observados, previstos e intervalo de confiança para a série de temperatura mínima do ar pelo método ARIMA (1, 1, 1).

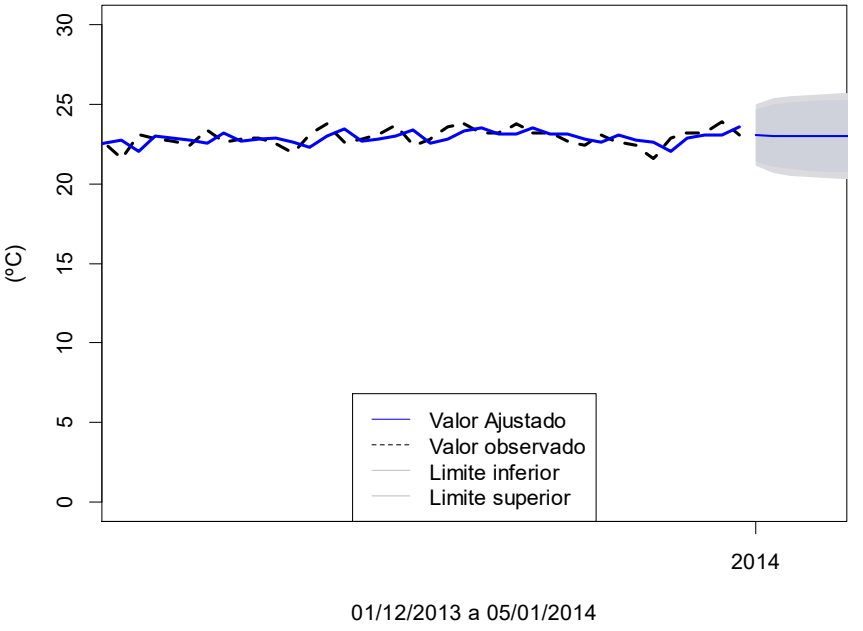

Figura 9 - Valores ajustados, observados, previstos e intervalo de confiança para a série de temperatura mínima do ar pelo método ARIMA (1, 1, 1), no período de 01/12/2013 a 05/01/2014.

3.5 COMPARAÇÃO DOS MODELOS BOX-JENKINS E REDE NEURAL GMDH: VALIDAÇÃO DAS PREVISÕES

Na Tabela 9 são apresentados os erros quadráticos médios dos modelos ajustados pelas técnicas Box-Jenkins e as redes neurais GMDH.

Tabela 9 – Valores dos erros quadráticos médios dos modelos ajustados nas técnicas Box-Jenkins e GMDH

Série Temporal	Box-Jenkins		GMDH	
	Modelo	EQM	Modelo	EQM
Temperatura Máxima do Ar	ARMA (3, 0, 1)	3,342	4 Entradas, 3 Camadas intermediárias e função RBF	3,314
Temperatura Mínima do Ar	ARIMA (1, 1, 1)	0,983	4 Entradas, 3 Camadas intermediárias e função RBF	0,946

Observa-se na Tabela 9, que nos modelos GMDH as temperaturas máximas e mínimas obtiveram menores valores do EQM 3,314 e 0,946, respectivamente. De um modo geral os resultados da modelagem por rede neural tipo GMDH, em comparação com a modelagem Box-Jenkins, foram satisfatórios visto que apresentaram menores erros quadráticos médios, correspondendo ao um ganho na predição de dados.

Os modelos foram ajustados com os dados do período de 01 de janeiro de 2011 a 31 de dezembro de 2013 e para as devidas comparações de valores observados e previstos foram separados os dados entre 01 e 05 de janeiro de 2014. Na Tabela 10 são apresentados os valores observados e previstos pelos modelos Box-Jenkins e GMDH.

Tabela 10 - Comparação dos valores previstos e observados no período de 01 a 05 de janeiro de 2014 para as séries temporais ajustadas

	Temperatura Máxima °C					Temperatura Mínima °C				
Dia	GMDH	ERP (%)	ARMA (3, 0, 1)	ERP (%)	VO	GMDH	ERP (%)	ARIMA (1, 1, 1)	ERP (%)	VO
01	30,10	2,59	29,53	4,34	30,9	23,14	2,84	23,06	2,49	22,5
02	30,33	7,53	29,79	9,18	32,8	23,09	0,04	23,03	0,30	23,1
03	30,09	10,18	30,05	10,30	33,5	22,96	3,53	23,02	3,28	23,8
04	30,05	5,50	30,16	5,16	31,8	22,87	1,64	23,01	2,27	22,5
05	29,98	5,13	30,23	4,34	31,6	22,79	0,91	23,01	0,04	23

Nota: Erro Relativo Percentual = ERP; Valor Observado = VO

Observa-se na Tabela 10, que os erros relativos percentuais para o modelo GMDH em relação à temperatura máxima do ar foram menores do que os valores dos erros no modelo ARMA (3,0,1) nas previsões dos três primeiros dias. Já para a temperatura mínima os valores dos ERP's se alternaram, sugerindo a proximidade dos valores ajustados dos modelos.

4. CONSIDERAÇÕES FINAIS

A modelagem por redes neurais tipo GMDH apresentou resultados similares aos resultados dos modelos Box-Jenkins, constituindo-se, assim, numa opção para previsão de séries temporais climáticas. Modelos GMDH com quatro entradas e três camadas intermediárias se mostraram mais eficientes, ou seja, menor erro quadrático médio. A função de transferência de base Radial (RBF) foi mais eficiente do que as funções sigmoide, polinomial e tangente.

Futuros estudos são sugeridos com a aplicação da modelagem GMDH revisada para estas séries e aplicação dos parâmetros regularizadores no método de mínimos quadrados.

AGRADECIMENTOS

À Fundação de Amparo ao Desenvolvimento das Ações Científicas e Tecnológicas e a Pesquisa de Rondônia - FAPERO, pelo auxílio financeiro através do Programa de bolsas de mestrado e doutorado para docentes (MS / DR / DOCENTE / FAPERO / CAPES - 07/2014).

REFERÊNCIAS BIBLIOGRÁFICAS

ACOCK, M. C.; PACHEPSKY, Y. A. A. Estimating Missing Weather Data for Agricultural Simulations Using Group Method of Data Handling. Journal of Applied Meteorology, v. 39, n. 1, p. 1176-1184, 2000.

AKAIKE, H. A new look at the statistical model identification. IEEE Transactions on Automatic Control, v. 19, n. 6, p. 716-723, Dec. 1974.

ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; GONÇALVEZ, J. L. M.; SPAROVEK, G. Köppen´s climate classification map for Brazil. Meteorologische Zeitschrift, v. 22, n. 6, p. 711-728, 2014.

BARRETO, J. M. Introdução às redes neurais artificiais. Florianópolis: UFSC, 2002.

BEVINGTON, P. R.; ROBINSON, D. K. Data reduction and error analysis for the physical sciences. 3ª Ed. New York: McGraw-Hill. 2003. 320p.

BOX, G. E. P.; JENKINS, G. E. Times series analysis: forecasting and control. Revise edition. California: Holden-Day, 1976.

BOX, G. E. P; JENKINS, G. M.; REINSEL, G. C. Time Series Analysis Forecasting and Control. 4 Edição. JOHN WILEY & SONS, INC., PUBLICATION: Hoboken, New Jersey. 2008. 746p.

BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. Redes neurais artificiais: Teoria e aplicações. 2ª Edição – [Reimpressão]. Rio de Janeiro: LTC, 2016. 225p.

CARVALHO, R. L. S.; NASCIMENTO, B. I. S.; QUERINO, C. A. S.; SILVA, M. J. G.; DELGADO, A. R. S. Comportamento das séries temporais de temperatura do ar, umidade e precipitação pluviométrica no município de Ariquemes (Rondônia-Brasil). Revista Brasileira de Climatologia, v. 18, n. 12, p. 123- 142, 2016.

COUTINHO, E. R.; SILVA, R. M.; DELGADO, A. R. S. Utilização de Técnicas de Inteligência Computacional na Predição de Dados Meteorológicos. Revista Brasileira de Meteorologia, v. 31, n. 1, p. 24-36, 2016.

DAG, O.; YOZGATLIGIL, C. GMDH: An R Package for Short Term Forecasting via GMDH-Type Neural Network Algorithms. The R Journal, v. 8, n. 1, p. 379-386, 2016.

DANTAS, D.; LUZ, T. M. O.; SOUZA, M. J. H.; BARBOSA, G. P.; CUNHA, E. G. S. Uso de Redes Neurais Artificiais na previsão da precipitação de períodos chuvosos. Revista Espinhaço, v. 5, n. 1, p. 11-18, 2016.

DUBREUIL, V.; FANTE, K. P.; PLANCHON, O.; SANT´ANNA NETO, J. L. Os tipos de climas anuais no Brasil: uma aplicação da classificação de Köppen de 1961 a 2015. Revista Franco-Brasileira de Geografia, v. 31, n. 37, p. 1-20, 2018.

FARIA, J. C.; GROSJEAN, P.; JELIHOVSCHI, E. G.; PIETROBON, R. FARIAS, P. S.; KRAMER, P. A. R.; LIMA, S. P. Tinn-R GUI/Editor for R Environment. 2016. Disponível em http://www.sciviews.org/Tinn-R/. Acesso em 01/11/2016.

FARLOW, S. J. The GMDH Algorithm of Ivakhnenko. The American Statistician, v. 35, n. 4, p. 210-215, 1981.

FERRAZ, R. C. Estimativa de evapotranspiração de referência utilizando redes neurais artificiais para o estado do Rio Grande do Sul. Revista Tecnológica, v. 23, n. 1, p. 25-31, 2014.

FERREIRA, L.; MOURA, G. L.; BORENSTEIN, D.; FISCHMANN, A. A. Utilização de redes neurais artificiais como estratégia de previsão de preços no contexto de agronegócio. RAI – Revista de Administração e Inovação, v. 8, n. 4, p. 6-26, out./dez. 2011.

HAKYKIN, S. Redes neurais - princípios e prática. 2ª Ed. Porto Alegre: Bookman, 2001. 902p.

HEATON, J. Programming Neural Networks with Encog 2 in Java. Heaton Research, Inc, 2010.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍTICA. Cidades. 2017. Disponível em < https://cidades.ibge.gov.br/brasil/ro/ariquemes/panorama>. Acesso em 11/12/2017.

INSTITUTO NACIONAL DE METEOROLOGIA. Dados do Monitoramento da estação Automática de Ariquemes/RO. Comunicação pessoal via e-mail: <sadmet.df@inmet.gov.br> em 18 de jun. 2014.

IVAKHNENKO, A. G. Group Method of Data Handling – A Rival of the Method of Stochastic Approximation. Soviet Automatic Control, v. 13, n. 1, p. 43-71, 1966.

IVAKHNENKO, A. G.; IVAKHNENKO, G. A. The review of problems solvable by algorithms of the Group Method of Data Handling (GMDH). Pattern Recognition and Image Analysis, v. 5, n. 4, p. 527-535, 1995.

IVAKHNENKO, A. G.; IVAKHNENKO, G. A. Problems of Further Development of the Group Method of Data Handling Algorithms. Part I. Pattern Recognition and Image Analysis, v. 10, n. 2, p. 187–194, 2000.

KONDO, T.; UENO, J. Feedback GMDH-type neural network and its application to medical image analysis of liver cancer. International Journal of Innovative Computing, Information and Control, v. 8, n. 3(B), p. 2285-2300, 2012.

KOVACS, Z. L. Redes Neurais Artificiais: Fundamentos e Aplicações. 4ª Edição, São Paulo: Livraria da Física, 2002. 174p.

MONTGOMERY, D. C.; RUNGER, G. C. Estatística aplicada e probabilidade para engenheiros. Tradução: Verônica Calado. 4ª. Edição – Rio de Janeiro: LTC, 2009.

MORETTIN, P. A.; TOLOI, C. M. Análise de séries Temporais. 2ª edição revisada e ampliada – São Paulo: Edgard Blucher, 2006.

MURTGATH, F. Neural networks and related "massively parallel" methods for statistics: a short review, International Statistical Review, v. 62, n. 3; p. 275-88, 1994.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2016. Disponível em <https://www.R-project.org/>. Acesso em 01/11/2016.

SAMSUDIN, R.; SAAD, P.; SHABRI, A. Combination of Forecasting Using Modified GMDH and Genetic Algorithm. International Journal of Computer Information Systems and Industrial Management Applications, v. 1, n. 1, p. 170-176, 2009.

SHABRI, A.; SAMSUDIN, R. A Hybrid GMDH and Box-Jenkins Models in Time Series Forecasting. Applied Mathematical Sciences, v. 8, n. 62, p. 3051 – 3062, 2014.

TIKHONOV, A. N. Solution of incorrectly formulated problems and the regularization method. Soviet Mathematics Doklady, v. 4, n. 1, p. 1035-1038, 1963.

TURCO, J. E. P.; CHAVES, L. H. Evapotranspiração de referência pelas equações de Hargreaves na mesorregião do Triângulo Mineiro e Alto Paranaíba – MG. Revista Brasileira de Agricultura Irrigada, v. 12, n. 5, p. 2964 – 2975, 2018.

WILKS, D. S. Statistical methods in the atmospheric sciences. 2ª Ed. Amsterdam: Academic Press - Elsevier. 2006. 627p