

ERRO TIPO I E “PODER” DE TESTES DE ADERÊNCIA PARA POSTOS PLUVIOMÉTRICOS NO CEARÁ

TYPE I ERROR AND “POWER” OF ADHERENCE TESTS FOR PLUVIOMETRIC GAUGES IN CEARÁ

Danilo Nogueira de Souza¹, Ticiane Marinho de Carvalho Studart², Iran Eduardo Lima Neto³

¹Doutorando no Programa de Pós-graduação em Engenharia Civil da Universidade Federal do Ceará, CE, Brasil. E-mail: danilomaxbr@yahoo.com.br

²Professora Doutora da Universidade Federal do Ceará, CE, Brasil. E-mail: ticiane@ufc.br

³Professor Doutor da Universidade Federal do Ceará, CE, Brasil. E-mail: iran@ufc.br

RESUMO

Devido a curta extensão das séries históricas de precipitação e vazão disponíveis, utilizam-se testes de aderência para verificar o ajuste de distribuições de probabilidade a estas séries. Quando aderidas, as distribuições podem estimar os valores críticos dos fenômenos e as suas ocorrências. Sendo assim, realizou-se a análise de quatro testes de aderência – Kolmogorov-Smirnov, Qui-Quadrado, Cramer-von Mises e Anderson-Darling –, quanto ao Erro Tipo I e ao “Poder”, determinou-se o comportamento de cada teste no ajuste de quatro distribuições de probabilidade: Gama, Normal, Lognormal e Gumbel. Tal estudo foi feito via Simulação Monte Carlo. Os resultados das análises dos testes foram utilizados na definição de qual distribuição de probabilidade se ajustava melhor às séries históricas de 24 postos pluviométricos no Estado do Ceará, Brasil. Verificou-se que, para o coeficiente de variação e a extensão dos dados dos postos, o teste com Erro Tipo I mais próximo do valor esperado e com maior “Poder” é o Qui-Quadrado.

Palavras-chave: séries históricas, fenômenos hidrológicos, séries sintéticas, simulação Monte Carlo.

ABSTRACT

Due to the short extent of available historical time series of rainfall and flow, adherence tests are carried out to verify the fitting of different probability distributions to such time series. When there is adherence, the distributions can estimate the critical values of the phenomena and their occurrence. Thus, the analysis of four adherence tests was conducted including - Kolmogorov-Smirnov, Chi-Square, Cramer-von Mises and Anderson-Darling – considering Type I Error and “Power”, in order to determine the behavior of each test regarding the fitting of four probability distributions: Gamma, Normal, Lognormal and Gumbel. Such study has been done via Monte Carlo simulation. The results of the analysis of the tests were used in the definition of which probability distribution was a better fit for the historical time series of 24 pluviometric gauges in the State of Ceará, Brazil. It was verified that, considering the coefficient of variation and the extent of data of the gauges, the Type I Error test that is closer to the expected values and has greater “power” is the Chi-Square.

Keywords: historical time series, hydrological phenomena, synthetic series, Monte Carlo Simulation.

1 – INTRODUÇÃO

A necessidade humana em garantir o seu suprimento de água é atendida pelas intervenções (obras de engenharia) na natureza: canais, aquedutos, barragens, poços, etc. Tais obras visam resolver problemas relativos a demandas futuras de abastecimento de água. Todavia, para os projetos destas estruturas, dispõe-se somente de séries de dados que representam o passado.

No Brasil, tanto os primeiros postos fluviométricos (ANA, 2007) como os primeiros postos pluviométricos (TUCCI, 2009) são do fim do século XIX.

As séries históricas nacionais de precipitação e vazão podem ser obtidas na página da Agência Nacional de Águas – ANA, a qual opera 4.543 estações hidrometeorológicas (ANA, 2012). Logo, vê-se que as extensões das séries de precipitação e de vazão são curtas – 113 anos de dados –, não descrevendo, portanto, com precisão, o comportamento dos fenômenos. Assim, para tratar do problema das séries curtas, adota-se o

procedimento de ajustar uma distribuição de probabilidade aos dados medidos por meio de testes de aderência.

Como exemplo, os trabalhos de Sampaio *et al.* (2007); Silvino *et al.* (2007) e Deka *et al.* (2009) estudaram o assunto.

Este artigo faz uma análise comparativa entre quatro testes de aderência, utilizando séries sintéticas com distribuição de probabilidade conhecida, observando a taxa de Erro do Tipo I e o “Poder” dos testes, de acordo com o coeficiente de variação e a extensão dessas séries.

Em seguida, estes testes foram aplicados para definir a distribuição que melhor adere aos dados de 24 postos pluviométricos do Estado do Ceará. Estas análises foram realizadas via simulação Monte Carlo.

2 – MATERIAL E MÉTODOS

Para determinar o Erro Tipo I e o “Poder” dos quatro testes de aderência estudados – Kolmogorov-Smirnov, Qui-Quadrado, Anderson-Darling e Cramer-von Mises – foram geradas séries que seguiam uma distribuição de

probabilidade Gama II. As distribuições de probabilidade utilizadas foram: Gama (verdadeira), Normal, Lognormal e Gumbel (falsas).

As séries dividiam-se em quatro valores de coeficientes de variação (CV), dado pela razão entre o desvio-padrão e a média da série: 0,1; 0,5; 1,0 e 1,5. Para cada CV, foram geradas 30 séries, cujas médias eram iguais a 100, com extensões de 50, 100, 200, 300, 400, 500, 600, 700, 800, 900 e 1.000 dados.

O Erro Tipo I é o número de vezes que os testes rejeitaram a aderência da distribuição Gama às séries sintéticas. Já o “Poder” dos testes é o número de rejeições de aderência das distribuições Normal, Lognormal e Gumbel às séries sintéticas. Os testes foram realizados a um nível de significância (α) de 5%, que, pela teoria, deve ser igual à taxa de Erro Tipo I.

Como resultado, tem-se a determinação do teste de aderência cujo resultado (rejeição ou não rejeição) seja o mais confiável para o ajuste de uma das distribuições testadas a uma dada série histórica (intensidade pluviométrica e vazão), considerando o CV e a extensão da série.

Foram analisados, ainda, os impactos nos valores críticos (valores correspondentes a certos períodos de retorno) ao se usar uma distribuição distinta da que melhor aderiu à série histórica.

Após a análise das séries sintéticas foi feito um estudo em séries de dados reais de precipitação para 24 postos pluviométricos no Estado do Ceará.

2.1 Testes de aderência

São métodos matemático-estatísticos que avaliam se uma série teórica de dados (distribuição de probabilidade) descreve o comportamento de uma série real de dados. Na sequência são apresentadas as quatro metodologias aplicadas neste estudo.

2.1.1 Teste de Kolmogorov-Smirnov (KS)

A metodologia do teste baseia-se na diferença absoluta máxima (D) entre os valores obtidos de uma distribuição empírica de tamanho N e os valores obtidos de uma distribuição teórica específica. A estatística do teste de Kolmogorov-Smirnov é calculada pela Equação (1):

$$D = \text{Max } |F_e(x) - F_o(x)| \quad (1)$$

Em que: D é a diferença máxima absoluta entre os valores esperados e observados;

$F_e(x)$ são os valores esperados (distribuição assumida);

$F_o(x)$ são os valores observados (série de dados real).

Este teste não se aplica ao ajuste de distribuições discretas (JOHNSON, 2000). Trabalhos como os de Sampaio *et al.* (2007), Hartmann *et al.* (2011) e Razali & Wah (2011) utilizaram o teste.

2.1.2 Teste do Qui-Quadrado (QQ)

É um teste de aderência clássico, sendo, talvez, o mais utilizado. A estatística do teste é calculada pela Equação (2):

$$\chi^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i} \quad (2)$$

Com k – m graus de liberdade

Em que: o_i são as frequências observadas;

e_i são as frequências esperadas;

k é o número de classes;

m é o número de parâmetros da distribuição.

Rahman *et al.* (2006), Ribeiro *et al.* (2007), Silvino *et al.* (2007) e Araújo *et al.* (2010) usaram o teste.

2.1.3 Teste de Cramer-von Mises (CVM)

A estatística do teste de Cramer-von Mises é calculada pela Equação (3):

$$W^2 = \frac{1}{12n} + \sum_{i=1}^n \left[F(x_i) - \frac{(2i-1)}{2n} \right]^2 \quad (3)$$

Em que: W^2 é o valor da estatística do teste;

n é o tamanho da amostra;

$F(x_i)$ é o valor segundo a distribuição presumida.

2.1.4 Teste de Anderson-Darling (AD)

Este teste originou-se de uma modificação do teste de Cramer-von Mises. O valor da estatística do teste de Anderson-Darling é calculado pela Equação (4):

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n [(2i-1) \text{Ln} Z_i + (2n+1 - 2i) \text{Ln}(1 - Z_i)] \quad (4)$$

Em que: Z_i é a função de distribuição acumulada a ser testada;

n é o tamanho da amostra.

Por exemplo, o teste de Anderson-Darling foi usado por Razali; Wah (2011).

2.2 Distribuições de probabilidade

Uma distribuição de probabilidade aderida a uma série de dados (intensidade pluviométrica e vazão) pode estimar os valores futuros, principalmente os extremos, usados em projetos de obras hidráulicas, além de estimar, também, a sua frequência ou período de retorno.

2.2.1 Distribuição gama

A função densidade de probabilidade da distribuição Gama é calculada pela Equação (5):

$$f(x) = \frac{x^{(\alpha-1)} e^{\left(-\frac{x}{\beta}\right)}}{\beta^\alpha \Gamma(\alpha)} \quad (5)$$

Para $x > 0$, $\alpha > 0$, $\beta > 0$; e $f(x) = 0$ para outros valores.

A função $\Gamma(\alpha)$ é o valor da distribuição Gama, a qual é calculada pela Equação (6):

$$\Gamma(\alpha) = \int_0^{\infty} x^{(\alpha-1)} e^{(-x)} dx \quad (6)$$

Os parâmetros α e β são estimados por meio da média (μ) e da variância (σ^2) da distribuição Gama (Yevjevich, 1972). Assim:

$$\mu = \alpha\beta \quad (7)$$

e

$$\sigma^2 = \alpha\beta^2 \quad (8)$$

A distribuição Gama ajusta-se bem (descreve o comportamento) à precipitação no Estado do Paraná (LONGO *et al.*, 2006); às vazões anuais de 1.221 rios estudados por McMahon *et al.* (2007a), distribuídos pelos cinco continentes; à precipitação em Garanhuns, Pernambuco (ANDRADE *et al.*, 2008).

2.2.2 Distribuição Gumbel (ou extremo tipo I)

A função densidade de probabilidade da distribuição Gumbel é calculada pela Equação (9):

$$f(x) = e^{(-e^{(-\alpha(x-\beta))})} \quad (9)$$

Para um valor de $\alpha > 0$.

Os parâmetros α e β são, respectivamente, de posição e de escala.

A distribuição Gumbel adere bem às vazões máximas anuais do córrego Capetinga, Distrito Federal (SILVA, 2003); à precipitação diária máxima, períodos mensal e anual, em Lavras, Minas Gerais (BEIJO *et al.*, 2005).

2.2.3 Distribuição normal

A distribuição Normal é calculada pela Equação (10):

$$f(x) = \frac{e^{(-\frac{(x-\mu)^2}{2\sigma^2})}}{\sigma\sqrt{2\pi}} \quad (10)$$

Para todo valor de $X \in \mathbb{Z}$.

Os valores de μ e σ são, respectivamente, a média e o desvio-padrão da distribuição Normal.

A distribuição Normal adere bem às temperaturas diárias do ar máxima e mínima em Iguatu, Ceará (ARAÚJO *et al.*, 2010).

2.2.4 Distribuição lognormal

A distribuição Lognormal é calculada pela Equação (11):

$$f(x) = \frac{e^{(-\frac{1}{2} \frac{(\ln x - \mu_n)^2}{\sigma_n^2})}}{x\sigma_n\sqrt{2\pi}} \quad (11)$$

Para $x > 0$. Para $x \leq 0$, $f(x) = 0$.

As relações entre a média (μ_n) e o desvio-padrão (σ_n) da distribuição Lognormal e a média (μ) e o desvio-padrão (σ) da variável aleatória (x) são calculadas, respectivamente, pelas Equações (12) e (13):

$$\mu = e^{(\mu_n + \frac{1}{2}\sigma_n^2)} \quad (12)$$

$$\sigma^2 = \mu^2(e^{\sigma_n^2} - 1) \quad (13)$$

A distribuição Lognormal adere bem às vazões anuais da parte continental dos Estados Unidos (VOGEL *et al.*, 1999).

2.3 Recursos computacionais utilizados

Os programas usados para os cálculos estatísticos foram: EasyFit versão 5.5, Statgraphics versão 16.1.11 e XLSTAT 2012. As séries foram geradas no programa SimRes versão 1.2, desenvolvido no Departamento de Engenharia Hidráulica e Ambiental – DEHA da Universidade Federal do Ceará – UFC.

2.4 Método de Monte Carlo

É um tipo de simulação que consiste basicamente na geração de várias séries sintéticas cujos valores seguem uma distribuição de probabilidade previamente definida. Tal método é usado ainda para avaliar a sensibilidade da resposta do modelo testado às mudanças feitas nos seus parâmetros, no estado das variáveis ou nos dados de entrada (WAGENER; KOLLAT, 2007).

Esse método é especialmente útil quando há uma limitação no número disponível de dados de uma determinada variável estudada. Com um grande número de séries geradas, com diferentes extensões de dados, pode-se determinar os limites de aplicação do método avaliado.

Assim, este artigo estudou a sensibilidade dos testes de aderência às modificações nas séries sintéticas geradas, tais como: tipo de distribuição, coeficiente de variação e extensão das séries.

2.5 Aplicação em casos práticos

Os vinte e quatro postos pluviométricos do Estado do Ceará estudados foram: Fortaleza (posto Pici), Farias Brito, Frecheirinha, Graça, Guaraciaba do Norte, Itapajé, Milagres, Hidrolândia, Ibiapina, Icó, Iguatu, Jaguaruana, Jardim, Limoeiro do Norte, Maranguape, Missão Velha, Mucambo, Novo Oriente, Palhano, Parambu, Paramoti, Porteiras, Senador Pompeu e Tamboril. Com exceção de Fortaleza, todos os postos se encontram inseridos na região semiárida cearense.

Foram estudadas as séries históricas de precipitações máximas anuais e totais anuais para cada um dos postos. Calcularam-se, ainda, as precipitações máximas e totais, referentes a períodos de retorno de 5, 10, 100, 1.000 e 10.000 anos, segundo as distribuições ajustadas e as não ajustadas. As precipitações correspondentes a períodos de 1.000 e 10.000 anos são utilizadas nos projetos hidrológicos de barragens.

resultados da aderência das distribuições estudadas às séries dos postos pluviométricos.

3.1 Erro do Tipo I dos testes

Na Tabela 1 são mostradas as taxas de Erro Tipo I relativo apenas à distribuição Gama, dos testes de aderência para cada extensão de série, segundo cada valor de CV.

3 – RESULTADOS E DISCUSSÃO

A seguir são apresentados os resultados das análises de Erro Tipo I e “Poder” dos testes de aderência e os

Tabela 1 – Taxas de erro tipo I dos testes de aderência para cada CV e extensão de série

(A) Erro Tipo I – CV = 0,1											
Testes	Extensões das séries (anos)										
	50	100	200	300	400	500	600	700	800	900	1.000
KS	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
QQ	0,0333	0,0333	0,0667	0,0333	0,0333	0,0000	0,1000	0,0333	0,0333	0,0333	0,1000
AD	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
CVM	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
(B) Erro Tipo I – CV = 0,5											
Testes	Extensões das séries (anos)										
	50	100	200	300	400	500	600	700	800	900	1.000
KS	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
QQ	0,0000	0,0333	0,0000	0,0667	0,0333	0,0000	0,0667	0,0333	0,0667	0,0333	0,1333
AD	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
CVM	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
(C) Erro Tipo I – CV = 1,0											
Testes	Extensões das séries (anos)										
	50	100	200	300	400	500	600	700	800	900	1.000
KS	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
QQ	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
AD	0,0000	0,0000	0,0000	0,0333	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0333
CVM	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
(D) Erro Tipo I – CV = 1,5											
Testes	Extensões das séries (anos)										
	50	100	200	300	400	500	600	700	800	900	1.000
KS	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
QQ	1,0000	1,0000	1,0000	0,0333	0,0333	0,0333	0,0333	0,0667	0,0333	0,0667	0,0000
AD	0,1333	0,0333	0,2667	0,1000	0,0667	0,1333	0,1333	0,0667	0,1000	0,1000	0,0667
CVM	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Para séries com CV = 0,1 (Tabela 1A), o Qui-Quadrado é o único que apresenta erro tipo I mais próximo ao nível de significância adotado (5%), logo é o mais confiável. Logo, como exemplo de sua aplicação, poderia ser utilizado em testes de aderência nas séries de vazões anuais dos Rios Yenisei (Rússia), Mackenzie (Canadá) e Amazonas (Brasil), pelos dados de Nijssen *et al.* (2001); e nas séries de vazões anuais no Norte do Canadá, pelos dados de Déry e Wood (2005).

Para séries com CV = 0,5 (Tabela 1B), tem-se um comportamento semelhante ao descrito no Gráfico 1A. Aplica-se, por exemplo, o Qui-Quadrado, o teste mais

confiável, às séries de precipitação anual da cidade de Atenas (Grécia), segundo dados de Koutsoyiannis e Baloutsos (2000); às séries de precipitações de alguns municípios do semiárido do Estado de Sergipe segundo dados de Santos *et al.* (2009); às séries de precipitação das regiões da zona da mata, do agreste, do meio norte e do semi-árido, todos no nordeste brasileiro, segundo dados de Silva *et al.* (2011).

Para séries com CV = 1,0 (Tabela 1C), os testes KS, CVM e AD apresentam erro tipo I nulo para todas as extensões, porém mais próximos à significância do que o teste QQ, que apresenta erro tipo I igual 100% para todas

as extensões de séries, inviabilizando o seu uso para este coeficiente de variação. Assim, podem ser empregados os testes KS, CVM e AD nas séries de vazões anuais do Zimbábue, segundo dados de McMahon et al. (2007b).

Para séries com CV = 1,5 (Tabela 1D), os testes KS e CVM apresentam erro tipo I nulo para todas as extensões. Já o teste QQ apresenta erro tipo I igual 100% para séries de até 200 dados, o que inviabiliza o seu uso para este coeficiente de variação, nas extensões das séries atuais. A partir de n = 300, entretanto, os erros destes três testes passam a oscilar muito próximos da significância. Já o

teste AD é o que apresenta, em geral, erros tipo I maiores que a significância. Empregam-se, por exemplo, os testes KS e CVM, cujos erros são mais próximos à significância, no ajuste às séries de vazões afluentes anuais ao reservatório Pedras Brancas, no Ceará, segundo dados de Farias (2003).

3.2 “Poder” dos testes

Na Tabela 2 é mostrado o “Poder” dos testes de aderência para a distribuição Normal, segundo os valores de CV.

Tabela 2 – “Poder” da distribuição Normal

(A) "Poder" da distribuição Normal – CV = 0,1											
Testes	Extensões das séries (anos)										
	50	100	200	300	400	500	600	700	800	900	1.000
KS	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0333
QQ	0,0000	0,0000	0,0333	0,0000	0,0000	0,0667	0,1667	0,0667	0,0333	0,1667	0,1333
AD	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
CVM	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
(B) "Poder" da distribuição Normal – CV = 0,5											
Testes	Extensões das séries (anos)										
	50	100	200	300	400	500	600	700	800	900	1.000
KS	0,0000	0,0667	0,3667	0,8000	0,8667	0,9000	1,0000	1,0000	1,0000	1,0000	1,0000
QQ	0,1667	0,4667	0,6667	0,9667	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
AD	0,0000	0,0667	0,4000	0,9333	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
CVM	0,0000	0,0000	0,3333	0,7667	0,9333	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
(C) "Poder" da distribuição Normal – CV = 1,0											
Testes	Extensões das séries (anos)										
	50	100	200	300	400	500	600	700	800	900	1.000
KS	0,4333	0,9667	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
QQ	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
AD	0,4667	0,9667	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
CVM	0,4333	0,9333	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
(D) "Poder" da distribuição Normal – CV = 1,5											
Testes	Extensões das séries (anos)										
	50	100	200	300	400	500	600	700	800	900	1.000
KS	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
QQ	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
AD	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
CVM	0,9333	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Para séries com CV = 0,1 (Tabela 2A), os testes KS, AD e CVM apresentam “poder” nulo para todas as extensões de série, apresentando probabilidade nula de rejeitar a distribuição Normal quando ela é falsa. De um modo geral, observa-se que o teste com o maior valor de “poder” é o Qui-Quadrado. Para séries com CV = 0,5 (Tabela 2B), novamente, o teste Qui-Quadrado é o mais

“poderoso”. Depois, na sequência, têm-se os testes AD, KS e CVM.

Para séries com CV = 1,0 (Tabela 2C), repete-se o observado no caso anterior: o teste QQ é o mais “poderoso”. Na sequência, têm-se os testes AD, KS e CVM. Para séries com CV = 1,5 (Tabela 2D), todos os testes apresentam o mesmo “poder” (1,000) a partir de 100

dados. Assim, todos os testes rejeitariam a distribuição Normal quando ela é falsa.

Nota-se, para todos os casos, que o “poder” dos testes aumenta com o aumento do coeficiente de variação e da extensão da série, tendendo a um mesmo valor – 1,000.

Na Tabela 3 é mostrado o “Poder” dos testes de aderência para a distribuição Lognormal, segundo os valores de CV.

Tabela 3 – “Poder” da distribuição Lognormal

(A) "Poder" da distribuição Lognormal – CV = 0,1											
Testes	Extensões das séries (anos)										
	50	100	200	300	400	500	600	700	800	900	1.000
KS	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
QQ	0,0333	0,0333	0,0667	0,0667	0,0333	0,0333	0,0333	0,0667	0,0667	0,1333	0,1667
AD	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
CVM	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
(B) "Poder" da distribuição Lognormal – CV = 0,5											
Testes	Extensões das séries (anos)										
	50	100	200	300	400	500	600	700	800	900	1.000
KS	0,0000	0,0000	0,0000	0,0000	0,0667	0,2000	0,1000	0,2333	0,4667	0,4000	0,6333
QQ	0,0000	0,1667	0,3000	0,4000	0,1667	0,3667	0,3667	0,5333	0,7333	0,8333	0,7333
AD	0,0000	0,0000	0,0000	0,0000	0,1000	0,2333	0,2333	0,4667	0,4333	0,6667	0,7667
CVM	0,0000	0,0000	0,0000	0,0000	0,0333	0,1000	0,1333	0,1667	0,3000	0,4000	0,5333
(C) "Poder" da distribuição Lognormal – CV = 1,0											
Testes	Extensões das séries (anos)										
	50	100	200	300	400	500	600	700	800	900	1.000
KS	0,0000	0,1000	0,4333	0,8000	0,9000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
QQ	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
AD	0,0000	0,1000	0,6333	0,9000	0,9667	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
CVM	0,0000	0,0333	0,5333	0,7667	0,9667	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
(D) "Poder" da distribuição Lognormal – CV = 1,5											
Testes	Extensões das séries (anos)										
	50	100	200	300	400	500	600	700	800	900	1.000
KS	0,0667	0,2667	0,8333	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
QQ	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
AD	0,0667	0,4000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
CVM	0,0333	0,2667	0,9333	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Para séries com CV = 0,1 (Tabela 3A), os testes KS, AD e CVM apresentam “poder” nulo para todas as extensões de séries. Logo, nenhum deles rejeitará a distribuição Lognormal quando ela for falsa. O teste Qui-Quadrado é o mais “poderoso”, ainda que este “poder” seja baixo. Para séries com CV = 0,5 (Tabela 3B), o teste é o mais “poderoso” é o Qui-Quadrado, seguido, nesta ordem, pelo AD, KS e CVM.

Para séries com CV = 1,0 (Tabela 3C) assim como no caso anterior, os testes mais “poderosos”, nesta ordem, são:

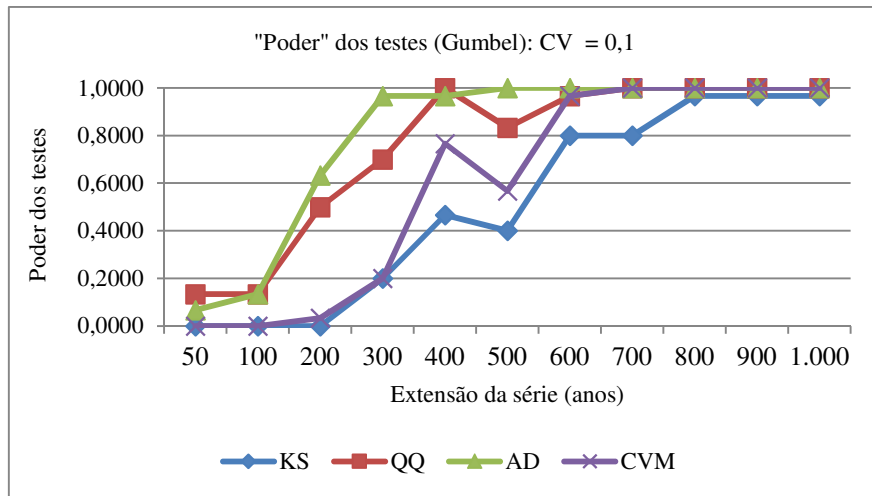
QQ, AD, KS e CVM. Para séries com CV = 1,5 (Tabela 3D), de novo, nesta ordem, os testes com maior “poder” são: QQ, AD, KS e CVM.

Nota-se, para todos os casos, que o “poder” dos testes aumenta com o aumento do coeficiente de variação e da extensão da série, tendendo a um mesmo valor – 1,000.

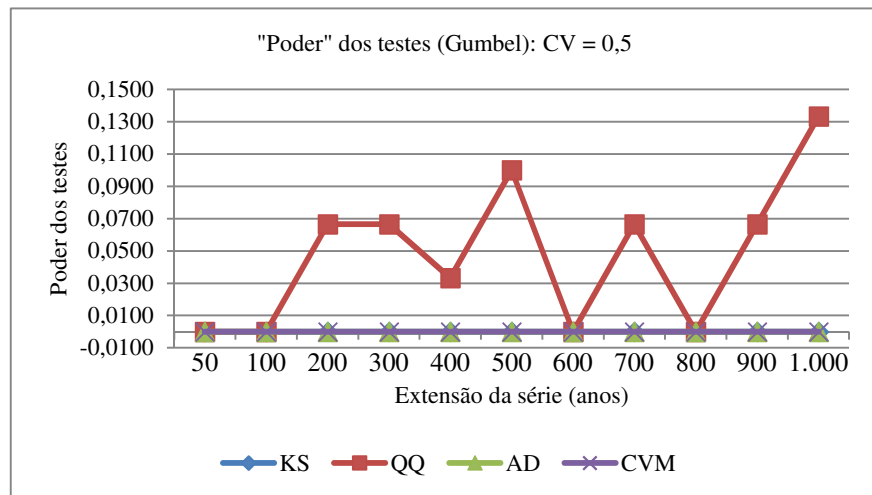
Na Figura 1 é mostrado o “Poder” dos testes de aderência para a distribuição Gumbel, segundo os valores de CV.

Figura 1 – “Poder” dos testes de aderência para a distribuição Gumbel: (A) CV = 0,1; (B) CV = 0,5; (C) CV = 1,0 e (D) CV = 1,5.

A.



B.



C.

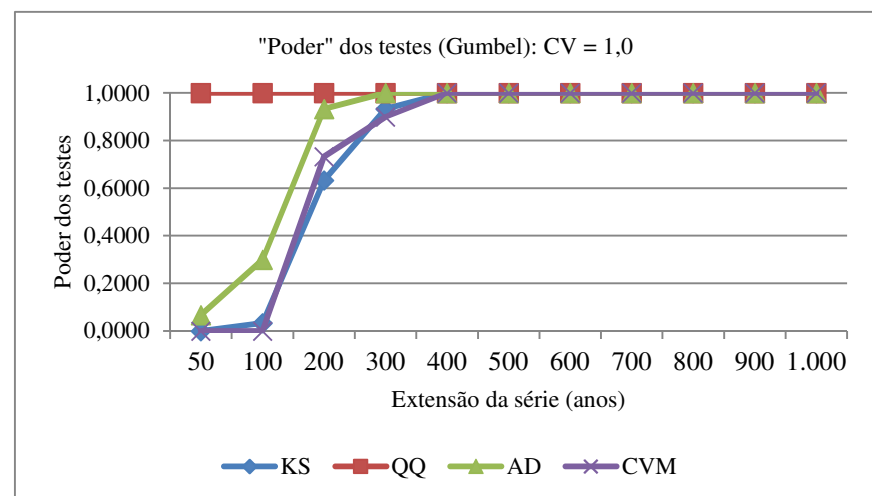
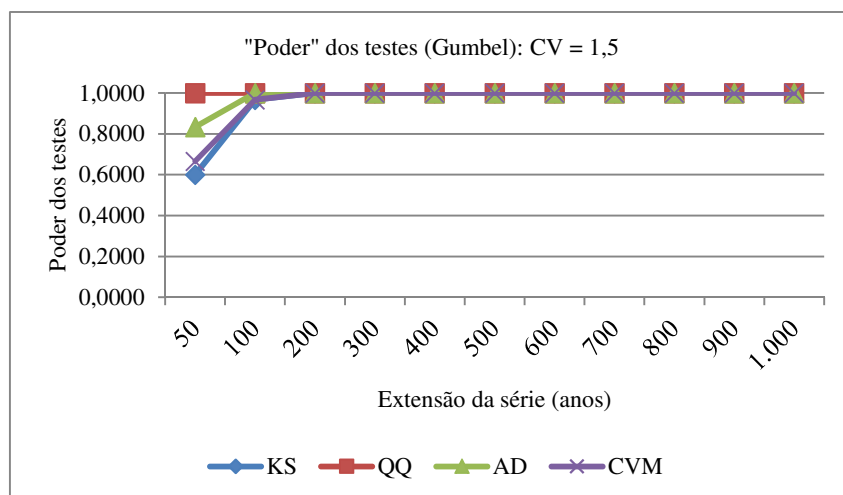


Figura 1 – “Poder” dos testes de aderência para a distribuição Gumbel: (A) CV = 0,1; (B) CV = 0,5; (C) CV = 1,0 e (D) CV = 1,5 (cont.) D.



Para séries com CV = 0,1 (Figura 1A), o teste mais “poderoso” é o Anderson-Darling. Em seguida, tem-se o QQ, CVM e KS. Para séries com CV = 0,5 (Figura 1B) os testes KS, AD e CVM apresentam “poder” nulo para todas as extensões de séries. Assim, nenhum deles rejeitará a distribuição Gumbel quando ela for falsa. O teste Qui-Quadrado é o mais “poderoso”, ainda que este “poder” seja baixo.

Para séries com CV = 1,0 (Figura 1C) os testes mais poderosos são, na ordem: QQ, AD, KS e CVM. Para séries com CV = 1,5 (Figura 1D), os testes mais “poderosos” são, na ordem: QQ, AD, CVM e KS.

Nota-se, para todos os casos, que o “poder” dos testes aumenta com o aumento do coeficiente de variação e da extensão da série, tendendo a um mesmo valor – 1,000.

3.3 Valores críticos

Admite-se o erro máximo entre os valores críticos calculados das distribuições falsas e os calculados da distribuição verdadeira igual a 5%. Calcularam-se os valores críticos só para séries com extensões de até 200 anos de dados.

Cv = 0,1

Para todos os períodos de retorno (5, 10, 100, 1.000 e 10.000 anos), apenas os valores críticos calculados pela distribuição Lognormal e pela distribuição Normal apresentam erros pequenos em relação aos valores calculados pela distribuição Gama (verdadeira). Logo, estas duas distribuições podem substituir a distribuição Gama no cálculo de valores críticos.

Cv = 0,5

Para períodos de retorno iguais a 5 e 10 anos, apenas os valores críticos calculados pela distribuição Lognormal e os calculados pela distribuição Normal apresentam erros pequenos em relação aos valores calculados pela distribuição Gama. Para os demais períodos de retorno (100, 1.000 e 10.000 anos), os erros aumentam e inviabilizam o uso destas duas distribuições.

Cv = 1,0

Para períodos de retorno iguais a 10 anos, apenas a distribuição Normal pode substituir a Gama no cálculo dos valores críticos. Para períodos de retorno iguais a 5 anos, indica-se a distribuição Lognormal, porém só para séries com 200 dados. Atualmente, nenhuma série histórica do mundo possui esta extensão.

Para períodos de retorno iguais a 100, 1.000 e 10.000 anos, todas as distribuições apresentaram erros consideráveis.

Cv = 1,5

Para períodos de retorno iguais a 10 anos indica-se só a distribuição Normal em substituição à distribuição Gama no cálculo dos valores críticos. Mas, só para séries de até 100 anos. As distribuições Lognormal e Gumbel apresentam erros sempre superiores ao máximo, para todos os períodos de retorno e para todas as extensões.

3.4 Testes de aderência nos postos pluviométricos

Os resultados das séries sintéticas que interessam aos postos pluviométricos são os referentes a coeficientes de variação até o valor de 0,5 e a extensões de até 100 dados, pois estas são as características dos postos estudados.

Para as séries dos postos, o teste do Qui-Quadrado é o mais confiável porque apresenta erro tipo I mais próximo da significância. O “poder” do Qui-Quadrado, porém, é nulo para a distribuição Gumbel, não sendo confiável nenhum dos resultados de não rejeição para esta distribuição. Os resultados do Qui-Quadrado para a Gumbel foram, em todos os postos, de rejeição nas Máximas Anuais e de não rejeição nos Totais Anuais. Logo, consideram-se os resultados do Qui-Quadrado não confiáveis para a distribuição Gumbel.

Para a Lognormal, o Qui-Quadrado apresenta “poder” baixo (0,1667). Isto, apesar de tornar os seus resultados de não rejeição sujeitos a grandes erros, faz com que seja o único a ser utilizado. Para a Normal, o “poder” do Qui-Quadrado é maior (0,4667), sendo este, pois, o teste mais confiável para os casos de não rejeição.

Nas Tabelas 4 e 5 são mostrados, respectivamente, os resultados do teste do QQ para os dados de precipitações

máximas anuais e de precipitações totais anuais, dos postos pluviométricos.

Tabela 4 – Resultados do teste Qui-Quadrado para as máximas anuais dos postos pluviométricos.

Postos Pluviométricos	Gama	Normal	Lognormal	Gumbel
Fortaleza/Pici	Não Rejeita		Não Rejeita	
Farias Brito				
Frecheirinha	Rejeita		Rejeita	
Graça				
Guaraciaba do Norte	Não Rejeita	Rejeita		
Itapajé				
Milagres	Rejeita			
Hidrolândia			Não Rejeita	
Ibiapina	Não Rejeita			
Icó		Não Rejeita		
Iguatu				
Jaguaruana	Rejeita			
Jardim		Rejeita	Rejeita	Rejeita
Limoeiro do Norte				
Maranguape	Não Rejeita			
Missão Velha		Não Rejeita	Não Rejeita	
Mucambo				
Novo Oriente	Rejeita			
Palhano		Rejeita	Rejeita	
Parambu	Não Rejeita		Não Rejeita	
Paramoti	Rejeita		Rejeita	
Porteiras		Não Rejeita		
Senador Pompeu	Não Rejeita	Rejeita	Não Rejeita	
Tamboril				

Observa-se, na Tabela 4, por exemplo, que as distribuições Gama e Lognormal descrevem as precipitações máximas anuais dos postos de Guaraciaba do

Norte e Fortaleza/Pici. Já a distribuição Normal não adere às máximas anuais da maioria dos postos.

Tabela 5 – Resultados do teste Qui-Quadrado para os totais anuais dos postos pluviométricos.

Postos Pluviométricos	Gama	Normal	Lognormal	Gumbel
Fortaleza/Pici	Não Rejeita		Não Rejeita	
Farias Brito	Rejeita	Rejeita	Rejeita	Não Rejeita
Frecheirinha				
Graça	Não Rejeita	Não Rejeita	Não Rejeita	
Guaraciaba do Norte			Rejeita	
Itapajé		Rejeita		
Milagres		Não Rejeita	Não Rejeita	
Hidrolândia				
Ibiapina	Rejeita	Rejeita	Rejeita	
Icó				
Iguatu	Não Rejeita	Não Rejeita	Não Rejeita	
Jaguaruana	Rejeita		Rejeita	
Jardim		Rejeita	Não Rejeita	
Limoeiro do Norte	Não Rejeita	Não Rejeita		
Maranguape			Rejeita	Não Rejeita
Missão Velha	Rejeita			
Mucambo				
Novo Oriente	Não Rejeita		Não Rejeita	
Palhano	Rejeita	Rejeita	Rejeita	
Parambu				
Paramoti	Não Rejeita		Não Rejeita	
Porteiras	Rejeita		Rejeita	
Senador Pompeu		Não Rejeita		
Tamboril	Não Rejeita	Rejeita	Não Rejeita	

Na Tabela 5, posto de Iguatu, por exemplo, as distribuições Gama, Normal e Log-normal descrevem o comportamento das precipitações totais anuais.

3.5 Valores críticos nos postos pluviométricos

Os valores críticos de precipitação foram calculados para os períodos de retorno iguais a 5, 10, 100, 1.000 e 10.000 anos.

O cálculo dos valores críticos de precipitação, tanto para máximas anuais quanto para totais anuais, independem da distribuição utilizada, para períodos de retorno de 5 e 10 anos. Para períodos de retorno de 100, 1.000 e 10.000 anos, utiliza-se a distribuição Gumbel para as máximas anuais e a distribuição Normal para os totais anuais. Assim, define-se qual a distribuição de probabilidade a ser utilizada em futuras estimativas de precipitação.

CONCLUSÕES

Quando o coeficiente de variação é igual a 0,1 e 0,5; o teste mais confiável, quanto ao erro tipo I, é o Qui-Quadrado. Quando o coeficiente de variação é igual a 1,0; os testes mais confiáveis são Kolmogorov-Smirnov, Cramer-von Mises e Anderson-Darling. E para coeficiente de variação igual a 1,5; os testes mais confiáveis são Kolmogorov-Smirnov e Cramer-von Mises.

De um modo geral, para todas as distribuições estudadas, os “poderes” de todos os testes crescem com o aumento da extensão e do valor do coeficiente de variação das séries sintéticas e, por isso, convergem a um mesmo valor – 1,000.

Para séries com coeficiente de variação igual a 0,1; o teste do Qui-Quadrado é o mais “poderoso”. Para a distribuição Gumbel, o teste mais “poderoso” é o de Anderson-Darling.

Para o cálculo dos valores críticos, só as distribuições Normal e Lognormal podem substituir a distribuição Gama, em todos os períodos de retorno.

Para séries com coeficiente de variação igual a 0,5; o teste do Qui-Quadrado é o mais “poderoso” para as distribuições Normal, Lognormal e Gumbel.

Apenas os valores críticos calculados pelas distribuições Normal e Lognormal são próximos aos valores da distribuição Gama. Porém, só para períodos de retorno de até 10 anos.

Para séries com coeficiente de variação igual a 1,0; novamente, o teste do Qui-Quadrado é o mais “poderoso” para as distribuições Normal, Lognormal e Gumbel.

Apenas a distribuição Normal substitui a distribuição Gama no cálculo dos valores críticos, mas somente para períodos de retorno iguais a 10 anos.

Para séries com coeficiente de variação igual a 1,5; nos ajustes com a distribuição Normal, os testes Qui-Quadrado, Kolmogorov-Smirnov e Anderson-Darling são igualmente “poderosos”. Para as distribuições Lognormal e Gumbel, o teste mais “poderoso” é o Qui-Quadrado.

Só a distribuição Normal gera valores críticos próximos aos valores da distribuição Gama. Porém, apenas para períodos de retorno de 10 anos.

Para os postos pluviométricos estudados, o teste mais confiável é Qui-Quadrado.

O cálculo dos valores críticos de precipitação, tanto para máximas anuais quanto para totais anuais, independe da distribuição utilizada, para períodos de retorno de 5 e 10 anos. Para períodos de retorno de 100, 1.000 e 10.000 anos, usa-se a distribuição Gumbel para as máximas anuais e a Normal para os totais anuais.

REFERÊNCIAS

AGÊNCIA NACIONAL DE ÁGUAS- ANA. Superintendência de Administração da Rede Hidrometeorológica. **Evolução da rede hidrometeorológica nacional**. Brasília: ANA, 2007.

AGÊNCIA NACIONAL DE ÁGUAS – ANA.

Disponível em: <<http://www.ana.gov.br>>. Acesso em: 10 set. 2012.

ANDRADE, A. R. S. de; PAIXÃO, F. J. R.; AZEVEDO, C. A. V. de; GOUVEIA, J. P. G. de; OLIVEIRA JÚNIOR, J. A. S. de. Estudo do comportamento de períodos secos e chuvosos no município de Garanhuns, PE, para fins de planejamento agrícola. **Pesquisa aplicada e Agrotecnologia**. v. 1, n. 1 set./dez. 2008.

ARAÚJO, E. M.; ARAÚJO, E. M.; OLIVEIRA, J. B. de; SILVA, M. G.; VIANA, P. C.; ALVES, A. S. Análise da aderência de distribuições de probabilidade aos dados de temperatura máxima e mínima do ar em Iguatu, CE, **Revista Caatinga**, Mossoró, v. 23, n. 3, p. 104-109, jul./set. 2010.

BEIJO, L. A.; MUNIZ, J. A.; CASTRO NETO, P. Tempo de retorno das precipitações máximas em Lavras (MG) pela distribuição de valores extremos do tipo I, **Ciência. agrotec.**, Lavras, v. 29, n. 3, p. 657-667, maio/jun. 2005.

DEKA, S.; BORAH, M.; KAKATY, S. C. Distributions of Annual Maximum Rainfall Series of North-East India. **European Water**, n. 27/28, p. 3-14, 2009.

DÉRY, S. J.; WOOD, E. F. Decreasing river discharge in northern Canada. **Geophysical research letters**, v.32, L10401. 2005. <https://doi.org/10.1029/2005GL022845>.

FARIAS, J. A. M. **Métodos de geração de vazões mensais e suas influências sobre a curva de garantia versus vazão regularizada em reservatórios do Estado do Ceará**. 2003, 115p. Dissertação (Mestrado), Universidade Federal do Ceará, Fortaleza, 2003.

HARTMANN, M.; MOALA, F. A.; MENDONÇA, M. A. Estudo das precipitações máximas anuais em Presidente Prudente, **Revista Brasileira de Meteorologia**, v. 26, n. 4, p. 561-568, 2011. <https://doi.org/10.1590/S0102-77862011000400006>.

JOHNSON, R. A. **Probability and statistics for engineers**. 6 ed. The United States: Prentice-Hall, chap. 12, p. 312-329, 2000.

- KOUTSOYIANNIS, D.; BALOUTSOS, G. **Analysis of a Long Record of Annual Maximum Rainfall in Athens, Greece, and Design Rainfall Inferences**. Natural Hazards, cap. 29, p. 29-48, 2000.
- LONGO, A. J.; SAMPAIO, S. C.; QUEIROZ, M. M. F. de.; SUSZEK, M. Uso das distribuições gama e log-normal na estimativa de pre-cipitação provável quinzenal, **Varia Scientia**. v. 06, n. 11, p. 107-118, 2006.
- McMAHON, T. A.; VOGEL, R. M.; PEEL, M. C.; PEGRAM, G. G. S. Global streamflows – Part 1: Characteristics of annual streamflows, **Journal of Hydrology**, n. 347, p. 243-259, 2007a. <https://doi.org/10.1016/j.jhydrol.2007.09.002>.
- McMAHON, T. A.; VOGEL, R. M.; PEEL, M. C.; PEGRAM, G. G. S. Global streamflows – Part 3: Country and climate zone characteristics. **Journal of Hydrology**, v. 347, p. 272-291, 2007b. <https://doi.org/10.1016/j.jhydrol.2007.09.013>. <https://doi.org/10.1016/j.jhydrol.2007.09.002>
- NIJSSEN, B.; O'DONNELL, G. M.; HAMLET, A. F.; LETTENMAIER, D. P. Hydrologic sensitivity of global rivers to climate change, **Climatic Change**, v. 50, p. 143-175, July 2001. <https://doi.org/10.1023/A:1010616428763>.
- RAHMAN, M.; PEARSON, L. M.; HEIEN, H. C. A Modified Anderson-Darling Test for Uniformity. **Bulletin of the Malaysian Mathematical Sciences Society**, v. 29, p. 11-16, 2006.
- RAZALI, N. M.; WAH, Y. B. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. **Journal of Statistical Modeling and Analytics**, v. 2, n.1, p.21-33, 2011.
- RIBEIRO, B. T.; AVANZI, J. C.; MELLO, C. R. de; LIMA, J. M. de; SILVA, M. L. N. Comparação de distribuições de probabilidade e estimativa da precipitação provável para a região de Barbacena, MG. **Ciência Agrotecnologia**, Lavras, v. 31, n. 5, set./out. 2007.
- SAMPAIO, S. C.; QUEIROZ, M. M. F. de; FRIGO, E. P.; LONGO, A. J.; SUSZEK, M. Estimativa e distribuição de precipitações decendiais para o Estado do Paraná. **Irriga**, Botucatu, v. 12, n. 1, p. 38-53, jan./mar. 2007.
- SANTOS, M. J. dos; ARAÚJO, L. E.; OLIVEIRA, E. M.; SILVA, B. B. Seca, precipitação e captação de água de chuva no semi-árido de Sergipe. **Engenharia Ambiental**, Espírito Santo do Pinhal, v. 6, n. 1, p. 055-073, jan./abr. 2009.
- SILVA, C. L. da. Análise estatística das características de vazão do córrego Capetinga. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.7, n.2, p.311-317, 2003. <https://doi.org/10.1590/S1415-43662003000200021>.
- SILVA, V. P. R. da; PEREIRA, E. R. R.; AZEVEDO, P. V. de; SOUSA, F. A. S. de; SOUSA, I. F. de. Análise da pluviometria e dias chuvosos na região Nordeste do Brasil. **Revista Brasileira de Engenharia Agrícola e Ambiental**. v. 15, n. 2, p.131-138, 2011. <https://doi.org/10.1590/S1415-43662011000200004>.
- SILVINO, A. N. de O.; SILVEIRA, A.; MUSIS, C. R. de.; WYREPKOWSKI, C. C.; CONCEIÇÃO, F. T. da. Determinação de vazões extremas para diversos períodos de retorno para o Rio Paraguai utilizando métodos estatísticos. **Geociências**, São Paulo, v. 26, n. 4, p. 369-378, 2007.
- TUCCI, Carlos E. M. **Hidrologia: ciência e aplicação**. 4. ed. Porto Alegre: UFRGS, 2009.
- VOGEL, R. M.; WILSON, I.; DALY, C. Regional regression models of annual streamflow for the united states. **Journal of irrigation and drainage engineering**. May/June 1999. [https://doi.org/10.1061/\(ASCE\)0733-9437\(1999\)125:3\(148\)](https://doi.org/10.1061/(ASCE)0733-9437(1999)125:3(148)).
- WAGENER, T.; KOLLAT, J. Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox. **Environmental Modelling & Software** v. 22, p. 1021-1033. July 2007. <https://doi.org/10.1016/j.envsoft.2006.06.017>.
- YEVJEVICH, V. M. Probability and Statistics in Hydrology. **Water Resources Publications**, Colorado, chap. 7, p. 118-168, 1972.