APLICAÇÃO DO PACOTE ALEA-R PARA ANÁLISE DE FREQUÊNCIA LOCAL DE DADOS PLUVIOMÉTRICOS

Rachel Nolibos Ramos [1]; Lucas Guimarães Brighenti[2] & Vitor Matheus Passos Martins[3]

Palavras-Chave – ALEA-R, análise de frequência local, máximos anuais de precipitação.

INTRODUÇÃO

O ALEA (Análise de Frequência Local de Eventos Anuais) é um programa que associa integradamente ferramentas estatísticas e computacionais que permitem auxiliar o estudante e o profissional da área de engenharia de recursos hídricos na análise de frequência de observações máximas, médias e mínimas anuais de variáveis hidrológicas (DE LIMA e NAGHETTINI, 2001). Atualmente, a disponibilidade de extensas séries temporais provenientes de estações de monitoramento, sensores automáticos, sistemas de sensoriamento remoto e bases de dados integradas exige ferramentas capazes de processar grandes volumes de informações de forma eficiente, consistente e reproduzível. Além de aumentar a velocidade das análises, a implementação computacional permite automatizar rotinas complexas, reduzir erros associados a procedimentos manuais, garantir rastreabilidade dos resultados e facilitar a atualização contínua dos estudos à medida que novos dados se tornam disponíveis. Nesse contexto, a reprodução de metodologias consolidadas em ambientes de programação possibilita a aplicação de técnicas estatísticas mais robustas, promove a transparência científica e amplia a capacidade de análise necessária para atender às demandas atuais da hidrologia. Portanto, o ALEA-R (FERNANDES, 2026) implementa um fluxo de trabalho reproduzível do software ALEA em um pacote da linguagem R, com interface pública do pacote projetada para ser intencionalmente simples, focada na execução eficiente das etapas de análise de séries hidrológicas de máximos. O pacote inclui ajuste de distribuições de probabilidade, estimativa de quantis, cálculo de intervalos de confiança por *bootstrap*, avaliação da qualidade do ajuste, diagnósticos amostrais, suporte assistido por inteligência artificial para seleção de distribuições, análises em lote e geração de gráficos e arquivos de exportação prontos para publicação. e suporte às distribuições Generalizada de Valores Extremos (GEV), Pearson Tipo III (PE3), Lognormal de dois parâmetros (LN2), Lognormal de três parâmetros (LN3) e Gumbel (GUM), amplamente utilizadas em análises de frequência hidrológica e estimativa de eventos extremos.

METODOLOGIA

À fim de testar e implementar o pacote em R e verificar quanto à acurácia das respostas obtidas comparando-se ao programa ALEA, este estudo foi desenvolvido com base nos dados de monitoramento disponíveis na estação 02346076 do HIDROWEB/ANA. Foi adotado o Método dos Momentos (LMOM) que se baseia na equivalência entre os momentos calculados a partir da amostra e os momentos teóricos da população, permitindo a obtenção das estimativas dos parâmetros da distribuição analisada. Embora seja uma abordagem relativamente simples do ponto de vista matemático, o MOM apresenta desempenho satisfatório e, em amostras de pequeno tamanho, pode produzir estimativas comparáveis ou até superiores às obtidas por outros métodos de estimação (NAGHETTINI e PINTO, 2007). Visto que os dados da estação apresentaram inconsistências significativas, foi feita uma validação/consolidação de 72 anos de dados por meio de avaliação

1) Afiliação: Engenheira de Produção Civil pelo Centro Federal de Educação Tecnológica de Minas Gerais, ; rachelnramos@hotmail.com.
2) Afiliação: Engenheiro Civil pela Universidade Federal de Minas Gerais/UFMG Campus Pampulha, lucasgbrighenti3@gmail.com.
3) Afiliação: Engenheiro Civil e Mestre em Saneamento, Meio Ambiente e Recursos Hídricos pela Universidade Federal de Minas Gerais/UFMG Campus Pampulha, vitorm601@gmail.com.

minuciosa das séries históricas e a comparação com registros jornalísticos de eventos relevantes e, portanto, fez-se necessário ajustes nos dados amostrais.

RESULTADOS

Após ajustes nos dados da estação, os máximos diários obtidos para cada ano hidrológico (iniciando em outubro e finalizando em setembro) estão apresentados na Tabela 1:

Tabela 1 – Máximos analisados nos 72 anos amostrados da estação 02346076 (mm).

Ano hidrológico	Máximo Diário (mm)	Ano hidrológico	Máximo Diário (mm)	Ano hidrológico	Máximo Diário (mm)
1924-1925	225,80	1949-1950	191,10	1973-1974	164,50
1925-1926	202,70	1950-1951	173,70	1974-1975	179,00
1926-1927	136,10	1951-1952	144,80	1975-1976	286,30
1927-1928	260,60	1952-1953	260,60	1976-1977	166,70
1928-1929	362,00	1953-1954	188,20	1977-1978	148,10
1929-1930	309,90	1954-1955	130,30	1978-1979	136,00
1931-1932	231,60	1955-1956	251,90	1979-1980	124,40
1932-1933	130,30	1956-1957	162,20	1980-1981	198,50
1933-1934	234,50	1957-1958	199,80	1981-1982	102,60
1934-1935	304,30	1958-1959	165,10	1982-1983	133,30
1935-1936	231,60	1959-1960	124,50	1983-1984	148,20
1936-1937	130,30	1960-1961	167,90	1984-1985	171,00
1937-1938	225,80	1961-1962	115,80	1985-1986	180,60
1938-1939	121,50	1962-1963	147,60	1986-1987	106,00
1939-1940	156,40	1963-1964	75,20	1988-1989	214,70
1940-1941	240,30	1964-1965	193,90	1989-1990	218,10
1941-1942	254,80	1965-1966	202,70	1990-1991	89,60
1942-1943	95,50	1966-1967	225,80	1991-1992	228,60
1943-1944	280,90	1967-1968	110,00	1992-1993	295,10
1944-1945	127,50	1968-1969	101,30	1993-1994	283,90
1945-1946	130,30	1969-1970	173,70	1994-1995	110,00
1946-1947	289,60	1970-1971	126,50	1995-1996	114,00
1947-1948	150,60	1971-1972	214,30	1997-1998	218,10
1948-1949	121,50	1972-1973	119,20	1999-2000	210,70

As distribuições comparadas foram GEV, Gumbel (GUM), Lognormal de dois parâmetros (LN2), Lognormal de três parâmetros (LN3) e Pearson Tipo III (PE3), resultando nos parâmetros estatísticos da Tabela 2, que foram idênticos aos analisados no ALEA-R:

Tabela 2 – Parâmetros de local, escala e forma das distribuições para este estudo.

Distribuição	Local	Escala	Forma
GEV	154,232398	54,4702634	0,05979451
GUM	152,779304	51,6942863	NA
LN2	5,145671	0,3513571	NA
LN3	-52,822423	0,2714071	5,42462723
PE3	182,618056	64,8073695	0,80482466

Além dos parâmetros estatísticos, o ALEA-R apresenta graficamente os ajustes das funções de distribuições e os 72 dados amostrados, conforme Figura 1:

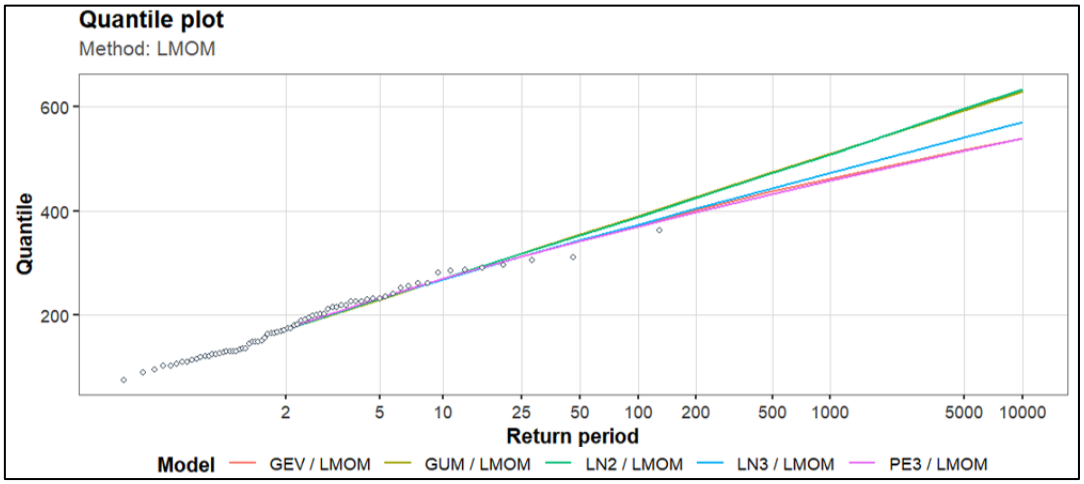


Figura 1. Análise de frequência. Fonte: ALEA-R.

Para escolha da distribuição mais aderente aos dados, espera-se obter menores valores para as estatísticas de Kolmogorov-Smirnov (KS), Cramér-von Mises (CVM), Anderson-Darling (AD), Akaike *information criterion* (AIC) e *Bayesian information criterion* (BIC), e maiores valores para a *Log-likelihood* (LOGLIK). Conforme mostrado na Tabela 3, a modelagem dos dados no ALEA-R mostrou uma melhor aderência das informações hidrológicas avaliadas para com a distribuição estatística PE3, de forma geral.

Tabela 3 – Resumo de resultados das análises estatísticas.

Distribuição	Estatística	Estimativa
GEV	KS	0,084366
GEV	CVM	0,067632
GEV	AD	0,424287
GEV	LOGLIK	-396,805
GEV	AIC	799,6107
GEV	BIC	806,4407
GUM	KS	0,081952
GUM	CVM	0,07733
GUM	AD	0,454565
GUM	LOGLIK	-396,726
GUM	AIC	797,452
GUM	BIC	802,0053
LN2	KS	0,079058
LN2	CVM	0,07118
LN2	AD	0,417984
LN2	LOGLIK	-396,411
LN2	AIC	796,8219
LN2	BIC	801,3752
LN3	KS	0,083602
LN3	CVM	0,066519
LN3	AD	0,418945

Distribuição	Estatística	Estimativa
LN3	LOGLIK	-396,758
LN3	AIC	799,5169
LN3	BIC	806,3469
PE3	KS	0,079812
PE3	CVM	0,058775
PE3	AD	0,373826
PE3	LOGLIK	-396,334
PE3	AIC	798,6677
PE3	BIC	805,4977

Assim como no ALEA-R, a distribuição Pearson Tipo III (PE3) foi considerada a mais aderente aos dados por apresentar o melhor desempenho na maior parte dos critérios diretamente associados à qualidade do ajuste. Essa distribuição obteve os menores valores das estatísticas Cramér-von Mises e Anderson-Darling, indicando menor discrepância global entre a distribuição teórica e os dados observados, bem como melhor representação das caudas da distribuição. Além disso, a PE3 apresentou a maior log-verossimilhança, sugerindo maior compatibilidade probabilística entre o modelo ajustado e a amostra analisada. Embora os critérios AIC e BIC tenham favorecido a distribuição LN2, esses indicadores incluem penalização pela complexidade do modelo. Portanto, considerando prioritariamente a aderência estatística aos dados, sobretudo em relação ao comportamento dos eventos extremos, a distribuição PE3 mostrou-se a alternativa mais adequada entre as distribuições avaliadas.

CONCLUSÕES

Conclui-se que o pacote ALEA-R reproduz fielmente o programa ALEA e as respostas obtidas para a amostragem desse estudo. Contudo, o pacote em R ainda passará por atualizações e atualmente suas limitações são a não inclusão da distribuição Log-Pearson Tipo III (LP3), a disponibilidade da API em português, *p-values* calibrados para testes de aderência, testes de aderência qui-quadrado, intervalos de confiança para os parâmetros estimados e para quantis obtidos por métodos assintóticos ou pelo método delta, nem acesso aos dados da plataforma HIDROWEB/ANA diretamente no pacote principal, sendo esta última uma melhoria que pode ser facilmente implementada pelo usuário.

REFERÊNCIAS

DE LIMA, A, A.; NAGHETTINI, M. *ALEA–uma ferramenta prática e consistente para a análise de freqüência local de eventos hidrológicos anuais*. 2001.

FERNANDES W.; _ALEA-R: *Hydrological Frequency Analysis in R_. R package version* 0.1.0, https://github.com/hydrostat/ALEA (2026).

NAGHETTINI, M.; PINTO, E. J. A. *Hidrologia estatística*. CPRM, 2007.