**Aplicando Séries Temporais em dados hidrológicos**

**Flávia Guimarães Gaia Paula1**

1Centro Universitário Instituto de Educação Superior de Brasília (IESB)

Brasília – DF – Brazil

flavia.paula@iesb.edu.br

1. **Introdução**

Uma série temporal, também denominada série histórica, é uma sequência de dados obtidos em intervalos regulares de tempo durante um período específico 1,2. Se a série histórica foi denominada como Z, o valor da série no momento t pode ser escrito como Zt(t=1,2,...,n). Denomina-se trajetória de um processo, a curva obtida no gráfico da série histórica e o conjunto de todas as possíveis trajetórias é denominado como um processo estocástico. Considera-se que uma série temporal é uma amostra deste processo. Uma série histórica pode ser composta por três componentes não observáveis: tendência (Tt), sazonalidade (St) e a variação aleatória denominada de ruído branco (at) (LATORRE; CARDOSO, 2001).

Um dos objetivos principais de uma série temporal é a utilização ou desenvolvimento de um modelo para descrever o fenômeno sob consideração. Podemos classificar os modelos para séries temporais em duas classes, segundo o número de parâmetros envolvidos em (MORETTIN; TOLOI, 2004):

· modelos paramétricos, em que esse número de parâmetros é finito;

· modelos não-paramétricos, que envolvem um número infinito

de parâmetros.

Na classe de modelos paramétricos, a análise é feita no domínio do tempo. Dentre esses modelos, os mais frequentemente usados são os modelos de erro (ou de regressão), os modelos auto-regressivos-médias móveis (ARMA) e os modelos auto-regressivos-integrados-médias móveis (ARIMA) e, para incorporar o comportamento de sazonalidade, utilizam se os modelos ARIMA sazonais multiplicativos (SARIMA) (MORETTIN; TOLOI, 2004).

Os modelos ARIMA exploram a autocorrelação entre os valores da série em instantes sucessivos, mas, quando os dados são observados em períodos inferiores a um ano, a série também pode apresentar autocorrelação para uma estação de sazonalidades. Os modelos que contemplam as séries que apresentam autocorrelação sazonal são conhecidos como SARIMA. O modelo SARIMA contém uma parte não sazonal com parâmetros (p,d,q) e uma sazonal com parâmetros (P,D,Q) (ESPINOSA; PRADO; GHELLERE, 2001).

O presente trabalho é sobre o uso de séries temporais, mais concretamente sobre a aplicação deste em dados hidrológicos, utilizando a linguagem R. É objetivo deste trabalho encontrar o melhor modelo de séries temporais para esse dataset e analisá-lo.

1. **Desenvolvimento**

Para a realização deste trabalho foram utilizados dados do Portal do Departamento de Águas e Energia Elétrica(DAEE) que gera e disponibiliza, via Internet (http://www.hidrologia.daee.sp.gov.br/), foram utilizados dados pluviométrico do município de São Paulo, em um intervalo entre janeiro de 1985 e dezembro de 2019.

O projeto foi desenvolvido em 9 etapas:

• Etapa 1: etapa que consiste na obtenção dos dados por meio do banco de dados da DAEE.

• Etapa 2: etapa de preparação dos dados para serem utilizados pelos algoritmos, limpeza e tratamento dos dados.

• Etapa 3: etapa de criação da série temporal com os parâmetros : data de início, data final e frequência mensal. Inicialmente a série não tem tendência mas tem sazonalidade.

• Etapa 4: etapa de análise inicial, em que é apurado como aprimorar a série temporal por meio da decomposição, suavização, normalidade e transformação.

• Etapa 5: o teste da estacionariedade de Phillips-Perron mostra que a série não é estacionária, sendo p <= 0.05. A diferenciação mostra que isso não pode ser modificado.

• Etapa 6: a etapa de autocorrelação indica que existe autocorrelação e uma grande sazonalidade nos dados.

• Etapa 7: etapa de criação de 2 modelos SARIMA, com sazonalidades diferentes para verificar qual dos 2 modelos é melhor.

• Etapa 8: etapa que consiste na avaliação e interpretação dos resultados obtidos. A RMSE e AIC indicam que o segundo modelo tende a ser mais favorável, a análise dos resíduos tende uma normalidade e não está correlacionada, ou seja, os dois modelos foram aprovados.

• Etapa 9: é feita a previsão dos modelos para 16 meses e avaliado por meio das métricas RMSE e desvio padrão do erro absoluto, que provam que o modelo 1 se saiu melhor nas previsões e o modelo 2 é melhor no ajuste com relação aos dados utilizados.

1. **Conclusão**

O trabalho desenvolvido neste artigo tem como principais resultados, a criação de dois modelos SARIMA com variação nos parâmetros de sazonalidade. Os mesmos possuem boas métricas de aceitação, com distinção que o primeiro modelo é mais adequado para previsões e o segundo modelo é ideal para o ajuste com relação aos dados deste projeto

Todos os objetivos deste trabalho foram cumpridos, uma vez foi construído um modelo de séries temporais adequado para prever a quantidade de chuva dos próximos 16 meses, com base em dados anteriores.

**Referências**

LATORRE, M.R.D.O.;CARDOSO,M.R.A. Análise de séries temporais em epidemiologia. **Rev. Bras. Epidemiol. Vol. 4, Nº 3, 2001**. Disponível em: <https://www.scielo.br/j/rbepid/a/KM9MndgpCGSnjSNDddSydCG/?format=pdf&lang=pt>. Acesso em: 20 de maio de 2022.

BLEIDORN, M.T.; PINTO, W.P; BRAUN,E.S.; LIMA, G.B; MONTEBELLER, C.A. .MODELAGEM E PREVISÃO DE VAZÕES MÉDIAS MENSAIS DO RIO JUCU, ES, UTILIZANDO O MODELO SARIMA. **Irriga, Botucatu, v. 24, n. 2, p. 320-335,** 2019. Disponível em: <https://revistas.fca.unesp.br/index.php/irriga/article/view/3207/2507>. Acesso em: 6 de junho de 2022.

MORETTIN, P. A.; TOLOI, C. M. C. Análise de séries temporais. 1 Ed. São Paulo. Editora Edgar Blücher LTDA, 535 p, 2004.

ESPINOSA, M.M.; PRADO, S.M; GHELLERE, M. Uso do modelo SARIMA na previsão do número de focos de calor para os meses de junho a outubro no Estado de Mato Grosso. **Periodicos UFSM,** 2010. Disponível em: <https://periodicos.ufsm.br/cienciaenatura/article/download/9482/5631>. Acesso em: 6 de junho de 2022.