



SOBRE ÁGUA NO SEMIÁRIDO BRASILEIRO

MODELAGEM PURAMENTE ESTOCÁSTICA DE PREVISÃO SAZONAL DE PRECIPITAÇÃO PARA O SEMIÁRIDO BRASILEIRO

Darlan Martines Schmidt ⁽¹⁾; Rosane Rodrigues Chaves ⁽¹⁾; Paulo Sérgio Lucio ⁽¹⁾

⁽¹⁾Universidade Federal do Rio Grande do Norte – Programa de Pós-Graduação em Ciências Climáticas - darlanmartiness@gmail.com

RESUMO

A região Nordeste Brasileira apresenta uma grande variabilidade espacial, sazonal e interanual na distribuição das chuvas, desde volumosos índices precipitados na região costeira como baixos volumes nas regiões interioranas, além da má distribuição espacial, existe ainda o fator de que as chuvas ocorrem em poucos meses do ano, principalmente na região semiárida, na qual o período chuvoso se resume a apenas três a quatro meses do ano, que climatologicamente se estende de fevereiro a maio. Essa variabilidade na região está associada as variações de padrões de ventos locais e Temperatura da Superfície do Mar (TSM) sobre os oceanos tropicais, que afetam o posicionamento e intensidade da Zona de Convergência Intertropical (ZCIT), Linhas de Instabilidade (LI), distúrbios Ondulatórios de Leste (OL) e sistemas convectivos locais. A região Nordeste, frequentemente é atingida por eventos extremos, tanto secas periódicas ou eventos ocasionais de inundações. Apresentando índices pluviométricos que variam em média de 1600 mm a menos de 500 mm anuais, no sentido do litoral ao interior. A possibilidade de previsibilidade do comportamento futuro de variáveis meteorológicas é de fundamental importância, principalmente se tratando da previsão de precipitação, variável de suma importância para o planejamento e gestão dos recursos hídricos em uma região com grande variabilidade e incertezas anuais. O objetivo do estudo foi construir um modelo numérico de previsão de precipitação, puramente estocástico, baseado na metodologia Box e Jenkins, para o Município de Assú – RN, localizado na região do Semiárido Nordeste Brasileiro. A modelagem estocástica é recomendada quando se estuda variáveis climáticas pois preservam a memória histórica da série e acoplam as principais características do comportamento da mesma. Para a escolha do modelo apropriado foram feitas análises gráficas e testes estatísticos pertinentes a metodologia adotada. A previsão deve ser verificada com dados observados na mesma localidade para um mesmo período, isso serve para avaliar a capacidade preditiva do modelo construído, além de permitir a análise dos erros referentes a previsão efetuada. Foi utilizada a série de precipitação mensal no período de 1911 a 2014. O modelo selecionado foi um Auto Regressivo Integrado de Média Móvel Sazonal (SARIMA $(2,1,1) \times (1,1,2)_{12}$), tendo como critério de escolha testes de significância e análise de resíduos, além dos valores do índice AIC. Foram realizadas previsões de temperatura para o ano de 2015 e comparadas com os valores reais observados de janeiro até junho de 2015, permitindo concluir que o modelo SARIMA escolhido resultou em uma alternativa viável de previsão sazonal de precipitação para a região semiárida Brasileira.

Palavras-chave: Semiárido Brasileiro. Box e Jenkins. Previsão Sazonal de Chuvas.

INTRODUÇÃO

O registro e manipulação de dados meteorológicos são essenciais ao entendimento do comportamento do clima. A marcante variabilidade sazonal e interanual da pluviometria sobre a





SOBRE ÁGUA NO SEMIÁRIDO BRASILEIRO

região Nordeste do Brasil, mais precisamente sobre a região Semiárida, é um dos principais fatores para a ocorrência de eventos extremos tanto secas periódicas ou eventos ocasionais de inundações. Essa variabilidade na região está associada as variações de padrões de ventos locais e Temperatura da Superfície do Mar (TSM) sobre os oceanos tropicais, que afetam o posicionamento e intensidade da Zona de Convergência Intertropical (ZCIT), Linhas de Instabilidade (LI), distúrbios Ondulatórios de Leste (OL) e sistemas convectivos locais. A possibilidade de previsibilidade do comportamento futuro de variáveis meteorológicas é de fundamental importância, principalmente se tratando da previsão de precipitação, variável de suma importância para o planejamento e gestão dos recursos hídricos em uma região com grande variabilidade e incertezas anuais. O objetivo do estudo foi construir um modelo numérico de previsão de precipitação, puramente estocástico, a partir da metodologia Box e Jenkins. Segundo SCHMIDT (2014) a precipitação média anual da região é de 600 mm, com grande variabilidade espacial e temporal. Pois preservam a memória histórica da série e acoplam as principais características do comportamento da mesma. Para a escolha do modelo apropriado foram feitas análises gráficas e testes estatísticos pertinentes a metodologia adotada. Após verificada a adequabilidade do modelo, é possível gerar a previsão dos índices de precipitação para um determinado período. A previsão deve ser verificada com dados observados na mesma localidade para um mesmo período, isso serve para avaliar a capacidade preditiva do modelo construído, além de permitir a análise dos erros referentes a previsão efetuada. Neste trabalho utilizamos um modelo SARIMA (BOX et al., 1994), modelo sazonal autorregressivo integrado à médias móveis. Com os ajustes do modelo SARIMA foi possível realizar previsões de padrões de comportamento futuro dessa variável.

MATERIAL E MÉTODOS

A área de estudo compreende à bacia do rio Piranhas-Assú, à jusante do reservatório Armando Ribeiro Gonçalves. Região de clima semiárido e relevo pouco acidentado. A economia da região é baseada em atividades do setor primário como pesca, seguindo-se as atividades referentes ao setor de serviços e ao setor secundário como exploração petrolífera, ambas sensíveis a falta ou excesso de água. O conjunto de dados utilizados neste estudo provém da série de dados pluviométricos observados da estação climatológica localizada no município de Assú – RN, obtidos da Agencia Nacional de Águas (ANA). Considerando uma série histórica de dados pluviométricos mensais de 1911 a 2014, total de 103 anos.

Um dos modelos mais utilizados que consideram sazonalidade de uma determinada série temporal, é o modelo ARIMA (LEE; KO, 2011), ARIMA sazonal ou SARIMA. Estes modelos são importantes pois levam em consideração a sazonalidade estocástica dos dados. Tal modelo foi proposto por Box e Jenkins na década de 1970 e tem origem nos modelos auto regressivo (AR), médias móveis (MA) e da combinação dos modelos AR e MA (ARMA).

Os modelos AR (p), MA (q) e ARMA (p,q) são utilizados quando a série é estacionária, ou seja, suas propriedades estatísticas básicas, como média, variância e covariância, permanecem constantes ao longo do tempo (HANKE et al, 2001).





SOBRE ÁGUA NO SEMIÁRIDO BRASILEIRO

Entretanto, quando a série é não estacionária, ela é transformada em uma série estacionária por meio da aplicação de diferenças na série de dados. Um modelo clássico para séries temporais supõe que a série contenha os seguintes componentes: tendência, ciclo e sazonalidade; e a construção de valores previstos para a série é realizado por meio da reunificação de cada um desses componentes (SOUZA et al, 2008).

Para lidar com as séries que apresentam autocorrelação sazonal, Box e Jenkins (1976) generalizaram o modelo ARIMA e definiram o modelo ARIMA sazonal, conhecido como SARIMA. Estes modelos são importantes pois levam em consideração a sazonalidade estocástica dos dados.


Quando o período for igual a 12 (doze) por exemplo, o modelo denominado SARIMA de ordem $(p,d,q) \times (P,D,Q)_{12}$, é dado por:

$$\varphi(X)\Phi(X^{12})\Delta^d\Delta_{12}^D Z_t = \theta(X)\Theta(X) \delta a_t \quad Eq:(1)$$

em que: $\varphi(X)$ é o operador autorregressivo (AR) de ordem p , $\theta(X)$ é o operador médias móveis (MA) de ordem q , $\Phi(X)$ é o operador AR-sazonal de ordem P , $\Theta(X)$ é o operador MA-sazonal de ordem Q . Δ é o operador diferença, Δ_{12}^D é o operador diferença sazonal e δa_t é o ruído branco ou resíduo.

A metodologia Box-Jenkins é definida por um ciclo composto de cinco etapas iterativas: Identificação do Modelo; Estimativa dos Parâmetros; Verificação; e Previsão, que compreende a aplicação e verificação do Modelo.

Um índice utilizado para a seleção dos modelos é o critério de informação de Akaike – Akaike Information Criterion (AIC), sendo uma função de verossimilhança do modelo ARIMA ajustado e do número de parâmetros do modelo (EHLERS, 2009). Ao comparar dois ou mais modelos, o escolhido será aquele que apresentar o menor valor de AIC, valores menores indicam modelos mais próximos à realidade, ou que têm menos perda de informação em relação à realidade. Além de observar se nas FAC's dos resíduos com lag = 1 são estatisticamente iguais a zero, o que indica que os resíduos são não correlacionados, e observando se na estatística de Ljung-Box dos modelos se os valores-p são altos, através do tsdiag.

Todo o tratamento estatístico dos dados foi feito com o software R (R DEVELOPMENT CORE TEAM, 2015),  versão 3.1.3.

RESULTADOS E DISCUSSÕES

A aplicação foi realizada de acordo com as etapas iterativas da metodologia Box e Jenkins. Para a identificação do modelo, inicialmente (Figura 1) é apresentado o gráfico BoxPlot mensal, que apresenta as principais estatísticas descritivas referentes a cada mês da série, mínimo, máximo, mediana, média, desvio padrão, quartis e alguns valores “outliers” de forma gráfica. A série





SOBRE ÁGUA NO SEMIÁRIDO BRASILEIRO

temporal utilizada na estimação do modelo é composta por 1236 observações (janeiro de 1911 a dezembro de 2014).

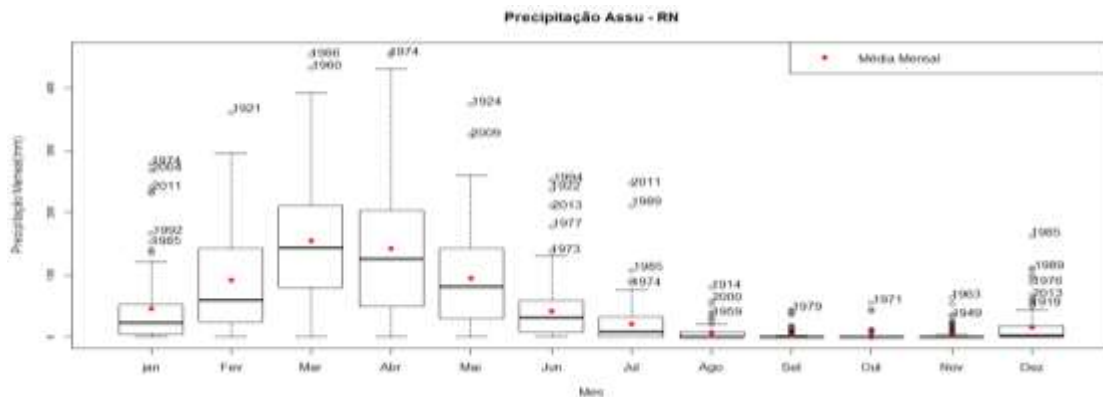


Figura 1 – Gráfico Box Plot da série original de precipitação para o município de Assú-RN.

Através do gráfico da decomposição da série original da série observada não é possível identificar facilmente se a série é ou não estacionária. Assim, partiu-se para a observação das características desmembradas, onde se pode observar claramente a existência de tendências, sazonalidade e ruídos que compõem a série original, denotando que a série é não estacionária.

Para confirmar a estacionariedade da série, optou-se por aplicar os testes estatísticos: Teste F que compara duas variâncias (var.test), no qual ao nível de significância de 5%, rejeita-se a hipótese nula de que a variância é homocedástica, pois o valor $p < 0.01$, logo a série é não estacionária. Assim como o Teste de Bartlett de homogeneidade de variâncias (bartlett.test), também para esse, a variância não estabilizou, portanto, para tentar estabilizar as variâncias, aplicou-se a transformação BoxCox, a qual permitiu pelos testes acima verificar que a série transformada se encontra com variâncias homocedástica, mas ainda existe sazonalidade e tendência.

Para induzir estacionariedade à série, aplicou-se uma diferença simples e uma diferença sazonal. Para confirmar a estacionariedade da série, aplicou-se os testes de raízes unitárias Dickey-Fuller (ADF) onde a hipótese nula de que a série é não estacionária foi rejeitada (p -valor = 0,01), mostrando estacionariedade da série com as diferenças e Phillips-Perron (PP), que assume as mesmas condições de hipóteses, o qual apresentou os mesmos valores, confirmando que na série foi induzida estacionariedade.

Por meio dos testes estatísticos e diagnóstico dos resíduos, entre os possíveis modelos, escolheu-se o modelo SARIMA (1,1,2)x(1,1,2)₁₂. Foi observado que esse modelo apresentou menor valor de AIC, e de acordo com a ACF, foi o único modelo em que a lag está próximo de 0 assim como os valores-p foram altos. E as estatísticas de Box-Ljung apresentaram valores altos. Além de apresentar todos os coeficientes significantes ao nível de 5%. A análise pelo Shapiro-Wilk indica que resíduos seguem uma distribuição normal (p -value = 4.554e-10).

De acordo com o modelo apropriado escolhido, os parâmetros da equação resultaram em:





SOBRE ÁGUA NO SEMIÁRIDO BRASILEIRO

Tabela 2: Estimativa de parâmetros do modelo – SARIMA(1,1,2)(1,1,2)12

Coefficiente	ϕ_{ar1}	Φ_{ma1}	Φ_{ma2}	θ_{sar1}	Θ_{sma1}	Θ_{sma2}
Estimativa	0.5489	-1.3510	0.3665	-0.1582	-0.8116	-0.1883
Erro Padrão δa_t	0.1272	0.1402	0.1351	0.6561	0.6451	0.6449

$AIC = 785.6.$

Com os parâmetros estimados, aplicando – os na Equação (1), resultou na Tabela 2, valores previstos para os primeiros 6 (seis) meses de 2015, além de expor os valores observados em 2015 e a média histórica climatológica de precipitação mensal para os mesmos meses em questão.

Tabela 2: Valores observados, previsões SARIMA (1,1,2)x(1,1,2)12 e Climatologia mensal.

Meses	Observado em 2015	Previsão 2015	Climatologia Mensal
Janeiro	2,8	25,38	45,41
Fevereiro	76,9	78,81	91,15
Março	154	194,15	154,74
Abril	70,65	163,17	141,14
Maiο	34,8	107,43	93,98
Junho	14,35	25,37	41,62

Para visualização gráfica elaborou se a Figura 3, no qual se observa os valores observados em 2015 (Janeiro a Junho), a climatologia mensal e os valores previstos para o ano de 2015 pela modelagem adotada. Verificando que o modelo construído se comportou de maneira satisfatória em relação aos valores médios mensais observados no período de 1911 a 2014, o modelo superestimou em torno de 20 mm os valores precipitados no período chuvoso (Março a Maio) para a região propriamente dito, e subestimando de certa forma em aproximadamente 20 mm, nos períodos de transição entre o seco - chuvoso e entre o chuvoso – seco. Porém, no período de estiagem (Agosto a Novembro), apresentou valores aproximados aos da média histórica. Comparando os valores observados com a previsão do modelo, de janeiro até junho de 2015, os valores modelados foram próximos aos observados nos três primeiros e no sexto mês. Nos meses de abril e maio de 2015 os valores observados foram bastante abaixo da climatologia média mensal e da previsão gerada pelo modelo.





SOBRE ÁGUA NO SEMIÁRIDO BRASILEIRO

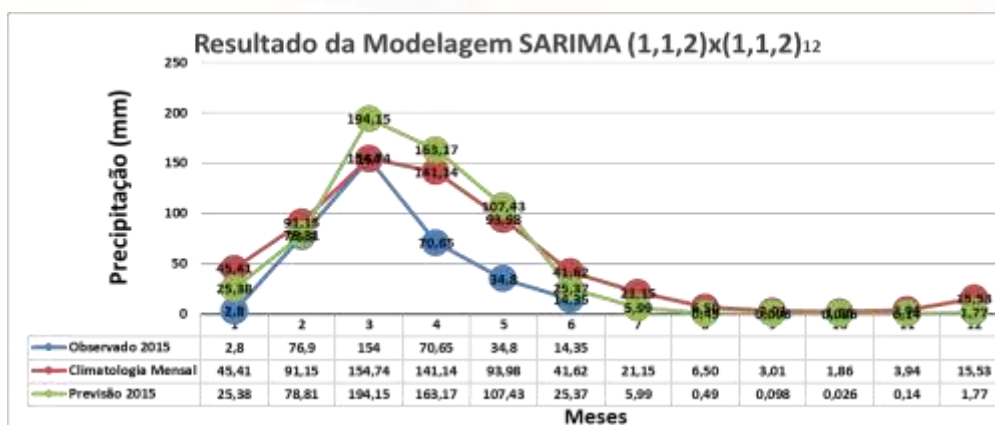


Figura 3 – Gráfico com os valores observados e previstos pelo modelo SARIMA(1,1,2)(1,1,2)₁₂.

Uma forma adicional de medir a capacidade preditiva de um modelo de previsão com as observações continuadas in loco, é comparar os valores previstos, com parte da série observada através de testes estatísticos, isso nos dá uma maior clareza dos erros decorrentes e a percepção do quanto foi boa qualitativamente minha previsão. Isso foi realizado pelo comando “verify” do próprio software “R”. Tal verificação chegou aos seguintes resultados: Erro médio absoluto = 0.2089; Erro médio = 0.1851 e Erro quadrático médio = 12.6. Logo, de acordo com os valores dos erros médios, a previsão de precipitação para o período aferido foi qualitativamente boa.

CONCLUSÃO

Os modelos de séries temporais baseados na metodologia Box e Jenkins mostraram ser uma excelente ferramenta para a previsão de precipitação sazonal para a região semiárida do Nordeste Brasileiro. Dentre os modelos testados, mesmo apresentando alguma flutuação, o modelo sazonal, SARIMA (1,1,2)x(1,1,2)₁₂, foi o selecionado e mostrou-se apropriado para previsão de um período de seis meses, com valores próximos aos da climatologia mensal local como também aproximados aos observados. Ressaltando ainda, que tais modelos não levam em consideração oscilações climáticas de grande escala como o fenômeno EL NINO por exemplo, que podem influenciar diretamente nos mecanismos provedores de chuvas em determinadas regiões.

REFERÊNCIAS

- BOX, G. E. P.; JENKINS, G. M. *Time series analysis forecasting and control*. San Francisco: H. Day, 1976.
- EHLERS, R. S. *Análise de séries temporais*. 2009. Disponível em: <<http://www.icmc.usp.br/ehlers/stemp/stemp.pdf>>. Acesso em: 22 fev. 2012.





SOBRE ÁGUA NO SEMIÁRIDO BRASILEIRO

HANKE, J. E.; WICHERN, D.; REITSCH, A. *Business forecasting*. 7. ed. New Jersey: Prentice Hall, 2001.

HAAN, C.T. **Statistical Methods in Hydrology**. Ames: The Iowa State University Press. 2ª Edição, 2002, 377p.

LEE, Cheng-Min; KO, Chia-Nan. Short-term load forecasting using lifting scheme and ARIMA models. *Expert Systems with Applications*, v. 38, p. 5902-5911, 2011.

SCHMIDT, D. M. *Dinâmica das configurações de formação e inibição das chuvas no Rio Grande do Norte: caracterização hidroclimática do estado* - Tese (Doutorado) – Universidade Federal do Rio Grande do Norte. Programa de Pós-Graduação em Ciências Climáticas. Natal, 2014. 136 p.

SOUZA, G. P.; SAMOHYL, R. W.; MIRANDA, R. G. *Métodos simplificados de previsão empresarial*. Rio de Janeiro: Ciência Moderna, 2008. 181 p.

AGRADECIMENTOS

Os autores agradecem ao Programa de Pós-Graduação em Ciências Climáticas da UFRN pelo apoio institucional, à Agência Nacional de Águas (ANA) pelos dados pluviométricos cedidos e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Programa Nacional de Pós-Doutoramento (PNPD) pelo auxílio financeiro durante o desenvolvimento desta pesquisa.

