

Análise de desempenho de modelos preditivos para determinação do comportamento padrão de séries pluviométricas

O presente estudo trata da verificação de diferentes modelos de análise preditiva a fim de explorar e entender o comportamento padrão de séries pluviométricas, utilizando como área de estudo o reservatório Engenheiro Ávidos. Primeiramente levantou-se os dados pluviométricos e foi realizado o preenchimento das falhas, juntamente com a análise de consistências da série de estudo. Logo após, foi aplicado o método de Thiessen para determinar a precipitação correspondente ao reservatório. Foi obtida uma série histórica equivalente a um período de 57 anos (1963-2019). A segunda etapa do estudo prosseguiu com a análise estatística dos dados históricos, realizada com o auxílio do software R Studio, e pôr fim a análise preditiva utilizando os modelos Auto-Regressive Integrated Moving Average Model (ARIMA), Suavização Exponencial (SE), Redes Neurais (RN) e Regressão Linear Múltipla (RLM). Com os resultados obtidos, observou-se que todos os modelos apresentaram necessidade de ajustes estatísticos, mas que não os tornaram inviáveis para utilização. A precipitação apresentou uma melhor aderência aos dados observados quando comparadas com a série teste em todos os modelos, sendo o RLM o mais adequado. Em poucas vezes as previsões de precipitação superaram os valores observados, sendo uma característica mais presente para os modelos ARIMA e RN. Dessa forma, a previsão de precipitação obtida é potencialmente útil para estudos que envolvem engenharia hídrica, como para modelos de simulação e otimização.

Palavras-chave: Ajustes estatísticos; Previsões de precipitação; Regressão linear; Séries pluviométricas.

Performance analysis of predictive models for determining standard behavior of pluviometric series

The present study deals with the verification of different predictive analysis models to explore and understand the standard behaviour of rainfall series, using the Engenheiro Ávidos reservoir as the study area. First, the rainfall data was collected, and the gaps were filled in, together with the analysis of the study series consistencies. Soon after, the Thiessen method was applied to determine the precipitation corresponding to the reservoir. A historical series equivalent to a period of 57 years (1963-2019) was obtained. The second stage of the study proceeded with the historical data statistical analysis, performed with the R Studio software, and finally, the predictive analysis using the Auto-Regressive Integrated Moving Average Model (ARIMA), Exponential Smoothing (SE), Networks Neural (RN) and Multiple Linear Regression (RLM). With the results obtained, it was observed that all models showed the need for statistical adjustments, but that did not make them unfeasible for use. The precipitation showed better adherence to the observed data when compared with the test series in all models, with the RLM being the most appropriate. Precipitation forecasts rarely exceeded the observed values, being a more present characteristic for the ARIMA and RN models. Thus, the precipitation forecast obtained is potentially useful for studies involving water engineering, as well as for simulation and optimization models.

Keywords: Statistical adjustments; Precipitation forecasts; Linear regression; Rainfall series.

Topic: **Engenharia de Recursos Hídricos**

Received: **03/04/2021**

Approved: **26/04/2021**

Reviewed anonymously in the process of blind peer.

Cícero Fellipe Diniz de Santana 

Universidade Federal de Campina Grande, Brasil
<http://lattes.cnpq.br/7916729062952119>
<http://orcid.org/0000-0002-9228-3903>
cicero.esa@hotmail.com

Aline Pessoa Bezerra 

Universidade Federal de Campina Grande, Brasil
<http://lattes.cnpq.br/2846009756893544>
<http://orcid.org/0000-0002-9012-1265>
alinebezerra2890@gmail.com

Virgínia da Costa Brito 

Universidade Federal de Campina Grande, Brasil
<http://lattes.cnpq.br/3293234957426500>
<http://orcid.org/0000-0002-5776-0612>
virginia.dcb@gmail.com

Wilson Fadlo Curi 

Universidade Federal de Campina Grande, Brasil
<http://lattes.cnpq.br/4941338019932170>
<http://orcid.org/0000-0002-7024-7127>
wfcuri@gmail.com

Valterlin da Silva Santos 

Universidade Federal de Campina Grande, Brasil
<http://lattes.cnpq.br/4740303724015282>
<http://orcid.org/0000-0002-4789-2937>
valterlin@yahoo.com.br



DOI: 10.6008/CBPC2179-6858.2021.004.0020

Referencing this:

SANTANA, C. F. D.; BEZERRA, A. P.; BRITO, V. C.; CURTI, W. F.; SANTOS, V. S.. Análise de desempenho de modelos preditivos para determinação do comportamento padrão de séries pluviométricas. **Revista Ibero Americana de Ciências Ambientais**, v.12, n.4, p.229-244, 2021. DOI: <http://doi.org/10.6008/CBPC2179-6858.2021.004.0020>

INTRODUÇÃO

O aumento da demanda hídrica ao longo dos anos, vem aumentando, conseqüentemente os desafios relacionando a gestão efetiva das águas, de maneira a diminuir as conseqüências em eventos hidrológicos extremos e assegurar a disponibilidade desse recurso para as futuras gerações, sendo esses objetivos da Política Nacional de Recursos Hídricos (BRASIL, 1997).

O ciclo hidrológico é um processo natural complexo com diversas variáveis e incertezas (LESPINAS et al., 2014). Assim como em estudos já publicados, a modelagem hídrica é uma ferramenta importante para auxiliar no monitoramento das águas superficiais (KLING et al., 2014), operação de reservatórios (PIANTADOSI et al., 2010), na gestão hídrica (DUTTA et al., 2013), sendo indispensável para a gestão adaptativa (FERNÁNDEZ-OLALLA et al., 2012), e em outros estudos hidrológicos (NOHARA et al., 2017). O manejo de reservatórios de usos múltiplos depende do balanço hídrico da região, daí a importância da criação de modelos de previsão hidrológicas baseados em dados de precipitação e vazão, considerando sua frequência e probabilidade ao longo do tempo, séries mensais ou anuais (MOKHTARI et al., 2016), permitem caracterização da bacia hidrográfica, e a análise seus comportamentos e tendências (ANA, 2016).

Há duas maneiras de obter previsões hidrológicas, utilizar modelos determinísticos, desenvolvidos a partir de equações empíricas, ou utilizar modelos estocásticos, que consideram processos probabilísticos (TUCCI, 1998). Previsões de precipitação já realizadas, à exemplo dos trabalhos de Bargagli et al. (2002), Wu et al. (2011) e Xu et al. (2020), mostram que quando se trata de previsão, umas das ferramentas efetivas no estudo dessas incertezas das séries de variáveis hidrológicas e parâmetros associados são os sistemas de simulação estocásticos.

A utilização desses métodos pode auxiliar na otimização e gestão das águas de reservatórios, proporcionando resultados de simulação baseados em estatística a curto e longo prazos, e a escolha do método mais adequado depende das características do problema, da disponibilidade de dados e do resultado pretendido (SUN et al., 2018).

O presente estudo trata particularmente sobre a análise de diferentes modelos de análise preditiva, usando aplicações já difundidas cientificamente, a fim de explorar e entender o comportamento padrão de séries pluviométricas, em especial do nordeste brasileiro devido a necessidade de eficiência no manejo da disponibilidade de água para as populações devido as condições climáticas da região. A premissa básica fundamenta-se na possível utilização dos modelos matemáticos como forma de contribuição para as estratégias e tomadas de decisão envolvendo reservatórios de acumulação para atendimento às demandas.

METODOLOGIA

O reservatório Engenheiro Ávidos, escolhido como área de estudo, localizado na bacia hidrográfica do Rio Piancó-Piranhas-Açu (Figura 1), no estado da Paraíba. Sua construção foi concluída em 1936, possuindo atualmente uma capacidade máxima de armazenamento de 293 milhões de m³, tendo como principal função o abastecimento urbano do município de Cajazeiras e distritos circundantes no interior da

Paraíba (ANA, 2018), correspondendo ao atendimento de aproximadamente 63.000 habitantes.

A bacia está localizada no sertão, onde possui clima semiárido, com precipitações médias anuais de 869,5mm para o reservatório (AES A, 2020). A pluviometria local possui períodos de chuvas concentradas em poucos meses do ano, com padrão sazonal interanual consecutivo, iniciando o período médio de altas precipitações a partir do mês de dezembro, com ápice do índice pluviométrico no mês de fevereiro, transitando para o período de estiagem durante o mês de maio. O período de estiagem segue de junho e vai até novembro.

A pluviosidade ao longo dos anos apresenta uma variabilidade distribuída de forma que os anos com pluviosidade acima da média são recorrentemente seguidos por ciclos de seca, o que faz com a disponibilidade hídrica para as demandas seja afetada (NEPOMUCENO, 2018; NETO, 2018).

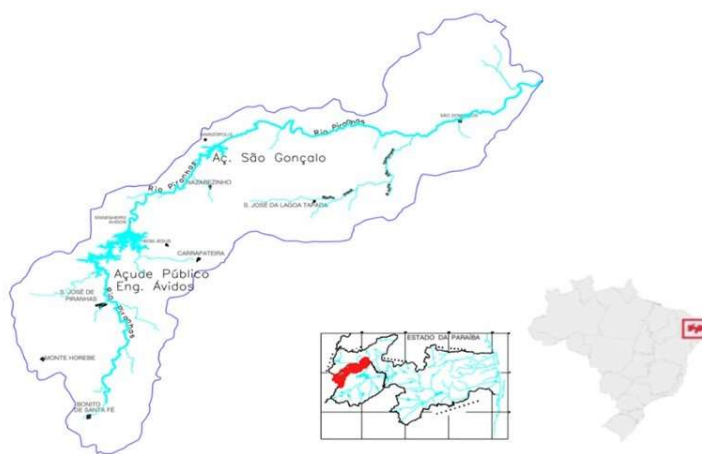


Figura 1: Localização dos Reservatórios Engenheiro Ávidos e São Gonçalo na Bacia Hidrográfica do Rio Piranhas - Fonte: Adaptado de SCIENTEC (1997).

Levantamento e tratamento dos dados

A primeira etapa da análise elaborada no estudo, foi iniciada com o levantamento dos dados da pluviometria incidente no reservatório, através do software *Microsoft Office Excel* para agrupamento dos dados obtidos nos postos que representam a alimentação do afluente. Inicialmente foram utilizados dados pluviométricos de postos em torno da área de estudo, esses dados foram disponibilizados pela AESA, e no Plano diretor da bacia do rio Piranhas (SCIENTEC, 1997), sendo estes, os postos: Bonito de Santa Fé-PB; São José de Piranhas; Serra Grande-PB; além dos postos nas barragens do Engenheiro Ávidos.

Após a coleta dos dados, foi realizado o preenchimento de falhas das séries, com base no método de ponderação regional integrado a regressão linear, conforme Oliveira et al. (2010), que consiste no estabelecimento de regressões lineares entre os postos com dados a serem preenchidos e cada um dos postos vizinhos.

Realizado o preenchimento de falhas, foram feitas as análises de consistências para os postos utilizando o método da dupla massa, consistindo na acumulação dos valores mensais do posto e na visualização do comportamento em um gráfico, que forma uma reta, obtida através da comparação com a

média dos postos vizinhos ou de um posto pluviométrico com dados confiáveis. No caso da ocorrência de desvios na reta, os valores não contínuos e consistentes são ajustados por meio da equação da reta ajustada, e novamente aplicado o método da dupla massa (CARVALHO et al., 2016).

Após o ajuste dos dados foi aplicado o método de Thiessen (THIESSEN, 1911), para determinar a precipitação correspondente ao reservatório. Conforme afirma Marciano et al. (2017), o polígono de Thiessen é formado pela interseção das linhas das mediatrizes, correspondendo à área de influência de cada estação, por fim calculando o valor da precipitação do ponto desejado.

Os postos pluviométricos e seus respectivos coeficientes de Thiessen utilizados foram: São José de Piranhas - 0,51; Bonito de Santa Fé - 0,23; Engenheiro Ávidos - 0,16; e Serra Grande - 0,10. Calculados os valores médios das precipitações, com base nas séries de dados previamente preenchidos e ajustados, foi obtida a série temporal histórica, em escala mensal, totalizando 684 meses, equivalente a um período de 57 anos, do ano de 1963 até 2019, conforme ilustrado na figura 2.

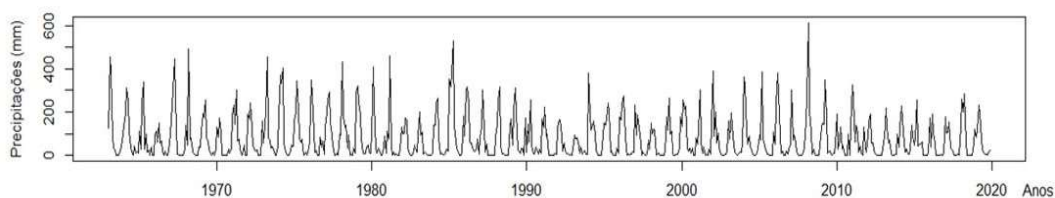


Figura 2: Dados mensais de precipitação na área de estudo de 1963 a 2019.

Análises estatísticas iniciais para as séries temporais

Toda a análise estatística, aplicação dos modelos e avaliação de desempenho dos resultados foi realizada com o auxílio do software *R Studio*, utilizando dos pacotes *forecast*, *tseries*, *seasonal*, entre outros pacotes auxiliares. Foram então realizadas as análises de estatística descritiva e tratamento dos dados *outliers*. Para análise do comportamento ao longo dos 57 anos, foi observada a estacionariedade e a distribuição da série. Foram aplicados os testes de *Ljung-box*, *Kwiatkowski-Phillips-Schmidt-Shin* (KPSS) e *Dickey-Fuller*.

O primeiro teste aplicado foi o *Ljung-Box* (LJUNG et al., 1978), em que utiliza estatística Q para verificar a autocorrelação entre os dados de acordo com o intervalo de confiança por meio da autocorrelação residual, expresso pela equação 1. O segundo teste aplicado foi o de KPSS (KWIATKOWSKI et al., 1992), teste estacionário baseado em regressão linear (equação 2). O terceiro teste aplicado foi *Dickey-Fuller* (DICKEY et al., 1979), que é baseado em uma função que depende do tempo adotado (equação 3). A aplicação do teste retorna um coeficiente que pode ser negativo ou positivo de acordo com a estacionariedade da série, sendo que quanto mais negativo for o resultado, maiores as chances de a série ser estacionária.

Os resultados obtidos para ambos os métodos *Ljung-Box* e KPSS, dentro do nível de confiança de 95%, devem retornar o p-valor maior que 0,05, o que indica que a série é estacionária, ou seja, não há indícios de que os dados sejam auto correlacionados, um valor que deve ser encontrado nos resíduos, sugerindo que não haja dados explanatórios faltantes na série temporal. O resumo de todas as equações aplicadas encontra-

se na Tabela 1.

Tabela 1: Resumo das equações dos testes aplicados a serie temporal.

Teste	Equação
<i>Ljung-Box</i>	$LB = N(N + 2) \sum_{k=1}^m \left(\frac{r_k^2}{N - k} \right) \quad (1)$
KPSS	$KPSS = \left(\frac{1}{N^2} \sum_{t=1}^N S_t^2 \right) / \lambda^2 \quad (2)$
<i>Dickey-Fuller</i>	$\Delta Y_t = \alpha + (\theta - 1)Y_{t-1} + \gamma T + u_t \quad (3)$

Aplicação dos modelos preditivos para previsão hidrológica

Devido as amplas aplicações já realizadas no cotidiano e desempenho, para a análise preditiva das séries temporais, foram escolhidos os modelos *Auto-Regressive Integrated Moving Average Model* (ARIMA), Suavização Exponencial (SE), Redes Neurais (RN) e Regressão Linear Múltipla (RLM).

O primeiro modelo aplicado foi o ARIMA (BOX et al., 1976), que é uma das mais importantes técnicas de análise de séries temporais. O modelo realiza a autocorrelação entre os valores da série em intervalos de tempo consecutivos associado a uma probabilidade de ocorrência dos eventos observados. Neste estudo, o ARIMA foi aplicado, apesar da série não possuir estacionariedade confirmada, o que é favorável ao modelo. Sua aplicação seguiu as etapas do ciclo de inferência: Identificação do modelo, estimativa dos parâmetros, previsão e análise dos resíduos. A aplicação do modelo foi realizada conforme a equação 4, sendo representado por 3 parâmetros não sazonais (p, d e q), e 3 parâmetros sazonais (P, D, Q), correspondendo ao ARIMA Sazonal. Onde p, d e q são parâmetros, sendo: p define a ordem da parte autorregressiva, com o número de dados anteriores interferentes (*lags*); d o grau de diferenciação e q a ordem da parte sazonal;

$$\phi(B) * \Phi(B^S) * (1 - B^S)^D * (1 - B)^d * Y_t = \theta(B) * \Theta(B^S) * a_t \quad (4)$$

Na expressão, B é o operador translação; ϕ é o coeficiente auto regressivo não sazonal; Φ é o coeficiente auto regressivo sazonal; θ é o coeficiente de média móvel não sazonal; Θ é o coeficiente média móvel sazonal. Para a obtenção do modelo foi utilizada a recomendação de Hyndman et al. (2018), usando comando *auto.arima* removendo os argumentos *stepwise* e *approximation*, que limitam e aproximam os ajustes da raiz unitária estacionária no ambiente R.

O segundo modelo de previsão aplicado foi o RLM, envolve aplicação de matrizes de acordo com o número de amostras, correlacionando coeficientes de regressão com variáveis independentes e dependentes, sendo obtido um gráfico (reta ou curva) dos dados correlacionados. No caso das séries temporais, os conjuntos de dados relacionados, são os *lags* para cada valor da série, conforme a equação 5:

$$\bar{Y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon \quad (5)$$

Sendo, $x_1, x_2 \dots x_k$ dados de entrada, ou seja, variáveis independentes; $\beta_0, \beta_1, \beta_2 \dots \beta_k$ coeficientes de regressão; e ϵ componente resíduo aleatório. O modelo RLM da série foi aplicado o comando *tslm*, usado para ajuste automático das regressões de séries temporais. A busca realizada no ambiente R, incluiu componentes de tendência e sazonalidade, obtendo como resultados os coeficientes de interseção (α), os

coeficientes sazonais (β) e o coeficiente de tendência (γ), sendo cada variável correspondente à um conjunto de dados oriundos da própria série.

A SE foi o terceiro modelo aplicado, especificamente o método de *Holt-Winters* (HOLT, 1957; WINTERS, 1960), que associa parâmetros de nível, tendência e sazonalidade em dados não estacionários, envolvendo médias aritméticas ponderadas. O modelo foi aplicado através do comando *ets* (*Exponential smoothing state space model*), baseado na busca automática do ajuste aos dados da série, identificando os parâmetros do modelo, previsão e intervalos de confiança. Adotou-se um modelo com tendência nula e sazonalidade aditiva, dado pelas equações 6, 7 e 8.

$$n_t = \alpha(Y_t - s_{t-m}) + (1 - \alpha)n_{t-1} \quad (6)$$

$$s_t = \gamma(Y_t - n_t) + (1 - \gamma)s_{t-m} \quad (7)$$

$$\hat{Y}_{t+k} = n_t + s_{t-m+k} \quad (8)$$

Por fim, foi aplicado o modelo de RN. No caso das séries temporais, os nós ou neurônios de entrada no modelo são respectivamente os tempos de autorregressão da série que recebem pesos, baseados nos ajustes dos *lags*, em que após feita a aplicação da função de saída da rede, a função de ativação estima o valor previsto usando um fator de tendência, porém, diferente dos modelos anteriores, a saída do modelo não possui intervalos de confiança. Os principais parâmetros a serem introduzidos no *R Studio* são os intervalos de entrada e o número de camadas intermediárias ocultas.

Para a aplicação foi utilizado o comando *nnetar*, que corresponde a um tipo de rede neural autorregressiva, que utiliza conjuntos de valores anteriores para treinamento da rede e ajuste dos pesos, sendo um modelo de camada única oculta e com entradas regressivas. Não foram adicionados limitantes ou outros recursos específicos do comando, sendo calculado automaticamente o parâmetro “*p*”, que representa o número de *lags* não sazonais.

Após a aplicação dos modelos, foi feita a checagem de resíduos para cada um deles, a diferença entre o valor observado e o valor ajustado ao verificar o comportamento da série, os diagramas das funções de autocorrelação (ACF) e autocorrelação parcial (PACF).

Análise de desempenho dos modelos testados

Antes da aplicação dos modelos para a previsão hidrológica, foram definidos os períodos de dados históricos, usando um procedimento de *Hold-Out*, ou seja, escolhendo os dados a serem utilizados para o ajuste dos modelos, nomeado de “período treino”, juntamente com outro período complementar, utilizado para comparação com os resultados obtidos pelo modelo para o mesmo intervalo de tempo, nomeado de “período teste”. O período treino escolhido foram os anos entre janeiro de 1963 e dezembro de 2009, enquanto o período teste foram os anos entre janeiro de 2010 e dezembro de 2019. A escolha destes períodos foi feita a partir da visualização da decomposição da série temporal e observação da tendência dos dados (figura 3), por meio da média móvel anual, onde foi possível identificar perda de sazonalidade dos ciclos de secas e cheias nos últimos dez anos.

Ao final dos procedimentos de modelagem, foram finalmente analisados quais os melhores resultados e modelos que descreveram melhor o padrão hidrológico da série de estudo, através das principais métricas utilizadas na análise de dados. Os indicadores de desempenho aplicados foram: *Standardized Mean Error* - ME, *Mean Absolute Error* - MAE, *Root Mean Squared Error* - RMSE, *Mean Absolute Squared Error* - MASE.

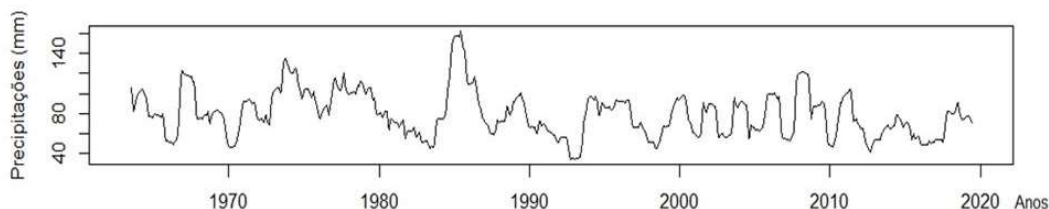


Figura 3: Tendência de dados das séries temporais.

Para todos os indicadores apresentados, os valores de \hat{y}_i são os dados observados para o período de estudo, enquanto y_i são os dados obtidos a partir da aplicação dos modelos. Em resumo, a previsão é avaliada através da diferença entre o previsto e o ocorrido.

O ME (Equação 9), calcula a média entre as diferenças dos dados previstos e observados. Devido ao ME ser susceptível a anulação entre os diferentes dados, podendo ser encontradas diferenças elevadas apesar do baixo valor do ME. O MAE (Equação 10), consegue resolver esse problema, pois avalia a média da diferença absoluta. O RMSE (Equação 11), calcula por sua vez o desempenho dos desvios da série, a partir do cálculo do desvio padrão da amostra da diferença entre os dados de treino e teste. Por fim, o MASE (Equação 12) compara o desempenho das séries temporais com diferentes escalas. Nas métricas citadas, os valores que representam o melhor desempenho estão o mais próximo possível do zero.

Tabela 3: Resumo das equações dos modelos de desempenho aplicadas a série temporal.

Modelo de desempenho	Equação
ME	$\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)$ (9)
MAE	$\frac{1}{n} \sum_{i=1}^n \hat{y}_i - y_i $ (10)
RMSE	$\sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$ (11)
MASE	$\frac{\sum_{i=1}^n y_n - \hat{y}_i / n}{\sum_{i=1}^n y_n - y_{n-1} / (n-1)}$ (12)

RESULTADOS

Como primeiros resultados para a análise da série pluviométrica foram obtidos os dados estatísticos (tabela 4) para Engenheiro Ávidos, entre os principais valores de média, desvio padrão, coeficiente de variação e o percentual de zeros.

Tabela 4: Dados estatísticos descritivos das séries pluviométricas.

Média	Desvio Padrão	Mediana	1º Quartil	3º Quartil	Coef. de variação	% Zeros
79,65	53,84	64,69	43,19	105,35	1,16	0,64

A figura 4 apresenta os gráficos de boxplot, histograma e distribuição de quartis da série

pluviométrica. O gráfico *boxplot* ilustrado apresenta a variabilidade dos dados históricos, demonstrando um comportamento macroscópico característico de padrão anual para as precipitações. Com relação ao histograma de distribuições, os mesmos são distribuídos a partir de um conjunto assimétrico, correspondendo à padrões gama e log-normal. Tal fato é constatado pela não linearidade das distribuições de quartis, sendo diferente da normal, característica de séries que não apresentam estacionariedade, aspecto esse recomendado para a análise preditiva, principalmente para modelos como ARIMA e SE.

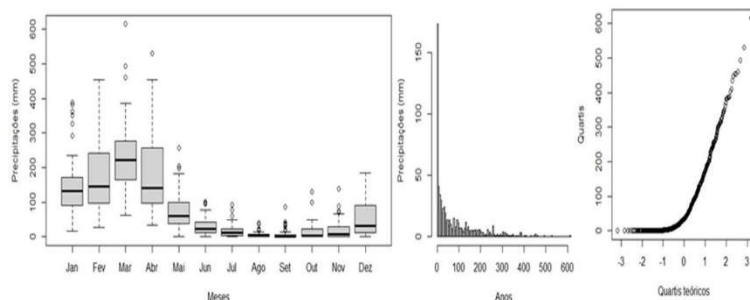


Figura 4: Gráficos de boxplot para os reservatórios Engenheiro Ávidos.

Os resultados encontrados nos testes LB, KPSS e DF (tabela 5), onde os testes realizados par a LB, alegam alta correlação, com interferência de variáveis explanatórias indefinidas. Já os KPSS e DF não são conclusivos quanto a estacionariedade, contudo, devido ao tipo de distribuição, é possível estimar valores de DF menores, caso fossem realizados ajustes matemático-estatísticos, como diferenciação e logaritmização.

Tabela 5: Resultados estatísticos da série temporal.

Testes	Parâmetro	Engenheiro Ávidos
LB	p-valor	$<2,2 \cdot 10^{-16}$
KPSS	Estatística LM	0,2779
DF	Estatística T	-9.7418

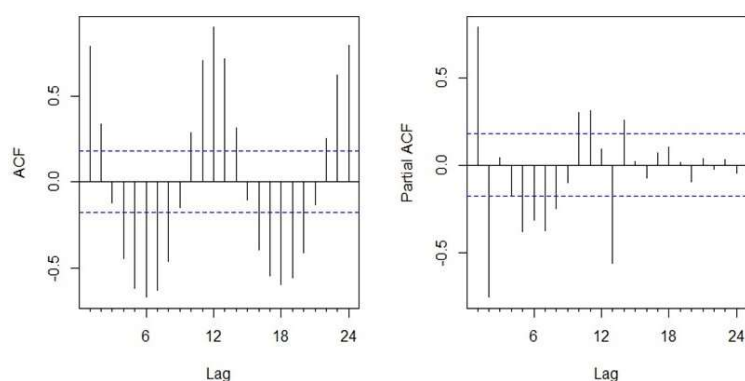


Figura 5: Gráficos da ACF e PACF para a série histórica.

Para os gráficos ACF e PACF (Figura 5), é possível notar que o ACF possui valores elevados nos primeiros *lags*, com uma queda rápida nos demais *lags*, porém retornando periodicamente aos valores mais altos, com forte sazonalidade. Essa condição pode ser confirmada pelo PACF, com autocorrelação presente no *lag* de ordem 1, com queda vertiginosa para os *lags* seguintes, com oscilação periódica ao longo dos *lags* de forma dissipada.

Os padrões sazonais com fraca tendência para as séries temporais, necessitam de análise mais complexa, devido à grande gama de padrões, o que indica a provável ineficácia de alguns modelos (MOHAMUDALLY, 2018). Optou-se por manter os aspectos naturais dos dados, com o intuito de reproduzir previsões com aspectos e distribuições semelhantes à série histórica.

ARIMA

Junto com a avaliação da ACF e PACF da série histórica, foi utilizado um modelo ARIMA(1,0,1)(2,0,1). A tabela 6, onde apresenta os coeficientes do modelo, o erro padrão, bem como a significância com relação aos resíduos. De acordo com o modelo, o AR1 indicou que os dados referentes ao mês imediatamente anterior possuem a maior significância, enquanto SAR2, apesar da pequena semelhança dos resultados, possuiu maior significância e forte sazonalidade.

Tabela 6: Coeficientes do ARIMA(1,0,1)(2,0,1) estimado para a série pluviométrica.

Parâmetro	Coefficiente	Erro Padrão	Teste t	p-valor
AR1	0,6496	0,12	-0,091065	0,9275
MA1	-0,5367	0,1343	-0,091719	0,927
SAR1	-0,6728	0,0411	-0,091794	0,9269
SAR2	-0,2918	0,0422	-0,091584	0,9271

A distribuição dos resíduos seguiu a mesma organização da série histórica, porém com valores simétricos, enquanto a ACF manteve 88,89% das defasagens não correlacionadas, estando os resíduos estabilizados não correlacionados (Figura 6). Para os ACF e PACF das médias dos intervalos de confiança da previsão, os resultados obtiveram alta congruência.

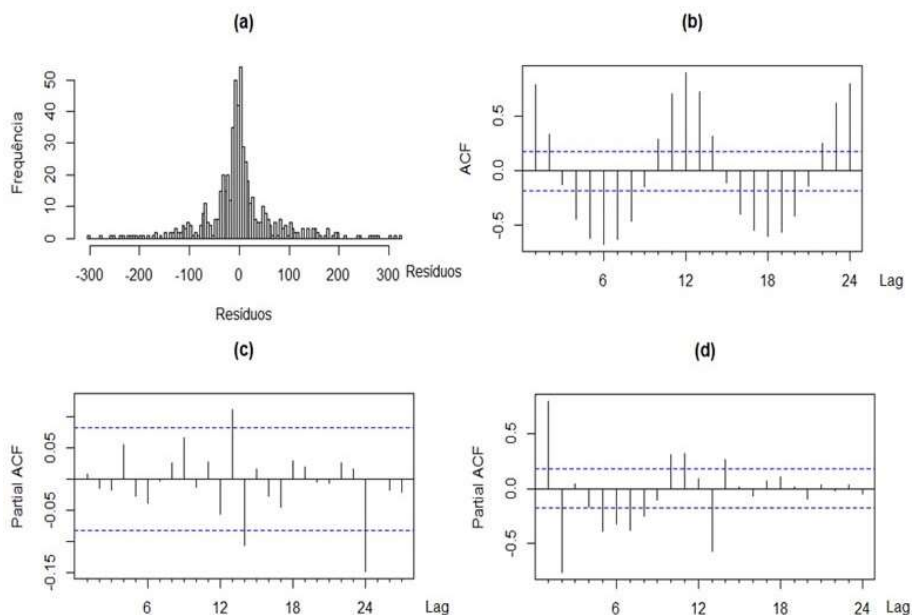


Figura 6: Histograma e ACF de resíduos do ARIMA.

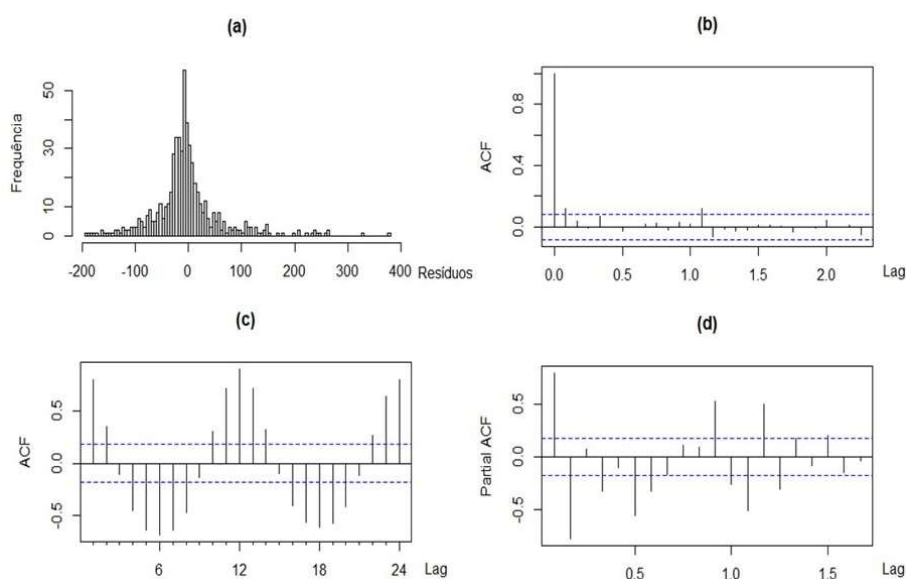
Regressão

A tabela 8, apresenta os coeficientes do modelo, o erro padrão, bem como a significância com relação aos resíduos.

Tabela 8: Coeficientes do modelo RLM para a série temporal.

Parâmetros		Coeficientes	Erro padrão	Teste T	p-valor
α_1		153,11206	11,06788	13,834	$<2,00 \cdot 10^{-16}$
	γ_1	-0,02562	0,01762	-1,454	0,14644
β_2		42,13438	14,04874	6,723	$2,83 \cdot 10^{-03}$
β_3		94,45045	14,04877	3,447	$4,46 \cdot 10^{-11}$
β_4		48,42913	14,04883	-4,817	$6,10 \cdot 10^{-04}$
β_5		-67,66937	14,04891	-8,181	$1,89 \cdot 10^{-06}$
β_6		-114,92968	14,04901	-9,129	$1,96 \cdot 10^{-15}$
β_7		-128,25790	14,04913	-9,940	$<2 \cdot 10^{-16}$
β_8		-139,65201	14,04927	-9,738	$<2 \cdot 10^{-16}$
β_9		-136,81235	14,04944	-9,493	$<2 \cdot 10^{-16}$
β_{10}		-133,37804	14,04962	-9,124	$<2 \cdot 10^{-16}$
β_{11}		-128,18898	14,0983	-6,813	$<2 \cdot 10^{-16}$
β_{12}		-95,72543	14,05007	2,999	$2,51 \cdot 10^{-11}$

Foi possível verificar que em todos os α e β , o p-valor do teste T foi inferior ao nível de 5% de significância, indicando que estes parâmetros são estatisticamente significativos ao modelo ajustado, exceto para γ , indicando a ausência de tendência estabelecida para as precipitações. O coeficiente de determinação ajustado foi de 0,5988, indicando que 59,88% da variabilidade das precipitações é aplicada pelo modelo ajustado. A estatística F também foi obtida, com o p-valor de $<2,0 \cdot 10^{-16}$, menor do que 5% de significância, indicando existência de variáveis explicativas estatisticamente significativas no modelo ajustado. Nota-se pelo ACF (Figura 7-b) que os resíduos aparentam estar distribuídos aleatoriamente em torno do zero e sua função de autocorrelação apresenta dois valores significativos, com exceção das defasagens em 0,1 e 1,1.

**Figura 7:** Distribuição (a) e ACF (b) dos resíduos, ACF (c) e PACF (d) da previsão.

Analisando os resíduos (Figura 7-a), para o teste LB, o p-valor foi de 0,002571, sendo menor que o nível de significância de 0,05%, não tratando-se do comportamento de ruído branco. No teste DF para os resíduos, obteve-se valor p de 0,01 que é menor que o nível de significância 0,05, então rejeita-se a hipótese nula de não estacionariedade, concluindo que a série temporal é estacionária. A análise da ACF (Figura 7-c) das médias dos intervalos de confiança da previsão, os resultados obtiveram comportamento semelhante ao da série histórica, representando um mesmo grau de autocorrelação. O comportamento da função PACF (Figura 7-d), também está de encontro com os dados históricos que apresentam movimento oscilatório

dissipado, com significância estatística em 45% dos *lags*.

Suavização Exponencial

O modelo SE escolhido foi o ETS(A,N,A), sendo o primeiro termo relacionado ao ajustes aditivo dos erros (A), o segundo como sem tendência (N), e terceiro relacionado a sazonalidade aditiva (A). Assim, dentre os parâmetros do modelo, foi encontrado o alfa (α) de 0,0132, que corresponde ao nível da suavização e gama (γ) de 0,0004, correspondente à sazonalidade, além dos componentes sazonais relativos aos meses do ano (Tabela 7).

Tabela 7: Componentes sazonais mensais.

Mês	Componente sazonal	Mês	Componente sazonal
Jan.	-32,0114	Jul.	-51,1386
Fev.	-63,2326	Ago.	-4,9525
Mar.	-69,3716	Set.	115,9775
Abr.	-73,0715	Out.	162,188
Mai.	-76,3431	Nov.	106,4627
Jun.	-63,5776	Dez.	49,0708

A figura 8 apresenta os gráficos referentes ao modelo SE. Assim como no modelo anterior, os resíduos não são normalmente distribuídos (Figura 8-a), contudo não apresentaram autocorrelação no PACF (Figura 8-b), com exceção dos *lags* em 0,1 e 1,1, com aderência de 91,67% não correlacionados. Para os ACF das médias dos intervalos de confiança da previsão (Figura 8-c), o gráfico possui o mesmo comportamento da amostra, entretanto o PACF não possui rápida dissipação (Figura 8-d), o que indica uma autocorrelação os meses mais recentes.

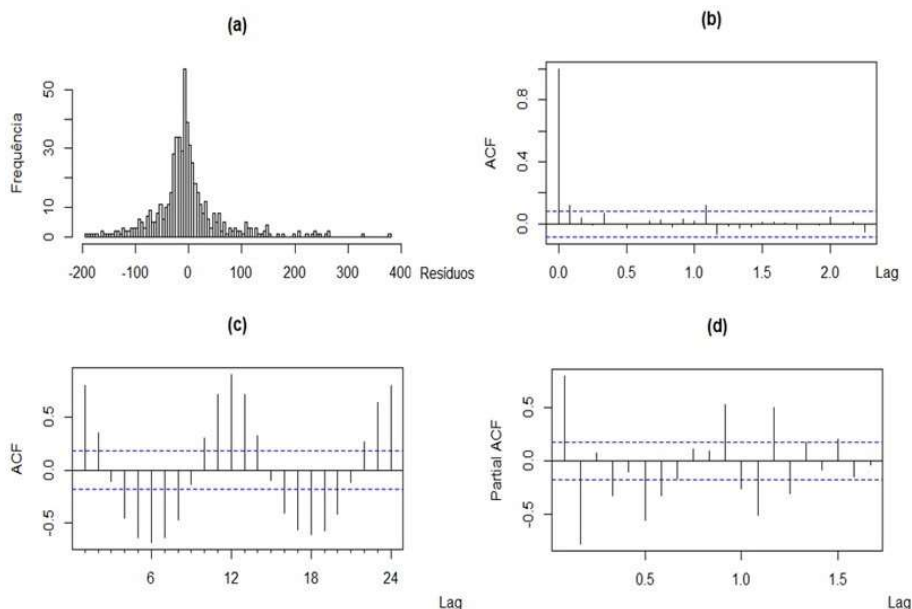


Figura 8: Histograma e ACF de resíduos do ARIMA.

Redes Neurais

Para o modelo RNA, o valor de p não foi previamente estabelecido, sendo o valor ideal encontrado automaticamente de acordo com o valor do AIC. O outro parâmetro encontrado foi o valor de " P ", que

representa o número de atrasos sazonais, sendo o valor padrão $P=1$ e valor escolhido $P=12$, representando a sazonalidade interanual.

A verificação da autocorrelação dos resíduos para o modelo (Figura 9-b) não apresentaram significâncias consideráveis ao nível de 0,05% de significância, sendo o único modelo que parece apresentar um comportamento de ruído branco. A realização do teste LB para os resíduos, apresentou o p-valor de 0,2513 que é maior que o nível de 0,05% de significância, levando a não rejeição da hipótese nula, também indicando o comportamento de ruído branco. Em relação ao teste KPSS, o valor p para os resíduos foi de 0,1.

Vê-se pelo ACF (Figura 9-b) que os resíduos aparentam estar distribuídos aleatoriamente em torno do zero, confirmando que os resíduos apresentam valores dentro dos limites de significância correspondente ao comportamento de ruído branco. A análise da ACF (Figura 9-c) das médias dos intervalos de confiança da previsão, também obtiveram comportamento semelhante ao da série histórica, assim como nos modelos anteriores. O comportamento da função PACF (Figura 9-d), também seguiram conforme os dados históricos, mantendo oscilação temporária para os primeiros *lags*.

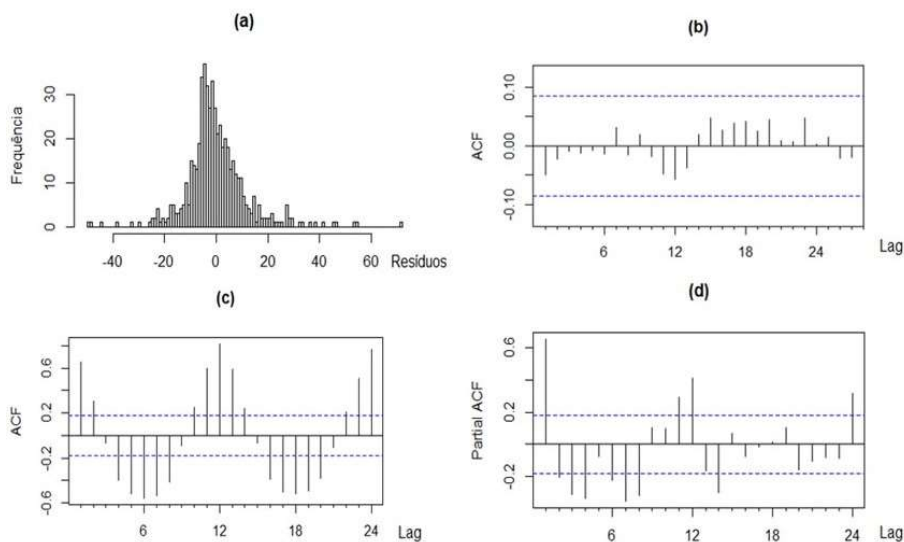


Figura 9: Distribuição (a) e ACF (b) dos resíduos, ACF (c) e PACF (d) da previsão.

Métricas de Desempenho

A análise dos indicadores de desempenho levou em consideração os dados dos valores de cada modelo, sendo o modelo com maior número de indicadores mais correspondentes, com maior número de aproximações, sendo realizada a análise dos melhores modelos em comparação com as séries treino e teste.

Tabela 9: Valores das métricas de desempenho dos modelos estudados para a série treino.

Modelos	ME	RMSE	MAE	MASE
ARIMA	-0,2939	76,2792	47,0474	0,8205
SE	-2,5363	67,7861	43,2870	0,7549
RLM	-1,598*10 ⁻¹⁵	67,3143	42,2203	0,7363
RNA	-0,0977	11,9356	08,0041	0,1396

Os resultados obtidos (tabela 9) para os indicadores que possui os valores mais próximos de zero são calculados para o RNA, com exceção do valor do ME, sendo este muito menor para o modelo RLM, contudo,

os modelos ARIMA e SE, obtiveram os valores de ME pelo menos 66,749% maiores que o modelo RNA. A média dos valores para o RMSE, entre ARIMA, SE e RLM de 70,4589, enquanto para RNA foi 11,93565, distando em 83,06%. Os valores do MAE e MASE para todos os modelos também possuíram diferenças consideráveis entre o RNA e os demais modelos, correspondendo em 81,88% para ambos os indicadores.

A tabela 10, apresenta os valores dos indicadores, obtidos para análise das aproximações feitas com a série teste.

Tabela 10: Valores das métricas de desempenho dos modelos estudados para a série teste.

Modelos	ME	RMSE	MAE	MASE
ARIMA	-29,0181	79,4894	53,9087	0,9402
SE	-15,8142	58,9168	39,6384	0,6913
RLM	-07,8976	56,6620	36,5447	0,6373
RNA	-12,6405	65,1643	43,3044	0,7552

Na comparação, os modelos apresentam valores mais elevados em todas as métricas, entretanto os valores do RNA obtiveram a maior discrepância enquanto os demais modelos sofreram aumento proporcional nas métricas, com exceção do MASE, que apresentou diminuição dos números, indicando um melhor ajuste dos resíduos em relação a previsão. Sendo assim, o modelo RLM obteve o melhor desempenho, considerando que os números desses modelos estão mais próximos de zero. A mudança dos valores com relação aos dados treino e teste no RNA, estão diretamente relacionadas com os ciclos existentes e distintos entre as duas séries, devido a maior presença de sazonalidade na série treino, enquanto a série teste possui um maior grau de estacionariedade, o que explica uma melhor desenvoltura do RLM para o conjunto teste. O valor de ME permaneceu como o mais baixo nas duas análises, significando uma maior neutralidade dos resíduos e a melhor adequação do RLM quanto a regularidade do ciclo anual de precipitações, com variação média das métricas em 41,54%, enquanto o RNA obteve a variação média de 79,31%.

Apesar do modelo RLM apresentar os melhores resultados para o período teste, os intervalos de confiança e as médias das previsões do RLM possuem em seus conjuntos numéricos valores negativos em alguns meses de escassez, que foram tratados e considerados como zeros, por se tratarem de precipitações. Da previsão extraiu-se valores estatísticos descritivos que poderão ser interpretados como dados padrão da série, estes por sua vez tão representativos quanto a estatística descritiva histórica já apresentada anteriormente, em detrimento da distribuição e da sazonalidade histórica. Os dados padrão da série temporal prevista são apresentados na tabela 11.

Tabela 11: Dados estatísticos da média das previsões para o modelo RLM.

Média	Desvio Padrão	Mediana	1º Quartil	3º Quartil	Coef. de variação	% Zeros
73,812	82,4769	31,721	6,825	148,472	0,895	0,59

Outras grandes diferenças entre os modelos podem ser visualizadas através dos gráficos das séries temporais previstas, sobrepostas sob a série histórica teste (Figura 10). É possível observar que as médias de previsão dos modelos ARIMA (Figura 10-a) e RNA (Figura 10-d), possuem uma maior variabilidade do que os

modelos SE (Figura 10-b) e RLM (Figura 10-c), o que pode ser considerado algo positivo, porém é pouco confiável que os dados históricos correspondam as amplitudes apresentadas dentro do ciclo sazonal de estudo, o que faz com que um comportamento padrão possa ser melhor definido pela regularidade e sincronia dos eventos de cheia e estiagem. Sendo assim, podemos confirmar que visualmente as SE e RLM, são os modelos com melhor definição característica, sendo dentre estes o mais aproximado da previsão padrão da série o modelo RLM.

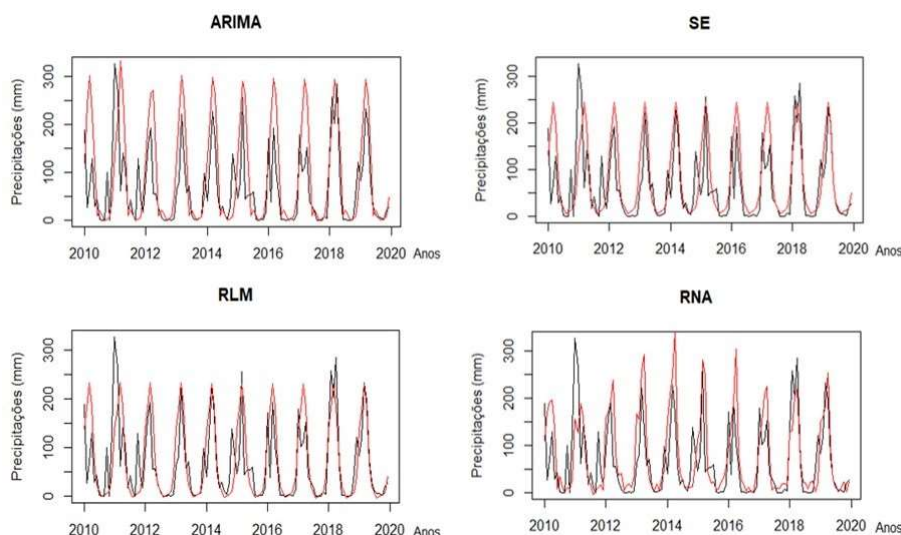


Figura 10: Gráficos das previsões (vermelho) sobre a série histórica teste (preto).

CONCLUSÕES

Foi possível identificar a forte presença de sazonalidade e fraca tendência entre os dados históricos e as séries modeladas. Contudo, todos os modelos, demonstraram necessidade de ajustes estatísticos, devido a presença de valores negativos, que foram convertidos em zero, mas que não tornou inviável a utilização dos modelos, se comparados os percentuais de zeros entre os dados observados e os modelados.

A precipitação prevista tende a apresentar uma melhor aderência aos dados observados quando comparadas com a série teste em todos os modelos, sendo os RLM e SE mais adequados, com destaque para o RLM. Para precipitações mensais mais intensas (superiores a 250 mm) o modelo RLM apresenta uma tendência em não subestimar as precipitações observadas. Nos casos estudados, em poucas vezes as previsões de precipitação superaram os valores observados, sendo uma característica mais presente para os modelos ARIMA e RN.

É importante ressaltar que, mesmo com imprecisões nas previsões de precipitação obtidas com a utilização dos modelos, pode-se concluir que essa previsão de precipitação é potencialmente útil como subsídio para o processo de previsão de precipitações mensais quando observados os limites de confiança probabilísticos em torno das médias de previsão.

Uma das particularidades interessantes sobre a análise, foi a possibilidade de aplicação deste estudo no contexto real e prático meteorológico, pois os valores das médias das previsões, juntamente com os intervalos de confiança para os dados de cada mês, servem de contribuição na identificação dos

comportamentos padrões de séries pluviométricas através de uma abordagem simples e robusta, proporcionando o melhor entendimento da influência das chuvas durante as tomadas de decisão em sistemas de recursos hídricos.

REFERÊNCIAS

AESA. Agência Executiva de Gestão de Águas. **Meteorologia, Chuvas**. AESA, 2020.

ANA. Agência Nacional de Águas. **Portal da Qualidade das Água**. ANA, 2016.

ANA. Agência Nacional de Águas. **Nota Técnica nº 11/2018/COMAR/SRE**. Marco Regulatório estabelecendo condições de uso dos recursos hídricos no sistema hídrico Eng. Ávidos, São Gonçalo e rio Piranhas, no Estado da Paraíba. Brasília: ANA, 2018.

BARGAGLI, A.; CARILLO, A.; PISACANE, G.; RUTTI, P. M.; STRUGLIA, V.; TARTAGLIONE, N.. An Integrated Forecast System over the Mediterranean Basin: Extreme Surge Prediction in the Northern Adriatic Sea. **Monthly Weather Review**, v.130, n.5, p.1317-1333, 2002. DOI: [http://doi.org/10.1175/1520-0493\(2002\)130<1317:AIFSOT>2.0.CO;2](http://doi.org/10.1175/1520-0493(2002)130<1317:AIFSOT>2.0.CO;2)

BOX, G. E. P.; JENKINS, G. M.. **Time Series Analysis: Forecasting and Control**. San Francisco: Holden-Day, 1976.

BRASIL. **Lei nº 9.433, de 8 de janeiro de 1997**. Institui a Política Nacional de Recursos Hídricos, cria o Sistema Nacional de Gerenciamento de Recursos Hídricos, regulamenta o inciso XIX do art. 21 da Constituição Federal, e altera o art. 1º da Lei nº 8.001, de 13 de março de 1990, que modificou a Lei nº 7.990, de 28 de dezembro de 1989. Brasília: DOU, 1997.

CARVALHO, H. P.; RUIZ, M. V. S.. Avaliação da Consistência de Séries Históricas de Chuva da Bacia Hidrográfica do Rio Araguari, em Minas Gerais. **Fórum Ambiental da Alta Paulista**, v.12, n.6, p.76-84, 2016. DOI: <http://dx.doi.org/10.17271/1980082712620161480>

DICKEY, D. A.; FULLER, W. A.. Distribution of the estimators for autoregressive time series with a unit root. **Journal of the American Statistical Association**, v.74, p.427-431, 1979. DOI: <https://doi.org/10.1080/01621459.1979.10482531>

DUTTA, D.; WILSON, K.; WELSH, W. D.; NICHOLLS, D.; KIM, S.; CETIN, L.. A new river system modelling tool for sustainable operational management of water resources. **Journal of Environmental Management**, v.121, p.13-28, 2013. DOI: <https://doi.org/10.1016/j.jenvman.2013.02.028>

FERNÁNDEZ-OLALLA, M.; MARTÍNEZ-ABRAÍN, A.; CANUT, J.; GARCÍA-FERRÉ, D.; AFONSO, I.; GONZÁLEZ, L. M.. Assessing different management scenarios to reverse the declining trend of a relict capercaillie population: A modelling approach within an adaptive management framework. **Biological Conservation**, v.148, p.79-87, 2012. DOI: <https://doi.org/10.1016/j.biocon.2012.01.047>

HOLT, C. C.. Forecasting seasonal and trends by

exponentially weighted moving averages. **Office of Naval Research, Research Memorandum**, v.52, 1957.

HYNDMAN, R. J.; ATHANASOPOULOS, G.. **Forecasting: principles and practice**. 2 ed. Otexts, 2018.

KLING, H.; STANZEL, P.; PREISHUBER, M.. Impact modelling of water resources development and climate scenarios on Zambezi River discharge. **Journal of Hydrology: Regional Studies**, v.1, p.17-43, 2014. DOI: <https://doi.org/10.1016/j.ejrh.2014.05.002>

KWIATKOWSKI, D.; PHILLIPS, P. C. B.; SCHMIDT, P.; SHIN, Y.. Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. **Journal of Econometrics**, v.54, p.159-178, 1992.

LESPINAS, F.; LUDWIG, W.; HEUSSNER, S.. Hydrological and climatic uncertainties associated with modeling the impact of climate change on water resources of small Mediterranean coastal rivers. **Journal of Hydrology**, v.511, p.403-422, 2014. DOI: <https://doi.org/10.1016/j.jhydrol.2014.01.033>

LJUNG, G. M.; BOX, G. E. P.. On a Measure of Lack of Fit in Time Series Models. **Biometrika**, v.65, p.297-303, 1978. DOI: <https://doi.org/10.1093/biomet/65.2.297>

MARCIANO, A. G.; BARBOSA, A. A.; SILVA, A. P. M.. Cálculo de precipitação média utilizando método de Thiessen e as linhas de cumeada. **Ambiente & Água**, v.13, n.1, p.1-9, 2017. DOI: <http://dx.doi.org/10.4136/ambi-agua.1906>

MOHAMUDALLY, N.. **Time series analysis and applications**. IntechOpen, 2018.

MOKHTARI, E. H.; REMINI, B.; HAMOUDI, S. A.. Modelling of the rain-flow by hydrological modelling software system HEC-HMS - watershed's case of wadi Cheliff-Ghrib, Algeria. **Journal of Water and Land Development**, v.30, p.87-100, 2016.

NEPOMUCENO, T. C.. **Modelo de simulação de nitrogênio e fósforo em sistemas de recursos hídricos, aplicação em região do alto curso do Rio Piranhas/PB**. Dissertação (Mestrado em Engenharia Civil e Ambiental) - Universidade Federal de Campina Grande, Campina Grande, 2018.

NETO, E. D. S.. **Modelo de cobrança pelo uso dos recursos hídricos em bacias hidrográficas controladas por reservatórios**: proposta metodológica e simulação na bacia hidrográfica dos Rios Piancó-Piranhas-Açu. Tese (Doutorado em Recursos Naturais) - Universidade Federal de Campina Grande, Campina Grande, 2018.

NOHARA, D.; HORI, T.. Integrated Reservoir Operation Considering Real-Time Hydrological Prediction for Adaptive

Water Resources Management. In: KOLOKYTHA, E.; OISHI, S.; TEEGAVARAPU, R. S. V.. **Sustainable Water Resources Planning and Management Under Climate Change**. Springer, 2017. p.101-132. DOI: https://doi.org/10.1007/978-981-10-2051-3_5

OLIVEIRA, L. F. C.; FIOREZE, A. P.; MEDEIROS, A. M. M.; SILVA, M. A. S.. Comparação de metodologias de preenchimento de falhas de séries históricas de precipitação pluvial anual. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.14, n.11, p.1186-1192, 2010. DOI: <http://dx.doi.org/10.1590/S1415-43662010001100008>

PIANTADOSI, J.; HOWLETT, P. G.; BEAN, N. G.; BEECHAM, S.. Modelling systems of reservoirs using structured Markov chains. **Water Management**, v.163, p.407-416, 2010. DOI: <https://doi.org/10.1680/warna.900053>

SCIENTEC. Associação para Desenvolvimento da Ciência e Tecnologia. **Plano Diretor de Recursos Hídricos da Paraíba: Bacias do Rio Piancó e do Alto Piranhas - SEPLAN**. João Pessoa: SCIENTEC, 1997.

SUN, Y.; ZHU, F.; CHEN, J.; LI, J.. Risk Analysis for Reservoir Real-Time Optimal Operation Using the Scenario Tree-Based

Stochastic Optimization Method. **Water**, v.10, n.5, p.606, 2018. DOI: <https://doi.org/10.3390/w10050606>

THIESSEN, A. H.. Precipitation averages for large areas. **Monthly Weather Review**, v.39, n.7, p.1082-1089, 1911.

TUCCI, C. E. M.. **Modelos hidrológicos**. UFRGS, 1998.

WINTERS, P. R.. Forecasting sales by exponentially weighted moving averages. **Management Science**, v.6, p.324-342, 1960. DOI: <https://doi.org/10.1287/mnsc.6.3.324>

WU, L.; SEO, D.; DEMARGNE, J.; BROWN, J. D.; CONG, S.; SCHAAKE, J.. Generation of ensemble precipitation forecast from single-valued quantitative precipitation forecast for hydrologic ensemble prediction. **Journal of Hydrology**, v.399, p.281-298, 2011. DOI: <https://doi.org/10.1016/j.jhydrol.2011.01.013>

XU, B.; ZHONG, P.; LU, Q.; ZHU, F.; HUANG, X.; MA, Y.; FU, J.. Multiobjective stochastic programming with recourse for real-time flood water conservation of a multireservoir system under uncertain forecasts. **Journal of Hydrology**, v.590, p.1-15, 2020. DOI: <https://doi.org/10.1016/j.jhydrol.2020.125513>

A CBPC - Companhia Brasileira de Produção Científica (CNPJ: 11.221.422/0001-03) detém os direitos materiais desta publicação. Os direitos referem-se à publicação do trabalho em qualquer parte do mundo, incluindo os direitos às renovações, expansões e disseminações da contribuição, bem como outros direitos subsidiários. Todos os trabalhos publicados eletronicamente poderão posteriormente ser publicados em coletâneas impressas sob coordenação da **Sustenere Publishing**, da Companhia Brasileira de Produção Científica e seus parceiros autorizados. Os (as) autores (as) preservam os direitos autorais, mas não têm permissão para a publicação da contribuição em outro meio, impresso ou digital, em português ou em tradução.