

USO DE DISTRIBUIÇÕES DE PROBABILIDADE NA MODELAGEM DE INTENSIDADE DE SECAS OCORRIDAS EM LARANJEIRAS DO SUL, PR

Jailson de Araujo RODRIGUES¹
Ana Paula Coelho Madeira SILVA²
Jaime dos SANTOS FILHO³

- **RESUMO:** As secas são fenômenos naturais recorrentes que afetam grande parte do planeta, provocando impactos negativos nos diversos ramos das atividades agrícolas, industrial e urbana, refletindo diretamente na qualidade de vida da população. Dessa forma, diferentes modelos têm sido propostos na tentativa de um melhor entendimento e previsão desse fenômeno. Neste trabalho, foi avaliado o potencial das distribuições exponencial, gama, log-normal, Pareto e Weibull na modelagem da intensidade de secas ocorridas em Laranjeiras do Sul no Estado do Paraná. Foram utilizados dados de secas referentes a uma série histórica de 30 anos. A aderência das distribuições foi verificada utilizando o teste de Kolmogorov-Sminorv. Com exceção da distribuição Pareto, todas as distribuições analisadas apresentaram aderência. Entretanto, os resultados obtidos indicaram que o modelo gama ajustou-se mais adequadamente ao conjunto de dados observados.
- **PALAVRAS-CHAVE:** Critério de informação de Akaike; distribuição exponencial; distribuição gama; distribuição log-normal; distribuição Weibull; índice padronizado de precipitação.

1 Introdução

As secas são fenômenos naturais recorrentes que resultam da persistência de valores de precipitação abaixo da média. Contrariamente aos demais desastres

¹Instituto Federal da Bahia - IFBA, Campus de Feira de Santana, CEP: 44096-486, Bahia, Brasil. E-mail: *jailsondearaujo@yahoo.com.br*

²Universidade Federal de São João del-Rei - UFSJ, Campus de Sete Lagoas, CEP: 35701-970 Minas Gerais, Brasil. E-mail: *anapaula@ufsj.edu.br*

³Instituto Federal da Bahia - IFBA, Campus de Vitória da Conquista, CEP: 37200-000, Bahia, Brasil. E-mail: *jaime@ifba.edu.br*

naturais, que geralmente atuam de forma rápida e com impactos imediatos, a seca é um desastre natural de origem meteorológica e climatológica mais complexo, que afeta maior número de pessoas, com longos períodos de duração.

Medir dados de seca é muito importante em diversos contextos, tais como produtividade de culturas agrícolas, manejo dos recursos hídricos e avaliação ambiental. A obtenção da correta distribuição para intensidade de seca é relevante no planejamento agrícola no que diz respeito à instalação de culturas. Além da influência na agricultura, secas muito intensas afetam o nível de água dos mananciais e dos reservatórios das usinas hidrelétricas, trazendo problemas para a geração de energia elétrica e no abastecimento urbano.

A modelagem probabilística se destacou como um dos mecanismos eficazes na análise dos eventos hidrológicos. Por exemplo, Rodrigues et al. (2013) utilizou os modelos gama, Weibull e log-normal para estudar a distribuição mensal das chuvas ocorridas na cidade de Bento Gonçalves em um período de onze anos. Os resultados indicaram que o modelo gama ajustou-se mais adequadamente a pluviometria da região. Silva et al. (2007) fez uma análise comparativa entre as distribuições gama, Weibull, normal, exponencial e log-normal no ajuste de um conjunto de dados reais de precipitações pluviométricas ocorridas em Santa Maria no Rio Grande do Sul. Os modelos gama e Weibull apresentaram os melhores ajustes com relação aos dados analisados. Longo et al. (2006) avaliaram o uso das distribuições gama e log-normal na estimativa de precipitações pluviais quinzenais no Estado do Paraná. Os resultados mostraram que a distribuição gama é a mais adequada para descrever as condições pluviométricas do Estado.

Ao longo dos anos diferentes distribuições de probabilidade têm sido propostas para descrever a intensidade, o período e a magnitude de secas. Por exemplo, os modelos exponencial e gama foram utilizados por Shiau e Morrades (2009) para descrever respectivamente, a duração e a intensidade de secas ocorridas no sul e no norte do Irã. Nadarajah (2008) introduziu uma generalização do modelo Pareto para estudar dados de secas ocorridas nas oito regiões climáticas do Estado norte-americano de Nebraska, a qualidade de ajuste do modelo foi verificada utilizando o gráficos de probabilidade. Esse mesmo conjunto de dados foi analisado por Nadarajah e Gupta (2007) utilizando uma generalização do modelo gama, os resultados indicaram que a nova distribuição fornecia um ajuste melhor que o obtido pelo modelo gama usual. Shiau et al. (2007) empregou a distribuição gama para analisar a intensidade de secas ocorridas em Yellow River no Norte da China. A distribuição gama também foi empregada por Shiau (2006) para descrever a intensidade de secas ocorridas em Taiwan e Yang e Nadarajah (2006) utilizaram a distribuição exponencial para descrever o período de seca e as distribuições Weibull e log-normal para descrever sua intensidade.

O objetivo deste trabalho é avaliar as distribuições exponencial, gama, log-normal, Pareto e Weibull na modelagem da intensidade de secas ocorridas em Laranjeiras do Sul no Estado do Paraná. A estimação dos parâmetros foi feita via método da máxima verossimilhança. As respectivas aderências dos modelos foi verificada pelo teste de Kolmogorov-Smirnov ao nível de 5% de significância. A

escolha do modelo que apresentou melhor ajuste foi feita via Critério de Informação de Akaike corrigido, razão de verossimilhanças e gráficos de probabilidade.

2 Materiais e métodos

2.1 Índice padronizado de precipitação (SPI)

Na literatura, é possível encontrar vários índices que permitem determinar a ocorrência de seca. O mais utilizado, no Brasil, é o Índice Padronizado de Precipitação (SPI). Esse índice quantifica o déficit ou o excesso de precipitação em diferentes escalas de tempo. Essa característica torna o SPI uma valiosa ferramenta para todos os estudos de disponibilidade hídrica. A escala temporal mais analisada é a mensal (SPI-1 mês). À medida que a escala temporal aumenta, a resposta do SPI tornam-se mais lenta a mudanças na precipitação. O SPI é calculado com base na distribuição de probabilidades da precipitação do período escolhido. Desse modo, os valores do índice SPI obtidos dependem da função de distribuição escolhida (geralmente a distribuição gama), dos valores da amostra a partir dos quais são determinados os parâmetros da distribuição e, também, do método de estimação desses parâmetros.

O evento seca começa quando o SPI torna-se negativo e termina quando este volta a apresentar valores positivos. Dentro de sua escala, magnitudes menores ou iguais a -2 indicam seca extrema. Na Tabela 1 é apresentada essa classificação que foi desenvolvida por McKee et al. (1995).

Tabela 1. Classificação da severidade da seca.

Valores do SPI	Categoria de seca
0,00 a -0,99	Seca ligeira
-1,00 a -1,49	Seca moderada
-1,50 a -1,99	Seca severa
$\leq -2,00$	Seca extrema

Considerando que a seca ocorre quando o valor do SPI é menor que zero, o período de seca é a somatória dos meses consecutivos em que o SPI é negativo. Dessa forma, se P representa o período de seca, a intensidade da seca, denotada por S , é a soma dos valores do SPI no período de seca. Por conveniência, assumiremos a intensidade da seca como uma grandeza positiva e assim, pode-se escrever:

$$S = - \sum_{i=0}^P SPI_i$$

Os dados de SPI explorados neste trabalho foram coletados na estação meteorológica X2552009 com coordenadas geográficas -25.41667 de latitude, -52.41667 de longitude e 1045m de altitude, localizada na cidade de Laranjeiras do Sul no Estado do Paraná. A base de dados corresponde a uma série histórica de medições mensais do SPI no período de janeiro de 1976 até dezembro de 2005. A Figura 1 apresenta a série temporal dessas medições de SPI.

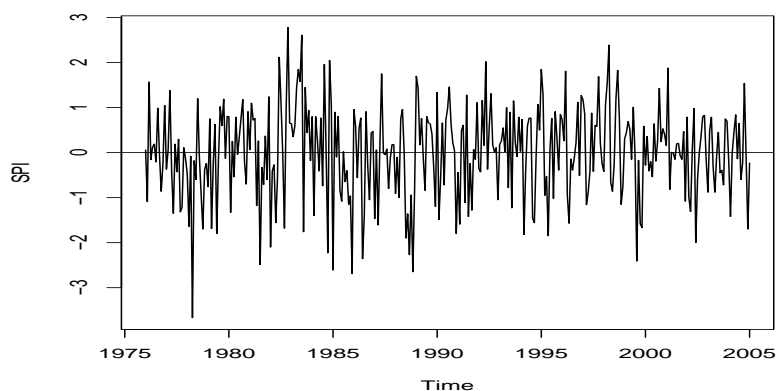


Figura 1 - Série temporal dos dados de SPI.

2.2 Modelos

Modelo exponencial: Uma variável aleatória tem distribuição exponencial quando sua função densidade de probabilidade (fdp) para $x > 0$ é dada por:

$$f_E(x) = \beta \exp(-\beta x) \quad (1)$$

sendo que $\beta > 0$ é um parâmetro de escala. A função de distribuição (fd) correspondente a (1) é

$$F_E(x) = 1 - \exp(-\beta x).$$

Se X_1, \dots, X_n é uma amostra aleatória de (1), a estimativa de máxima verossimilhança para o parâmetro β é dada por:

$$\hat{\beta} = \frac{n}{\sum_{i=0} x_i}.$$

Modelo gama: Uma variável aleatória tem distribuição gama quando sua fdp para $x > 0$ é dada por:

$$f_{GA}(x) = \frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} \exp(-\beta x) \quad (2)$$

sendo que $\alpha > 0$ é um parâmetro de forma, $\beta > 0$ é um parâmetro de escala e $\Gamma(\cdot)$ representa a função gama,

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} \exp(-t) dt.$$

A fd correspondente a densidade gama (2) pode ser expressa como:

$$F_{GA}(x) = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)}$$

sendo que $\gamma(\alpha, \cdot)$ representa a função gama incompleta,

$$\gamma(\alpha, x) = \int_0^x t^{\alpha-1} \exp(-t) dt.$$

O modelo exponencial apresentado anteriormente é um caso particular da distribuição gama quando $\alpha = 1$.

Se X_1, \dots, X_n é uma amostra aleatória de (2), as estimativas de máxima verossimilhança $\hat{\alpha}$ e $\hat{\beta}$ para os parâmetros α e β podem ser obtidas resolvendo o sistema de equações não-lineares,

$$\frac{n\hat{\alpha}}{\hat{\beta}} - \sum_{i=0}^n x_i = 0$$

$$n \log n + n \log \hat{\alpha} - n \log \left(\sum_{i=0}^n x_i \right) - n\psi(\hat{\alpha}) + \sum_{i=0}^n \log x_i = 0$$

sendo que $\psi(\cdot)$ é a função digama,

$$\psi(x) = \frac{d\Gamma(x)}{dx}.$$

Modelo log-normal: Uma variável aleatória tem distribuição log-normal quando sua fdp para $x > 0$ é dada por:

$$f_{LN}(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\log x - \mu)^2}{2\sigma^2}\right] \quad (3)$$

sendo que $-\infty < \mu < +\infty$ é um parâmetro de escala e $\sigma > 0$ é um parâmetro de forma. A fd correspondente a (3) é

$$F_{LN}(x) = \Phi\left(\frac{\log x - \mu}{\sigma}\right) \quad (4)$$

tal que $\Phi(\cdot)$ representa a fd da distribuição normal padrão.

Se X_1, \dots, X_n é uma amostra aleatória de (3), as estimativas de máxima verossimilhança para os parâmetros μ e σ^2 são:

$$\hat{\mu} = \frac{1}{n} \sum_{i=0}^n \log x_i \quad (5)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=0}^n \left(\log x_i - \frac{1}{n} \sum_{i=0}^n \log x_i \right)^2. \quad (6)$$

Modelo Pareto: Uma variável aleatória tem distribuição Pareto quando sua fdp para $x > 0$ é dada por:

$$f_{PA}(x) = \alpha \beta^\alpha x^{-(\alpha+1)} \quad (7)$$

sendo que β é um parâmetro de escala e α é um parâmetro de forma. A fd correspondente a (7) é

$$F_{PA}(x) = 1 - \left(\frac{\beta}{x} \right)^\alpha$$

em que $x \geq \beta$.

Se X_1, \dots, X_n é uma amostra aleatória de (3), as estimativas de máxima verossimilhança para os parâmetros α e β são:

$$\hat{\beta} = \min \{x_i\}_{i=1}^n$$

$$\hat{\alpha} = n \left[\sum_{i=0}^n \log \left(\frac{x_i}{\hat{\beta}} \right) \right]^{-1}.$$

Modelo Weibull: Uma variável aleatória tem distribuição Weibull quando sua fdp para $x > 0$ é dada por:

$$f_{WE}(x) = \theta \lambda (\lambda x)^{\theta-1} \exp \left[-(\lambda x)^\theta \right] \quad (8)$$

sendo que $\lambda > 0$ é um parâmetro de escala e $\theta > 0$ é um parâmetro de forma. A fd correspondente a (8) é

$$F_{WE}(x) = 1 - \exp \left[-(\lambda x)^\theta \right].$$

Se X_1, \dots, X_n é uma amostra aleatória de (8), as estimativas de máxima verossimilhança $\hat{\theta}$ e $\hat{\lambda}$ para os parâmetros θ e λ podem ser obtidas resolvendo o sistema de equações não-lineares,

$$\hat{\lambda} = \left(\frac{1}{n} \sum_{i=0}^n x_i^{\hat{\theta}} \right)^{-1/\hat{\theta}}$$

$$\hat{\theta} = \left[\left(\sum_{i=0}^n x_i^{\hat{\theta}} \log x_i \right) \left(\sum_{i=0}^n x_i^{\hat{\theta}} \right)^{-1} - \frac{1}{n} \sum_{i=0}^n \log x_i \right]^{-1}.$$

Em Johnson et al. (1995) são apresentadas todas as distribuições mencionadas juntamente com suas principais propriedades.

2.3 Teste de aderência de Kolmogorov-Smirnov

O teste estatístico de Kolmogorov-Smirnov (KS-teste) é um dos dispositivos mais utilizados para verificar a aderência de distribuições. Além de não depender do número de classes do agrupamento dos dados, esse teste oferece mais vantagens computacionais. O teste baseia-se na discrepância entre as distribuições, $D_n = \sqrt{n} \sup_x |F_n(x) - F_0(x)|$ em que $F_n(x)$ denota a distribuição teórica e $F_0(x)$ a distribuição ajustada. Essa estatística é usada para testar a hipótese nula $H_0 : F = F_0$ versus hipótese alternativa $H_1 : F \neq F_0$. O p-valor correspondente é dado por $P(K > d_n)$ sendo que d_n representa o valor observado de D_n e K é a variável aleatória especificada pela fd $P(K \leq x) = \sqrt{2\pi}/x \sum_{i=1}^{\infty} \exp\{-(2i-1)^2\pi^2/(8x^2)\}$ (KVAM, 2007). A hipótese nula é aceita quando o p-valor é maior que 0,05.

2.4 Gráficos de probabilidade

Os procedimentos gráficos são indicados, entre outras coisas, para verificar a qualidade de ajuste de modelos, eles mostram se dois conjuntos de dados provieram da mesma distribuição. Vários desses procedimentos gráficos são propostos na literatura especializada, um dos mais utilizados são os gráficos de probabilidade.

Os gráficos de probabilidade consistem em representar no plano cartesiano a probabilidade observada versus a probabilidade esperada. Segundo Chambers et al. (1983), a construção desses gráficos é feita a partir das seguintes etapas:

1. Dada uma amostra x_1, x_2, \dots, x_n , rearrumar a amostra em ordem crescente, $x_{(1)}, x_{(2)}, \dots, x_{(n)}$;
2. Determinar as probabilidades observadas $\tilde{P}_i = F(x_{(i)})$, $i = 1, 2, \dots, n$;
3. Determinar as probabilidades acumuladas $P_i = (i - 0,375)/(n - 0,25)$ correspondente ao valor $x_{(i)}$;
4. Plotar o gráfico de dispersão para os pares de pontos (\tilde{P}_i, P_i) .

Se a pressuposição sobre a distribuição está correta os pares de pontos estarão dispostos sobre uma linha reta ou próximos dela.

2.5 Critério de informação de Akaike

Ao se utilizar modelos probabilísticos, deve-se levar em consideração a sua condição de aproximação da realidade, ou seja, não há modelos probabilísticos

exatos, sempre ocorrerá perda de informação, característica inerente a presença da componente de aleatoriedade. Neste caso, é necessário selecionar o modelo que apresenta o melhor ajuste com relação ao conjunto de dados analisados. Tal escolha deve ser feita segundo princípios científicos (BURNHAM; ANDERSON, 2004). Dentre as metodologias empregadas para esse fim, uma das mais utilizadas é o critério de informação de Akaike (AIC). Utilizaremos como mecanismo de escolha uma correção desse critério proposta por Bozdogan (1987) chamada de critério de informação de Akaike corrigido (AICc). O AICc é baseado na teoria de decisão e é definido como a quantidade:

$$AICc = -2L + 2p + 2 \frac{p(p+1)}{n-p-1} \quad (9)$$

em que L representa o logaritmo do máximo da função de verossimilhança e p denota o número de parâmetros do modelo. De acordo com esse critério, o melhor dentre os modelos considerados na análise do problema é aquele que apresenta o menor valor de AICc.

Todas as análises foram realizadas utilizando o ambiente R: A Language and Environment for Statistical Computing (2012).

2.6 Razão de verossimilhanças maximizadas

Quando se busca a "correta" distribuição para um determinado conjunto de dados, podem ocorrer situações em que diferentes distribuições fornecem ajustes muito próximos. Nesse contexto, o método da razão de verossimilhanças generalizadas é um dos mais utilizados na literatura para fazer a discriminação entre duas distribuições (ELSHERPIENY et al., 2013).

Sejam X_1, X_2, \dots, X_n uma amostra aleatória de tamanho n obtida de uma das distribuições: exponencial, gama, log-normal, Pareto ou Weibull. Sob esses modelos, as funções de verossimilhança correspondentes à amostra aleatória observada são dadas, respectivamente, por:

$$\begin{aligned} L_E(\beta) &= \prod_{i=1}^n f_E(x_i; \beta) \\ L_{GA}(\alpha, \beta) &= \prod_{i=1}^n f_{GA}(x_i; \alpha, \beta) \\ L_{LN}(\mu, \sigma) &= \prod_{i=1}^n f_{LN}(x_i; \mu, \sigma) \\ L_{PA}(\alpha, \beta) &= \prod_{i=1}^n f_{PA}(x_i; \alpha, \beta) \\ L_{WE}(\theta, \lambda) &= \prod_{i=1}^n f_{WE}(x_i; \theta, \lambda) \end{aligned}$$

Considerando as funções de verossimilhança dos modelos analisados, pode-se definir o logaritmo da razão de verossimilhanças maximizadas como sendo:

$$T = \ln \left[\frac{L_{Modelo i}}{L_{Modelo j}} \right] \quad (10)$$

em que $L_{Modelo i}$ e $L_{Modelo j}$ representam os máximos das funções de verossimilhança dos modelos i e j respectivamente.

Com base no valor de T é possível realizar uma discriminação entre os modelos propostos (GUPTA; KUNDU, 2004; KUNDU et. al., 2005). Se $T > 0$ temos que o modelo i é mais adequado para descrever os dados, caso contrário opta-se pelo modelo j .

3 Resultados e discussão

A intensidade média das secas observadas na cidade de Laranjeiras do Sul foi de 1,606. Observou-se uma variabilidade muito grande nos dados, com medição mínima de 0,040, máxima de 7,00 e erro padrão de 1,325.

Na Tabela 2 são exibidas as estimativas de máxima verossimilhança encontradas nos ajustes das distribuições exponencial, gama, log-normal, Pareto e Weibull.

Tabela 1 - Estimativas dos parâmetros das distribuições

Modelos	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\theta}$	$\hat{\lambda}$
Exponencial	...	0,623
Gama	1,495	0,931
Log-normal	0,103	0,968
Pareto	0.301	0,040
Weibull	1,732	1,265

A aderência das distribuições foi verificada segundo o KS-teste. Com exceção da distribuição Pareto, para a qual foi obtido um p-valor inferior a 0,05, as demais distribuições apresentaram p-valores superiores a 0,05 indicando que essas distribuições descrevem satisfatoriamente os dados observados de intensidade de seca.

Na Tabela 3 são apresentados os valores de p do KS-teste, log verossimilhança e os valores de AICc. Segundo esses critérios, a distribuição gama foi a que apresentou melhor desempenho, com o menor valor de AICc e maior valor de p em relação às demais, seguida pelas distribuições weibull, exponencial e log-normal. O maior valor de AICc foi o da distribuição Pareto confirmando o resultado do KS-teste como modelo menos adequado.

Tabela 3. p-valores do KS-teste, log verossimilhança e AICc

Modelos	p-valores	LogVer	AICc
Exponencial	0,102	-131,161	267,367
Gama	0,978	-127,251	258,642
Log-normal	0,267	-132,684	269,508
Pareto	4.748×10^{-11}	-205,084	414,309
Weibull	0,903	-127,452	259,044

Como os AICc das distribuições gama e Weibull foram muito próximos, utilizou-se o método da razão de verossimilhanças maximizadas para realizar uma discriminação entre essas distribuições com o objetivo de determinar qual distribuição é a mais adequada para descrever os dados em estudo. A estatística T da razão de verossimilhanças maximizadas, $T = \ln \left[\frac{L_{GA}(\hat{\alpha}, \hat{\beta})}{L_{WE}(\hat{\theta}, \hat{\lambda})} \right] = 0,201$, confirmou que a distribuição gama fornece melhor ajuste com relação aos dados analisados.

Na Tabela 4 são apresentados os valores de intensidade de seca em Laranjeiras do Sul, estimadas pelas distribuições exponencial, gama, log-normal, Pareto e Weibull para os níveis de probabilidade de 1%, 5%, 10%, 90%, 95% e 99%. Estes níveis dizem respeito a ocorrência de seca com intensidade provável $P(X > x)$. Por exemplo, podemos verificar que a intensidade de seca estimada pelo modelo gama para o nível de probabilidade de 95% foi de 4,187, valor muito próximo ao estimado pelo modelo Weibull, 4,121. Observa-se que o modelo Pareto, o de pior ajuste, apresenta uma estimativa extremamente alta, 840,844, nesse mesmo nível de probabilidade.

Tabela 4. Intensidade de seca provável para o município de Laranjeiras do Sul

Modelos	1%	5%	10%	90%	95%	99%
Exponencial	0,016	0,082	0,169	3,697	4,811	7,395
Gama	0,061	0,187	0,311	3,348	4,187	6,081
Log-normal	0,116	0,225	0,320	3,838	5,458	10,562
Pareto	0,041	0,047	0,056	84,051	840,844	176617,700
Weibull	0,045	0,165	0,292	3,347	4,121	5,789

Na Figura 2 são apresentados os gráficos de probabilidade para os cinco modelos avaliados. Ao observar essas figuras, pode-se notar que a distribuição

Pareto foi a com pior ajuste, sendo acompanhada pelas distribuições exponencial e log-normal. Entretanto, os gráficos são inconclusivos em relação à distribuição que apresenta melhor ajuste. Dessa forma, os gráficos de probabilidade corroboram com os resultados obtidos anteriormente pelos testes analíticos.

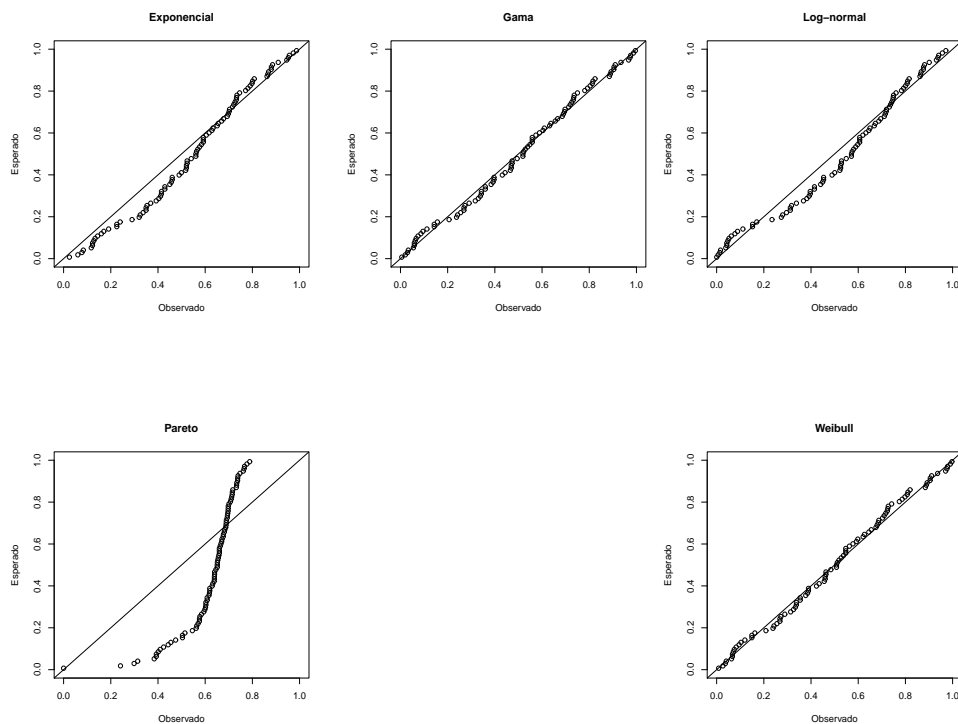


Figura 2 - Gráficos de probabilidade: exponencial, gama, log-normal, Pareto e Weibull.

Na Figura 3 pode-se comparar o histograma de frequências dos dados observados de intensidade da seca sobreposto pelas fdps exponencial, gama, log-normal, Pareto e Weibull que foram ajustadas. O que se pode notar é que as curvas de ajustes das distribuições exponencial, gama, log-normal e Weibull acompanham as condições de moda e assimetria, o que não ocorre com o modelo Pareto. Observa-se que, visualmente, as distribuições gama e weibull melhor se ajustaram aos dados observados.

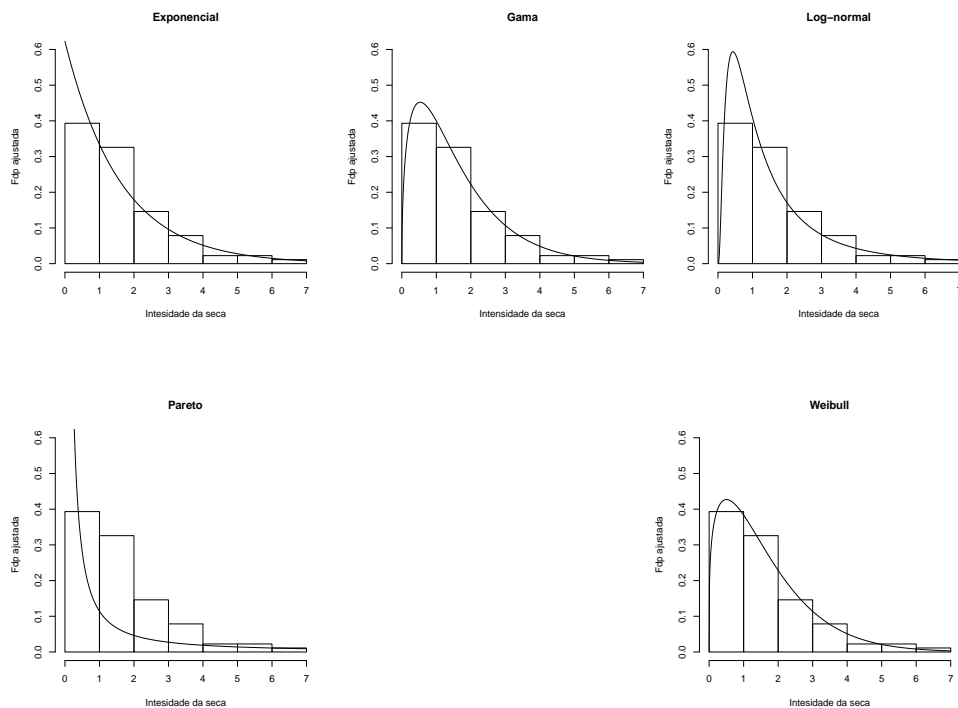


Figura 3 - Fdp ajustadas: exponencial, gama, log-normal, Pareto e Weibull.

4 Conclusões

Utilizando o teste de Kolmogorov-Smirnov no nível de 0,05 de significância, observou-se que com exceção da distribuição Pareto, as distribuições exponencial, gama, log-normal e Weibull apresentaram aderência. As distribuições gama e Weibull apresentaram uma maior aderência aos dados observados;

O critério de informação de Akaike corrigido indicou que a distribuição gama obteve melhor ajuste dos dados, seguida pela distribuição Weibull. O modelo de pior desempenho segundo esse critério foi o Pareto;

A discriminação entre as distribuições gama e weibull via razão de verossimilhanças indicou a distribuição gama como melhor modelo;

Os gráficos de probabilidade indicaram que as distribuições Pareto, exponencial e log-normal apresentaram os piores ajustes. No entanto, foi inconclusivo com relação ao melhor ajuste;

A comparação do histograma de frequências dos dados observados de seca sobreposto com as fdp exponencial, gama, log-normal, Pareto e Weibull que foram ajustadas, indicou que as curvas de ajustes dos modelos exponencial, gama, log-normal e Weibull acompanham as condições de moda e assimetria, o que não

ocorre com a a distribuição Pareto. As distribuições gama e Weibull, visualmente, obtiveram melhor ajuste.

RODRIGUES, J. A.; SILVA, A. P. M.; SANTOS FILHO, J. Use of probability distributions in the modeling of droughts intensity occurred in Laranjeiras do Sul, PR. *Rev. Bras. Biom.*, São Paulo, v.32, n.4, p.570-583, 2014.

- **ABSTRACT:** *Droughts are recurring natural phenomena that affect much of the planet, causing negative impacts on various branches of agriculture, industrial and urban, reflecting directly on the quality of life. Thus, different models have been proposed in an attempt to better understand and predict this phenomenon. In this study, it was evaluated the potential of exponential, gamma, log-normal, Pareto and Weibull distributions modeling of the intensity of droughts in Laranjeiras do Sul in Paraná state. Droughts data referring to a series of 30 years. Adherence to the estimated probabilities of observed frequencies was verified by means of Kolmogorov-Sminorv test. Except for the Pareto distribution, all distributions analyzed were adherent. However, the results indicated that the gamma model was adjusted more adequately to the observed data.*
- **KEYWORDS:** *Akaike information criterion; exponential distribution; Gamma distribution; lognormal distribution; Weibull distribution; standardized precipitation index.*

Referências

- BOZDOGAN, H. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, v.52, p.345-370, 1987.
- BURNHAM, K. P.; ANDERSON, D. R. Multimodel inference: understanding aic and bic in model selection. *Sociological Methods and Research*, v.33, n.2, p.261-304, 2004.
- CHAMBERS, J.; CLEVELAND, W.; KLEINER, B.; TUKEY, P. *Graphical Methods for Data Analysis*. Boston: Duxbury Press, 1983, 395p.
- ELSHERPIENY, E. A.; IBRAHIM, S. A. N.; RADWAN, N. U. M. M. Discriminating between Weibull and log-logistic distributions. *International Journal of Innovative Research in Science, Engineering and Technology*, v.2, n.8, p.3358-3371, 2013.
- GUPTA, R.; KUNDU, D. Discriminating between gamma and generalized exponential distributions. *Journal of Statistical Computation and Simulation*, v.74, p. 107-121, 2004.
- KVAM, P. H.; VIDA KOVIC, B. *Nonparametric statistics with applications to science and engineering*. New York: Wiley, 2007, 420p.
- KUNDU, D. ; GUPTA, R.; MANGLICK, A. Discriminating between the log-normal and generalized exponential distributions. *Journal of statistical planning and inference*, v.127, p.213-227, 2004.

- JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. *Continuous univariate distributions, vol 1*. New York: Wiley, 1995, 761p.
- LONGO, S. L.; SAMPAIO, C. S.; QUEIROZ, M. M. F.; SUSZEK, M. Uso das distribuições gama e log-normal na estimativa de precipitação provável quinzenal. *Revista Varia Scientia*, v.6, n.3, p.107-118, 2006.
- NADARAJAH, S. Generalized Pareto models with application to drought data. *Environmetrics*, v.19, p.395-408, 2008.
- NADARAJAH, S.; GUPTA, A. K. A generalized gamma distribution with application to drought data. *Mathematics and Computers in Simulation*, v.74, p.1-7, 2007.
- McKEE, T. B.; DOESKEN, N. J.; KLEIST, J. Drought monitoring with multiple time scales. In: 9th CONFERENCE ON APPLIED CLIMATOLOGY, American Meteorological Society, Boston, p.233-236, 1995.
- YANG, D. W.; NADARAJAH, S. Drought modeling and products of random variables with exponential kernel. *Stoch Environ Res Risk Assess*, v.21, p.123-129, 2006.
- R DEVELOPMENT CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. ISBN 3-900051-07-0, URL www.R-project.org. 2012.
- RODRIGUES, J. A.; SANTOS FILHO, J.; CHAVES, L. M. Funções densidade de probabilidade para estimativa de precipitação mensal. *Revista Semina: Ciências Exatas e Tecnológicas*, v.34, n.1, p.03-08, 2013.
- SHIAU, J. T. Fitting drought duration and severity with two-dimensional copulas. *Water Resources Management*, v.20, p.795-815, 2006.
- SHIAU, J. T.; FENG, S.; NADARAJAH, S. *Assessment of hydrological droughts for the Yellow River, China, using copulas*. *Hydrological Processes*, v.21, p. 2157-2163, 2007.
- SHIAU, J. T.; MORRADES, R. Copula-Based drought severity-duration-frequency analysis in Iran. *Meteorological Applications*, v.16, p.481-489, 2009.
- SILVA, J. C.; HELDWEIN, A.; MARTINS, F. B.; TRENTIN, G.; GRIMM, E. L. Análise de distribuição de chuva para Santa Maria, RS. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v.11, n.1, p.67-72, 2007.

Recebido em 07.03.2014.

Aprovado após revisão em 30.11.2014.