

Comparativo de desempenho entre os modelos ARIMA e redes neurais artificiais na predição do parâmetro pH da qualidade da água

Anderson Francisco de Sousa Almeida

Programa de Pós-Graduação em Computação Aplicada

Universidade Federal do Pará

andersonchico@gmail.com

Abstract. *Statistical and computational models are used in the data mining process to extract nontrivial knowledge from a set of data making it important for planning in decision making. Thus, the implementation of these models to predict parameters of water quality is necessary to scale the supply of a quality water. This paper aims to compare the performance of ARIMA models and artificial neural networks when used for self-regression of a time series in order to predict the pH parameter of water quality, based on the data mining technique. The results showed that the RNA model had the best error rates compared to the ARIMA model.*

Resumo. *Os modelos estatísticos e computacional são usados no processo de mineração de dados para a extração de conhecimento não trivial de um conjunto de dados tornando-se importante para o planejamento nas tomadas de decisões. Dessa forma, a implementação desses modelos para predição de parâmetros da qualidade da água é necessária para dimensionar a oferta de uma água qualidade. Diante disso, este artigo tem o objetivo comparar o desempenho dos modelos ARIMA e redes neurais artificiais quando utilizados para auto regressão de uma série temporal no intuito de prever o parâmetro pH da qualidade da água, baseado na técnica de mineração de dados. Os resultados demonstraram que o modelo de RNA apresentou as melhores taxas de erros do que o modelo ARIMA.*

1. INTRODUÇÃO

Os modelos estatísticos e computacional são usados no processo de mineração de dados para a extração de conhecimento não trivial de um conjunto de dados tornando-se

importante para o planejamento nas tomadas de decisões. Dessa forma, a implementação desses modelos para predição de parâmetros da qualidade da água é necessária para dimensionar a oferta de uma água qualidade. Liu et al. (2013) afirma que a predição de parâmetros da qualidade da água oferece aos gestores hídricos a oportunidade de tomar ações preventivas no tratamento da água.

Segundo Braga et al. (2015), o planejamento e a gestão dos recursos hídricos dependem das informações confiáveis para garantir qualidade na tomada das decisões, e consequentemente, preservar a sustentabilidade do ecossistema. Ressalta, que a falta de informação produz um custo maior do que adquirir informação por meios tecnológicos.

Braga et al. (2015), cita

[...] A informação sobre a qualidade da água é necessária para que se conheça a situação dos corpos hídricos com relação aos impactos antrópicos na bacia hidrográfica e é essencial para que se planeje sua ocupação e seja exercido o necessário controle dos impactos [...]

Assim, essas informações são produzidas pelo monitoramento hidrológico sobre a qualidade dos recursos hídricos superficiais, com a utilização de sensores automáticos ligados a uma plataforma de coleta de dados sobre os parâmetros básicos da qualidade da água como pH, temperatura, oxigênio dissolvido (OD), turbidez e condutividade elétrica, sendo transmitido para o centro de monitoramento (ANA, 2018).

Para que as informações contidas em um conjunto de dados sejam de fácil entendimento e aplicação, espera-se que a representação do conhecimento seja mais próxima da linguagem humana. Por isso, diversas metodologias de análises são utilizadas como ferramentas para agilizar o processo de análise de dados. Entre as metodologias está a denominada Descoberta de Conhecimento em Base de Dados, do inglês Knowledge Discovery in Databases (KDD). No contexto KDD está a Mineração de Dados (Data Mining) que tem o objetivo de extrair informação válida a partir de grandes bases de dados, apoiando a tomada de decisões (GUIMARÃES, 2005).

Segundo Solanki (2015), as técnicas de mineração de dados, modelos estatísticos e computacionais são usados para análise preditiva sobre uma série temporal de dados para obter conhecimento futuro sobre os parâmetros da qualidade da água.

O presente artigo tem como objetivo comparar o desempenho dos modelos ARIMA e redes neurais artificiais quando utilizados para auto regressão de uma série temporal no intuito de prever o parâmetro pH da qualidade da água, baseado na técnica de mineração de dados.

2. METODOLOGIA

2.1. Área de estudo

A região hidrográfica do Tocantins – Araguaia que possui uma área de 920 mil km², abrangendo os Estados do Pará, Goiás, Tocantins, Maranhão, Mato Grosso e o Distrito Federal, conforme figura 1, totalizando 409 municípios e com população estimada em 8.6 milhões de habitantes (IBGE, 2010). O principal corpo d'água da área de estudo é o rio Tocantins, que nasce no Estado de Goiás e desagua na Baía do Marajó no Estado do Pará, e tem como rios tributários o Araguaia, das Almas e Maranhão (ANA, 2015).

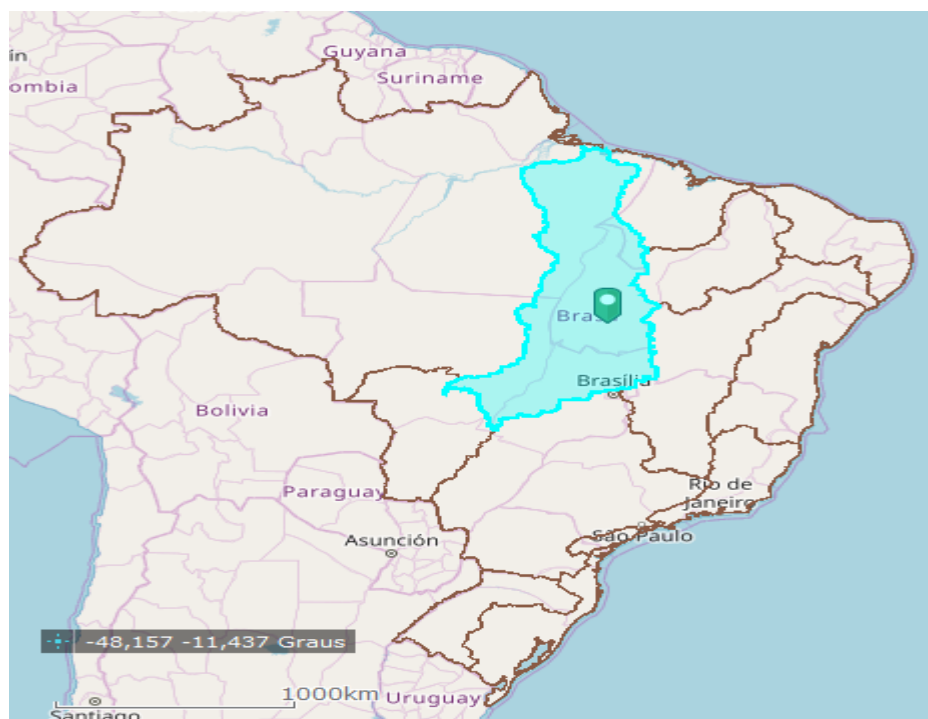


Figura 1: Região Hidrográfica Tocantins – Araguaia. **Fonte:** Sistema Nacional de Informações Hídricas – ANA.

A região hidrográfica do Tocantins – Araguaia possui como atividades socioeconômicas que podem afetar a qualidade da água na região são: a geração de energia elétrica com capacidade, aproximada, de 13 GW de potencial hidroelétrico (ANAEEL, 2013), a navegação com 4000 km de rios navegáveis, o cultivo irrigado de milho, arroz, feijão, soja e algodão, a mineração com a extração de minérios como ouro, amianto, cobre, níquel, bauxita, ferro, manganês, prata e cassiterita, a pesca de subsistência para o sustento dos ribeirinhos, a aquicultura para fins comerciais e o turismo (ANA, 2014).

2.2. Materiais

2.2.1. Ferramenta Hidroweb

Os dados utilizados na pesquisa estão disponíveis de forma pública na plataforma web Hidroweb versão 2.0.0.4, desenvolvida pela ANA e faz parte do Sistema Nacional de Informações sobre Recursos Hídricos (SNIRH). Nesta ferramenta é possível coletar dados das series históricas dos monitoramentos das estações, quanto aos aspectos pluviométrico e/ou fluviométrico, selecionando os critérios de busca como código da estação, nome da estação, bacia, sub-bacia, Estado (UF), responsável (Sigla) e operadora (Sigla), sendo exibido na tela os resultados da consulta com opção de *download* em formato de arquivos do tipo excel(.CSV), access (.MDB) ou texto (.TXT).

2.2.2. PHP (Hypertext Preprocessor) e Apache

Para as etapas referente ao processo de descoberta do conhecimento foram utilizadas as seguintes ferramentas: linguagem de programação PHP (Hypertext Preprocessor) que é uma linguagem de script de código aberto adequada para o desenvolvimento de aplicações web, de acordo com DALL’OGLIO (2018) “estima-se que o PHP seja utilizado em mais de 80% dos servidores web existentes[...]”. Para hospedar a aplicação web, utilizou-se o servidor web Apache que é o “[...] responsável por entregar as páginas quando o navegador solicitar. ” (BENTO, 2013), e está inserido no aplicativo XAMPP que contém um pacote de serviços facilitando o desenvolvimento de aplicações web com PHP.

2.2.3. Software R e IDE R-Studio

O software R é “um ambiente de software livre para computação estatística e gráficos” (R-PROJECT, 2019), que possui uma coleção de pacotes adicionais como

resultado de uma rede mundial colaborativa de vários pesquisadores responsáveis pela avaliação e atualização de novos pacotes (KATAOKA et al., 2008), contribuindo para outras áreas, além da estatística, como a Engenharia, Economia, Biologia, Ciência da Saúde e Ciências Sociais (KONRATH et al., 2018).

Para facilitar a usabilidade do software R é utilizado uma ferramenta de ambiente de desenvolvimento interativo (IDE), o R-Studio, que “é um software de código aberto, disponível gratuitamente na Internet para sistemas Windows, MacOS, Linux e Unix” (FOX e WEISBERG, 2019), que reúne diversas funcionalidades como importação de dados, visualização de comandos, funções, resultados e gráficos, além da geração de documentos (HENNING et al., 2016), tornando-se um ambiente atrativo para os discentes e pesquisadores, contribuindo para o ensino e a pesquisa (MINUTO et al., 2015).

2.3. Coleta de dados

Conforme mencionado, na ferramenta web Hidroweb foram selecionadas as estações pertencentes a região hidrográfica Tocantins – Araguaia, resultando em 265 arquivos no formato .CSV. Para facilitar a etapa de pré-processamento, foi desenvolvido uma aplicação em PHP para unir os arquivos, executando a leitura e cópia do conteúdo de cada arquivo .CSV e escrevendo no novo arquivo .CSV os campos data, que indica a sequência temporal de coleta dos valores dos parâmetros e os campos pH, turbidez, condutividade elétrica, DBO (demanda bioquímica de oxigênio) e OD (oxigênio dissolvido), que segundo CONAMA (2005), são parâmetros da qualidade da água.

2.4. KDD (Knowledge Discovery in Database)

FAYYAD et al. (1996) define KDD como “o processo, não trivial, de extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir dos dados armazenados em um banco de dados”. Portanto, o processo de descoberta de conhecimento usa técnicas de busca para produzir conhecimento novo e útil. As etapas do processo KDD são: pré-processamento (seleção, limpeza, codificação e transformação dos dados), mineração de dados (data mining) e interpretação/avaliação, conforme a figura 2.

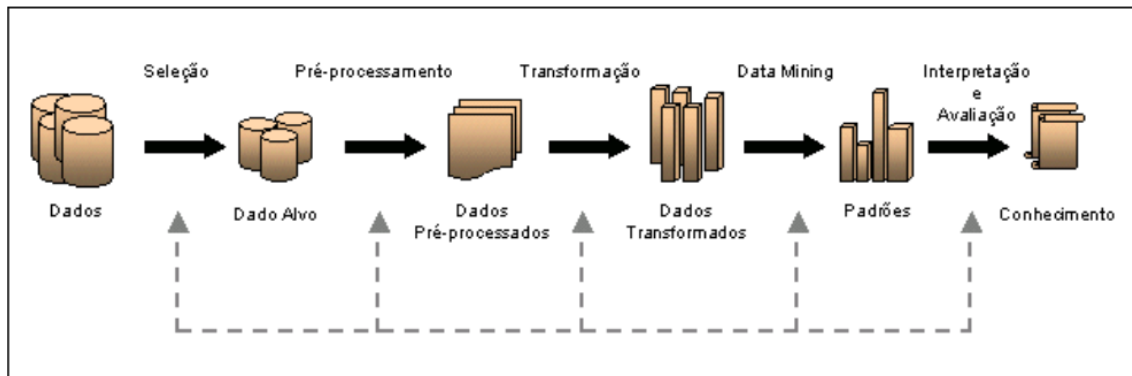


Figura 2: Etapas do processo KDD. **Fonte:** Adaptado de FAYYAD et al. (1996).

2.4.1. Pré-processamento dos dados

Nesta etapa para a descoberta do conhecimento, utilizamos a ferramenta R-Studio para importar e visualizar os dados da pesquisa, criar a série temporal, observar o resumo estatísticos da série e testar a condição de estacionariedade da série.

A seguir as etapas do pré-processamento de dados:

- **Importação dos dados:** Os dados foram carregados no R-Studio, função *read.csv()*, com os rótulos data, pH, turbidez, condutividade elétrica, DBO e OD, sendo possível observar a presença de registros com valores ausentes, a maioria para o rótulo condutividade elétrica e a minoria para o rótulo pH.
- **Criação da série temporal:** Os dados apresentam uma frequência temporal indicada pelas datas das coletas, assim, criou-se uma série temporal, função *ts()*, com 505 amostras considerando o período de 1976 a 2017 com frequência mensal, rótulo pH e limpeza dos dados ausentes substituindo os valores ausentes pelo valor constante 1, conforme figura 3.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1976	231	28	309	260	260	11	237	237	237	151	217	248
1977	217	11	227	309	207	266	237	237	291	151	203	227
1978	207	260	327	222	227	237	281	318	314	97	66	150
1979	146	324	337	273	325	140	170	165	223	184	337	1
1980	182	152	318	271	145	165	145	207	241	237	131	289
1981	265	165	196	243	306	212	310	199	106	279	319	316
1982	165	221	223	254	248	149	223	73	195	185	206	101
1983	160	228	259	176	198	272	189	313	154	268	209	189
1984	209	242	118	251	198	320	272	236	280	53	39	39
1985	291	237	107	75	28	140	309	300	39	75	75	18
1986	18	53	13	1	1	1	184	184	13	227	217	
1987	217	68	309	260	271	248	68	151	96	96	28	28
1988	68	143	159	150	117	217	251	184	257	257	253	1
1989	116	331	323	295	309	309	303	311	314	297	1	259
1990	284	325	1	1	217	271	214	1	281	323	302	259
1991	250	273	204	320	201	251	228	238	258	1	184	217
1992	182	284	268	270	1	248	67	260	304	190	101	225
1993	265	212	286	258	228	313	227	234	230	321	300	172
1994	292	251	296	208	204	1	239	234	257	234	193	227
1995	217	239	211	146	230	260	1	254	219	261	233	260
1996	242	222	248	276	1	302	253	170	233	319	1	284
1997	233	220	250	174	241	277	187	199	271	275	329	263

Figura 3: Parte da série temporal utilizada no experimento.

- **Resumo estatístico:** São observados os valores mínimo, do primeiro quadrante, da mediana, da média, do terceiro quadrante e o valor máximo da série temporal utilizando a função *summary()*.
- **Teste de série temporal estacionaria:** É observado se a serie possui a média e variância constante durante o tempo, utilizando a função *uk.kpss()* do pacote *urca*, pois a serie ser estacionaria é uma condicionante para aplicação do modelo auto regressão ARIMA.

2.4.2. Mineração de dados (Extração de Padrões)

Nesta etapa o objetivo é aplicar modelos estatístico e computacional, respectivamente, ARIMA e redes neurais artificiais, para extrair o conhecimento dos dados, através da tarefa de predição de valores futuros dos parâmetros da qualidade da água.

ARIMA – Autoregressive Integrated Moving Average

Segundo Cardoso e Cruz (2016), no modelo Arima, o processo é estacionário e o valor futuro de uma variável, é representado por uma função linear das observações

passadas e do erro aleatório, sendo este, uma variável independente distribuída com média igual a zero e variância constante.

O modelo ARIMA possui dois modelos aplicados para a série temporal estacionária que são: o Auto-regressivo (AR) e a Média Móvel (MA).

No Auto-regressivo (AR) a aproximação de uma série temporal é realizada quando os valores futuros estão relacionados linearmente com os componentes passados da série temporal e na Média Móvel (MA) “[...] reflete uma relação linear entre as componentes passadas do erro aleatório e a previsão dos próximos valores da série temporal [...]” (CARDOSO e CRUZ, 2016).

Para identificar o modelo ARIMA mais adequado a série temporal, pode-se usar as métricas Akaike Information Criteria (AIC e AICC) e Bayesian Information Criteria (BIC), na qual o melhor modelo é aquele que apresentar o menor valor (PINTO et. al., 2015). Essas métricas são executadas no R-Studio utilizando a função *auto.arima()* do pacote *forecast*.

Redes Neurais Artificiais - RNA

É uma técnica computacional que simula o processamento de informações como o cérebro humano, composta por neurônios artificiais unidas pelas conexões sinápticas (ZHANG, PATUWO, HU, 1998). O principal benefício da rede neural é a generalização, pois produz saídas adequadas para diferentes entradas durante o treinamento da rede e, com isso, pode ser aplicada para um grande conjunto de dados (HAYKIN, 2001). A estrutura da rede neural apresenta as entradas (x), pesos (w), função de ativação (linear, escada ou sigmoide) e a saída (y) (RUSSEL E NORVING, 2013), conforme a figura 4. No R-Studio pode ser executada pela função *nnetar()* do pacote *forecast*, neste caso, a rede neural é do tipo auto-regressiva.

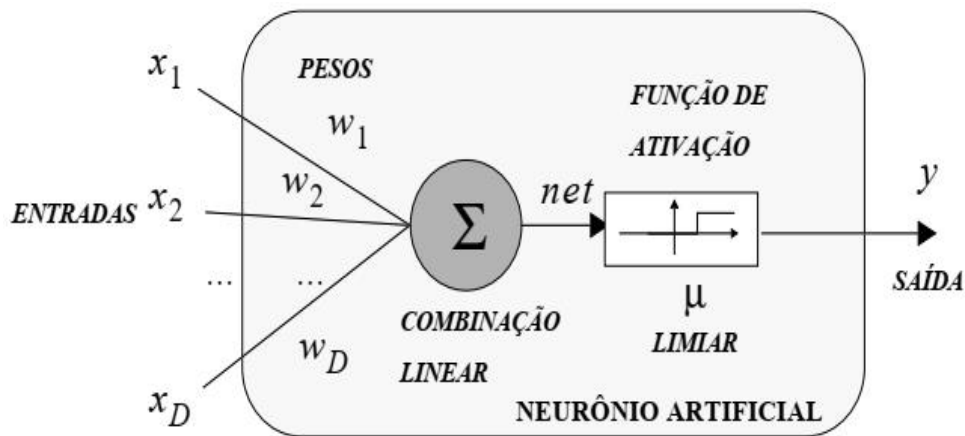


Figura 4: Modelo de um neurônio. **Fonte:** Adaptado de RAUBER (2005).

2.4.3. Validação dos modelos ARIMA e RNA

Neste experimento foi utilizado o método Hold Out, na qual separa o conjunto de dados em dois subgrupos um para treino e outro para teste, e possui a característica de rápida aprendizagem (SCHREIBER, 2017).

Em resumo, a série temporal contendo os anos de 1976 a 2017 foi dividida em duas subséries temporal: uma subsérie para treinamento, contendo os anos de 1976 a 2015 e outra subsérie para teste contendo os anos de 2016 a 2017, com objetivo testar a precisão de acerto da série temporal usado na série de treinamento. Ambos os modelos foram aplicados sob as mesmas condições tanto para treinamento quanto para teste e comparadas as métricas de erros na subsérie de teste.

Para verificar as métricas de erros dos modelos foi utilizada a função *accuracy()*, pertencente ao pacote *forecast* no R-Studio, que apresenta as seguintes métricas de erro: ME (Erro Médio), MAPE (Erro Percentual Médio Absoluto), MAE (Erro Médio Absoluto), RMSE (Erro Quadrático Médio), MPE (Erro Percentual Médio) e MASE (Erro Escalonado Médio Absoluto).

3. RESULTADOS E DISCUSSÃO

3.1. Comportamento da Série Temporal

A análise visual na figura 5 permite observar preliminarmente o comportamento da série temporal em estudo. A tabela 1 apresenta, quantitativamente, o resumo estatístico sobre os dados do parâmetro pH.

Mínimo	1º Quadrante	Mediana	Media	3º Quadrante	Máximo
1.0	180.0	228.0	213.5	272.0	344.0

Tabela 1: Resumo estatístico da série temporal.

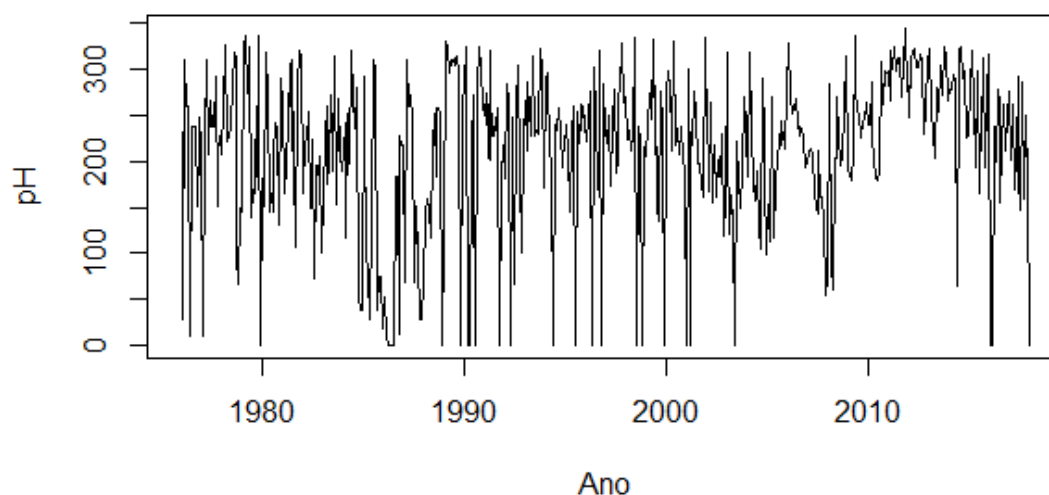


Figura 5: Comportamento da série temporal.

A série temporal é estacionária quando o valor do teste de estacionariedade é menor que 0.05. O resultado no primeiro teste foi de 0,9266 caracterizando que a serie não é estacionária, desse modo, foi aplicado a técnica da diferenciação para ajustar a serie temporal, e um segundo teste, resultando no valor de 0,0228, conforme a tabela 2, que contém os valores dos testes. Assim, quanto ao comportamento da série temporal utilizada no estudo, não é estacionaria, sendo necessário ajuste com a técnica de diferenciação.

Teste	Valor
1º	0,9266
2º	0,0228

Tabela 2: Resultado do teste de estacionariedade.

3.2. Comparativo de desempenho entre os modelos ARIMA e RNA

Na tabela 3, as métricas de avaliação de erro indicam que o modelo de RNA apresentou melhor desempenho de acurácia em relação ao modelo ARIMA. Ressalta-se, que entre as seis métricas apresentadas, em apenas duas (MAE e MASE), o modelo ARIMA apresentou melhor avaliação de acurácia.

	ME	RMSE	MAE	MPE	MAPE	MASE
ARIMA	-47.9450254	92.28130	62.56932	-2163.7545	2168.7606	0.7254774
RNA	-25.779651973	90.014474	63.817876	-2008.37689	2022.36249	0.73995406

Tabela 3: Métricas de desempenho entre os modelos ARIMA e RNA.

A figura 5, apresenta a previsão do modelo ARIMA para a série temporal do pH, na qual é observado que o ponto de predição (linha azul) está dentro do intervalo de predições que indicando que o modelo foi adequado para série temporal experimentada. Na figura 6, apresenta a previsão do modelo RNA para a série temporal do pH, na qual é observado que a predição do modelo RNA é pontual, não existindo intervalo de predições. A figura 7, apresenta a predição autoregressiva dos modelos ARIMA e RNA na série temporal do parâmetro pH, indicando que ambos estão próximos dos valores reais para o período de teste.

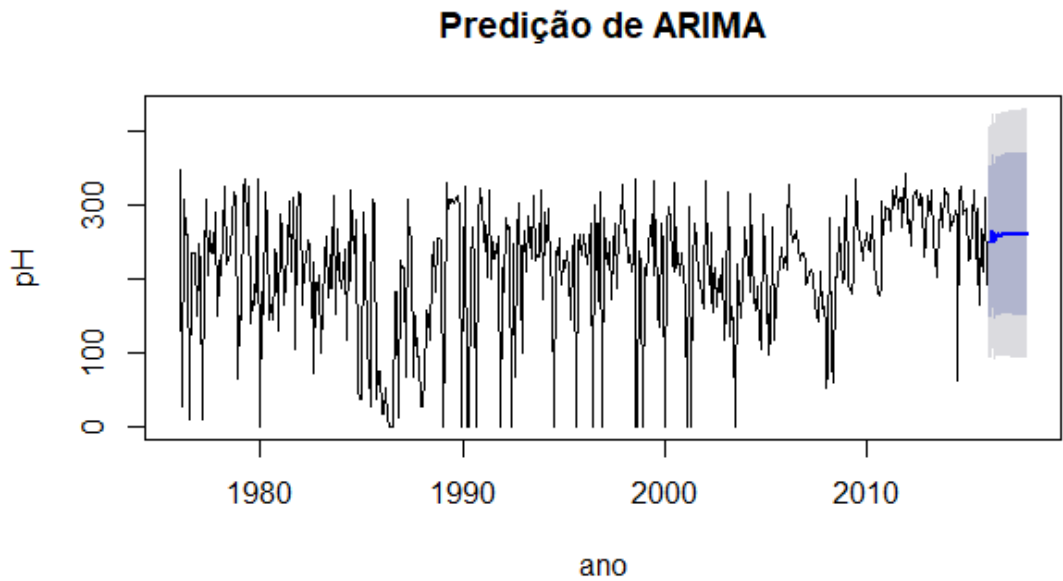


Figura 5: Predição da série pH para os próximos 24 meses.

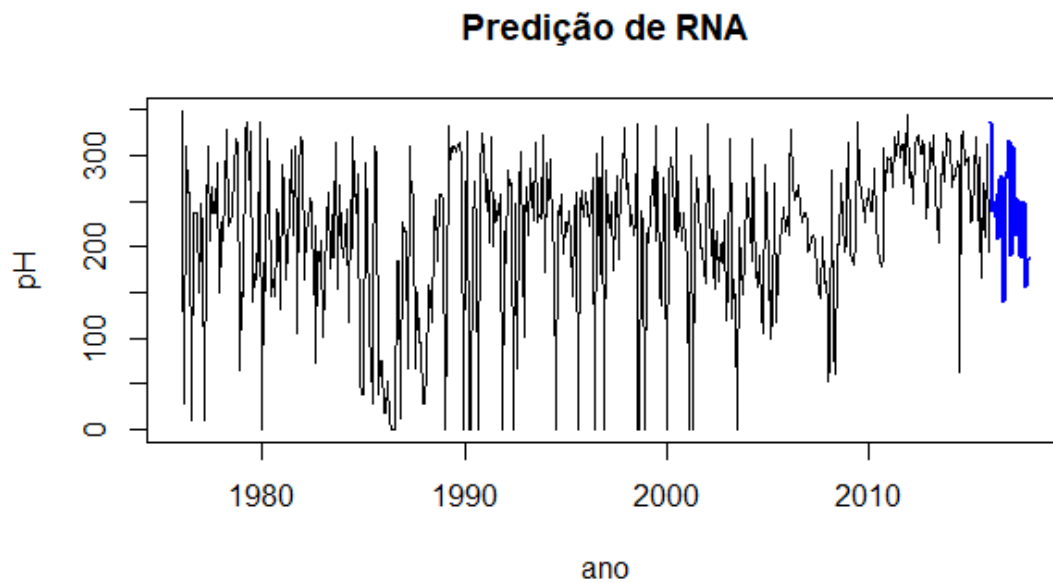


Figura 6: Predição da série pH para os próximos 24 meses.

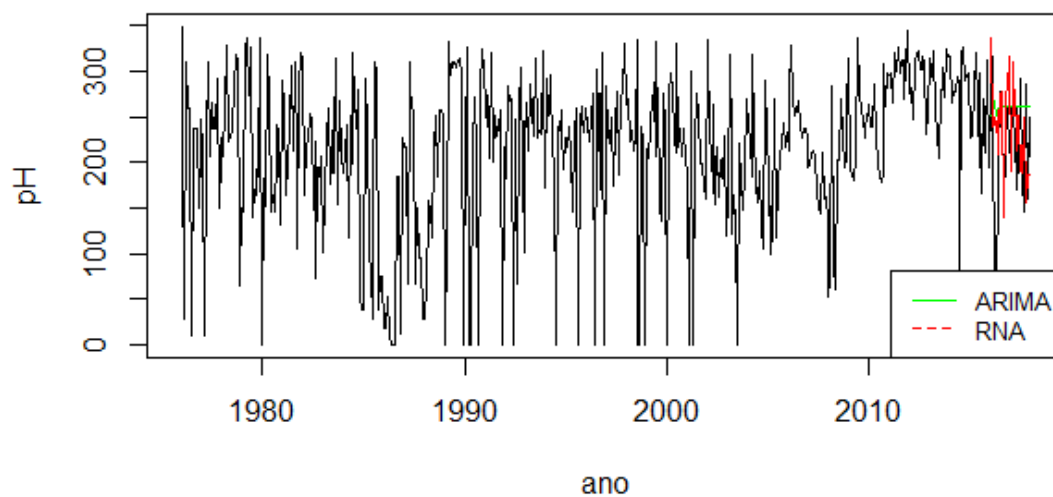


Figura 7: Predição autoregressiva na série temporal dos modelos ARIMA e RNA.

4. Conclusão

Neste artigo, avaliou-se o desempenho dos modelos ARIMA e RNA quando aplicados na autoregressão de uma série temporal do parâmetro pH da

qualidade da água da região hidrográfica do Tocantins-Araguaia. De acordo com os resultados apresentados, o modelo de RNA apresentou as melhores taxas de erros do que o modelo ARIMA. Quanto a predição tanto o modelo RNA quanto o ARIMA, apresentaram razoabilidade nas suas predições com valores preditos próximos da realidade. Como parte de um estudo posterior, sugere-se realizar a predição do valor do parâmetro pH para um intervalo de tempo futuro.

5. REFERÊNCIAS

- AGÊNCIA NACIONAL DE ÁGUAS (BRASIL). **Conjuntura dos recursos hídricos no Brasil: regiões hidrográficas brasileiras**. Brasília: ANA, 2014.
- AGÊNCIA NACIONAL DE ÁGUAS (BRASIL). **Conjuntura dos recursos hídricos no Brasil 2018: relatório pleno / Agência Nacional de Águas**. Brasília: ANA, 2018.
- AGÊNCIA NACIONAL DE ÁGUAS (BRASIL). **Manual de usos consultivos da água no Brasil / Agência Nacional de Águas**. Brasília: ANA, 2019.
- AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA (Brasil). **Banco de Informações de Geração (BIG)**. Disponível em: <
<http://www2.aneel.gov.br/aplicacoes/capacidadebrasil/capacidadebrasil.cfm>>.
Acesso em: jun. 2019.
- BAIRD, C; CANN, M. **Química ambiental**. 4 ed. Porto Alegre: Boolman, 2011.
- BENTO, Evaldo Junior. **Desenvolvimento web com PHP e MySQL**. São Paulo: Casa do Código, 2013.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. **Time Series Analysis, Forecasting and Control**. 4ed. Wiley. 2008.
- BRAGA, B.; PORTO, M.; TUCCI, C. E. M. Monitoramento de quantidade e qualidade das águas. In: BRAGA, B; TUNDISI, J. G.; TUNDISI, T. M.; CIMINELLI, V. S. T. (Org.). **Águas doce no Brasil: Capital ecológico, uso e conservação**. São Paulo: Escrituras, 2015. P. 127 – 142.
- BRANCO, S. M.; AZEVEDO, S. M. F. O.; HACHICH, E. M.; VAZOLLER, R. F.; TUNDISI, J. G. Água e saúde humana. In: BRAGA, B; TUNDISI, J. G.; TUNDISI, T. M.; CIMINELLI, V. S. T. (Org.). **Águas doce no Brasil: Capital ecológico, uso e conservação**. São Paulo: Escrituras, 2015. P. 319 – 358.
- BRASIL. Resolução CONAMA Nº 357, de 17 de março de 2005. Diário Oficial da República Federativa do Brasil, Poder Executivo, Brasília, DF, 18 MAR. 2005. P. 58 – 63.
- C. A. Villacorta Cardoso and G. Lima Cruz, Forecasting Natural Gas Consumption using ARIMA Models and Artificial Neural Networks, in **IEEE Latin America Transactions**, vol. 14, no. 5, pp. 2233-2238, May 2016. doi: 10.1109/TLA.2016.7530418.
- CARDOSO. C. A. V.; CRUZ. G. L.; Forecasting natural gas consumption using ARIMA models and artificial neural networks.
- DALL’OGLIO, P. **PHP Programando com orientações a objetos**. 4. ed. São Paulo: Novatec, 2018.
- EVSUKOFF, A.; EBECKEN, N. F. F. Mineração de dados de recursos hídricos. R. C. V. da Silva (editor). **Métodos numéricos em recursos hídricos**, vol. 6. ABRH/COPPETEC, 2003.

- FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From data mining to knowledge discovery: An overview. In: **Advances in Knowledge Discovery and Data Mining**, AAAI Press/The MIT Press, England, 1996, p.1-34.
- FOX, J. WEISBERG, S. **An R Companion to Applied Regression**. 3. ed. California: Sage, 2019.
- GUIMARÃES, A. M. **Aplicação de computação evolucionária na Mineração de dados físico-químicos da água e do solo**. São Paulo, 2005.
- HAYKIN, S. **Redes neurais: princípios e pratica**. [S.l.]: Bookman Editora, 2007.
- HENNING, E. et al. **Para além da computação estatística: o uso do ambiente R para o ensino de métodos numéricos**. RENOTE, v. 14, n. 1, 2016.
- HUAN, J.; CAO, W.; QIN, Y. Prediction of dissolved oxygen in aquaculture based on EEMD and LSSVM optimized by the Bayesian evidence framework. **Computers and electronics in agriculture**, v. 150, p. 257-265, 2018.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo Demográfico 2010: resultados do universo**. Disponível em: <
<https://sidra.ibge.gov.br/pesquisa/censo-demografico/demografico-2010/inicial> >.
 Acesso em 20 jul. 2019.
- KATAOKA, V. Y. et al. O uso do r no ensino de probabilidade na educação básica: Animation e teachingdemos (the use of r in probability teaching at basic education: Animation and teaching demos). **Proc. 18o Simpósio Nacional de Probabilidade e Estatística**, São Paulo, 2008.
- KONRATH, A. C.; SILVA, S. A.; HENNING, E.; SANTOS, L. M.; MIRANDA, R. G.; SAMOHYL, R. W. **Desenvolvimento de Aplicativos Web Com R e Shiny: inovações no ensino de Estatística**. Abakos, Belo Horizonte, v. 6, n. 2, p. 55-71, maio 2018 - ISSN: 2316-9451.
- LIU, S.; XU, L.; LI, D.; LI, Q.; JIANG, Y.; TAI, H.; ZENG, L. Prediction of dissolved oxygen content in river crab culture based on least squares support vector regression optimized by improved particle swarm optimization. **Computers and electronics in agriculture**, v. 95, p. 82-91, 2013.
- MINUTO, A.; PITTARELLO, F.; NIJHOLT, A. Smart material interfaces for education. **Journal of Visual Languages & Computing**, Elsevier, v. 31, p. 267–274, 2015.
- MONARD, M. C.; BARANAUSKAS, J. A. Conceitos Sobre Aprendizado de Máquina. In: RESENDE, S. O. (Coord.). **Sistemas Inteligentes Fundamentos e Aplicações**. 1 ed. Barueri-SP: Manole, 2003. p. 89--114. ISBN 85-204-168.
- PARMEZAN, A. R. S. **Predição de séries temporais por similaridade**. São Carlos: USP. 2016.
- PINTO, W. P.; LIMA, G. B.; ZANETTI, J. B. **Estudo de predição e previsão de temperatura médias diárias da cidade de Cariacica, Espírito Santo, Brasil, utilizando a metodologia de séries temporais**. Revista GEINTEC – ISSN: 2237-0722. São Cristóvão, 2015. Vol. 5. p. 1881 – 1895.

- RAUBER, T. W. **Redes neurais artificiais**. *Universidade Federal do Espírito Santo*, 2005.
- REBOUÇAS, A. C. A água doce no mundo e no Brasil. In: BRAGA, B; TUNDISI, J. G.; TUNDISI, T. M.; CIMINELLI, V. S. T. (Org.). **Águas doce no Brasil: Capital ecológico, uso e conservação**. São Paulo: Escrituras, 2015. P. 1 – 34.
- RUSSEL, S.; NORVIG, P. **Inteligência Artificial**. 3a Edição. Editora Campus, 2013.
- SCHREIBER, J. N. C.; BESKOW, A. L.; MÜLLER, J. C. T.; NARA, E. O. B.; SILVA, J. I.; REUTER, J. W. Técnicas de validação de dados para sistemas inteligentes: uma abordagem do software sdbayes. **XVII Colóquio Internacional de Gestão universitária. Mar del Plata, 2017**.
- SOLANKI, Archana; AGRAWAL, Himanshu; KHARE, Kanchan. Predictive Analysis of Water Quality Parameters using Deep Learning. **International Journal of Computer Applications**, v. 125, n. 9, p. 0975-8887, 2015.
- SILVA, G. A.; KULAY, L. A. Água na indústria. In: BRAGA, B; TUNDISI, J. G.; TUNDISI, T. M.; CIMINELLI, V. S. T. (Org.). **Águas doce no Brasil: Capital ecológico, uso e conservação**. São Paulo: Escrituras, 2015. P. 319 – 358.
- TELLES, D. D.; DOMINGUES, A. F. Água na agricultura e pecuária. In: BRAGA, B; TUNDISI, J. G.; TUNDISI, T. M.; CIMINELLI, V. S. T. (Org.). **Águas doce no Brasil: Capital ecológico, uso e conservação**. São Paulo: Escrituras, 2015. P. 319 – 358.