Revista Brasileira de Agrometeorologia, Santa Maria, v. 10, n. 1, p. 153-162, 2002
Recebido para publicação em 12/12/2001. Aprovado em 21/06/2002.

ISSN 0104-1347

Aplicação de cinco funções densidade de probabilidade a séries de precipitação pluvial no Estado de Minas Gerais

Evaluation of five probability density functions for rainfall data series in the State of Minas Gerais, Brazil

Márcio José Catalunha[1], Gilberto Chohaku Sediyama[2], Brauliro Gonçalves Leal[3],
Carlos Pedro Boechat Soares[4] e Arystides Ribeiro[2]

Resumo - *No presente trabalho foram analisadas as distribuições de probabilidade exponencial, gama, log-normal (a dois e três parâmetros), normal e Weibull. Os testes, não-paramétricos, de qui-quadrado e de Kolmogorov-Smirnov foram utilizados para verificar a aderência das probabilidades estimadas às freqüências observadas. Foram consideradas, para fins de análise a precipitação diária e total para os períodos decêndiais e mensais de janeiro a dezembro. Como período chuvoso, consideraram-se os meses de janeiro, fevereiro, março, outubro, novembro e dezembro, e como período seco, os meses de abril, maio, junho, julho, agosto e setembro. Para os valores diários de precipitação, observou-se a superioridade do ajustamento da distribuição Weibull, com exceção dos decêndios do período seco, em que predomina a distribuição exponencial. No caso dos valores totais de precipitação para o período seco, predominou a distribuição exponencial; no período chuvoso, prevalecendo as distribuições Weibull, exponencial, gama e normal, respectivamente; esta última aparecendo somente em dois meses.*

Palavras-chave*: precipitação pluvial, modelagem.*

Abstract - *This work was carried out (a) to test the probability density function that fits better the observed precipitation frequency (b) to analyze the precipitation data, considering its time variability, for annual and monthly rainy and dry periods, including the average number of days with rainfall in the annual and monthly basis. The Gamma, exponential, normal, Weibull, and log-normal probability distributions (with two and three parameters) were also analyzed. The non-parametric test of qui-square, at 5% of significance, and the Kolmogorov-Smirnov test, at 20% of significance, were used to verify the goodness of fit between estimated and observed rainfall data. For the daily precipitation values, the superiority of Weibull distribution function was demonstrated, except for ten dry days period, in which the exponential probability distribution function was better. For accumulated total precipitation for dry period, the exponential probability distribution prevailed. For the rainy period, the best probabilities distributions were the Weibull distribution, exponential, gamma, and normal, the later only in two months of the year.*

Key words*: precipitation, modelling.*

Introdução

O aproveitamento adequado dos recursos hídricos requer o uso de técnicas de planejamento que se baseiam na estimativa das probabilidades associadas a certas variáveis hidrológicas, no caso a precipitação. Nesses trabalhos, o principal interesse é prever, com base em funções de densidade de probabilidade aplicadas aos dados observados, as precipitações capazes de ocorrer em uma certa localidade.

[1]Doutorando em Eng. Agrícola – UFV/MG (catalunha@catalunha.eng.br).
[2]Professor Departamento de Engenharia Agrícola – UFV/MG (sediyama@mail.ufv.br, ribeiro@mail.ufv.br).
[3]Professor do Centro de Ciências Exatas – UNIVALE/MG (brauliro@uol.com.br).
[4]Professor Departamento Eng. Florestal – UFV/MG (csoares@mail.ufv.br).

O uso de funções densidade de probabilidade está diretamente ligado à natureza dos dados a que ela se relacionam. Algumas têm boa capacidade de estimação para pequeno número de dados, outras requerem grande série de observações. Devido ao número de parâmetros de sua equação, algumas podem assumir diferentes formas, enquadrando-se em um número maior de situações, ou seja, são mais flexíveis. Desde que respeitado o aspecto da representatividade dos dados, as estimativas dos seus parâmetros para uma determinada região podem ser estabelecidas como de uso geral, sem prejuízo da precisão na estimação da probabilidade.

Stern e Coe (1982), citados por ALMEIDA (1995), afirmam que o ajuste de modelos probabilísticos aos dados diários de chuva, além de fornecer um resumo sucinto destes dados, representa uma técnica eficiente para a análise dessas informações. Cada distribuição de freqüência apresenta uma certa forma, esta forma pode ser aproximada através da utilização de equações de densidade de probabilidade com alguns parâmetros extraídos da amostra em questão. A utilização ou não de uma distribuição reside na capacidade da mesma em estimar os dados observados, com base em seus parâmetros, e esta capacidade é medida através de testes de aderência.

Com base na necessidade e importância dos dados pluviométricos, o presente trabalho teve por objetivo: testar o ajuste de diferentes funções densidade de probabilidade aos dados pluviométricos do Estado de Minas Gerais.

Material e métodos

As distribuições utilizadas na análise foram:

a) Distribuição exponencial: A distribuição exponencial é geralmente aplicada a dados com forte assimetria, ou seja, apresentando uma forma de "J" invertido. Sua função densidade de probabilidade é assim descrita (KITE, 1978):

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & ; \ x > 0 \\ 0 & ; \ x < 0 \end{cases} \qquad (1)$$

sua função de distribuição acumulada é do tipo:

$$F(x) = \int_0^\infty f(x) = 1 - e^{-\lambda x} \qquad (2)$$

O único parâmetro da distribuição (λ) é estimado pelo inverso da média.

b) Distribuição gama: Esta distribuição foi testada para modelar as quantidades diárias de chuva, em localidades da Jordânia, Nigéria, Botswana e Sri Lanka, tendo obtido resultados satisfatórios conforme cita ALMEIDA (1995) nos trabalhos de Stern e Coe (1982). FARIA (1998) utilizou a distribuição gama para estimar a precipitação dependente ao nível de 75% de probabilidade, obtendo boa aderência. Na literatura como pode ser observado tem-se utilizado com freqüência esta distribuição na estimação da probabilidade de precipitação.

Se x for uma variável aleatória contínua, tal que ($0 < x < \infty$), com distribuição gama de parâmetros $\alpha > 0$ e $\beta > 0$, então a sua função densidade de probabilidade é definida como:

$$f(x) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-\frac{x}{b}} \quad para\ 0 < x < \infty \qquad (3)$$

Algumas formas de estimar os parâmetros da distribuição gama foram desenvolvidas, contribuindo, junto com a sua flexibilidade de formas, para sua utilização em diversas áreas (HAAN, 1977). O principal método para estimar seus parâmetros é o método de máxima verossimilhança; que para satisfazer a condição $\alpha > 0$ (por definição) e após algumas considerações matemáticas na equação original, Equação 4, a solução que interessa será:

$$\hat{a} = \frac{1 + \sqrt{1 + 4(\ln(\overline{x}) - x_g)/3}}{4(\ln(\overline{x}) - x_g)} \qquad (4)$$

O estimador do parâmetro β, poderá ser obtido por

$$\hat{\beta} = \frac{\overline{x}}{\hat{\alpha}} \qquad (5)$$

sendo $\overline{x}$ a média aritmética e x_g a média geométrica das observações.

Sendo F(x) a probabilidade de ocorrência de um evento menor ou igual a x, pode-se escrever que a função de distribuição acumulada de probabilidade é representada pela função gama incompleta, segundo THOM (1958):

$$F(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^x u^{\alpha-1} e^{-\frac{u}{\beta}} du \qquad (6)$$

em que, F(x) é a probabilidade de ocorrer um valor menor ou igual a x; x a variável aleatória contínua; $\Gamma(\alpha)$ a função gama do parâmetro alfa; α o parâmetro de forma da variável aleatória x; β o parâmetro de escala da variável aleatória x; e base do logaritmo neperiano (2,718281828...) e u é a variável aparente utilizada para integração.

A função de distribuição acumulada da distribuição de probabilidade gama possui integral, que pode ser resolvida por métodos numéricos ou pelo desenvolvimento em série de uma expressão exponencial. Para isso, utilizando-se de transformação de variáveis, tal que $t = u/\beta \rightarrow du = \beta dt$, não mudando o limite inferior zero e substituindo x por t, e após as simplificações adequadas, tem-se, da Equação 4:

$$F(t) = \frac{1}{\Gamma(\alpha)}\int_0^t t^{\alpha-1}e^{-t}dt \qquad (7)$$

Considerando o desenvolvimento em série de e^t, multiplicando-se e dividindo-se a equação anterior por e^{-t}, reunindo, do produto resultante, os termos em t, t^2, t^3,... e os demais termos correspondentes aos diversos expoentes, tem-se, segundo ASSIS et al. (1996):

$$F(t) = \frac{t^a}{a\Gamma(a)e^t}\left[1 + \frac{t}{a+1} + \frac{t^2}{(a+1)(a+2)} + \right.$$
$$\left. + \frac{t^3}{(a+1)(a+2)(a+3)} + \ldots\right] \qquad (8)$$

em que, t é x/β, é uma aproximação da distribuição gama por expansão em série.

c) Distribuição log-normal: Uma outra distribuição, testada por HUF & NEIL (1959) num trabalho de comparação entre vários métodos para analisar freqüência de precipitação, é a distribuição log-normal, a qual assume que os logaritmos das variáveis aleatórias são normalmente distribuídos.

Conforme MIRSHAWKA (1971), a função densidade da distribuição log-normal a dois parâmetros e a três parâmetros são representadas pela seguinte equação:

$$f(x) = \frac{1}{(x-a)\sigma\sqrt{2\pi}}\exp\left(-\frac{[\ln(x-a)-\mu]^2}{2\sigma^2}\right) \qquad (9)$$

em que f(x) é a função densidade de probabilidade da variável; *e* a base dos logaritmos neperianos; x o valor da variável aleatória; μ a média dos logaritmos da variável x; σ o desvio-padrão dos logaritmos da variável x; e a o limite inferior da amostra.

Para encontrar a probabilidade de que uma variável aleatória x tendo distribuição log-normal, assuma valores entre a e b ($a \leq x \leq b$), tem-se

$$F(a \leq x \leq b) = \int_a^b \frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{[\ln(x-a)-\mu]^2}{2\sigma^2}\right)dx \qquad (10)$$

O valor de "a" pode ser zero, quando se considera a distribuição log-normal a dois parâmetros, ou um valor mínimo da série, quando se considera log-normal a três parâmetros. O valor de "b" pode ser o da variável aleatória, quando se considera a probabilidade cumulativa de ocorrência daquele valor (HASTINGS e PEACOCK, 1975).

d) Distribuição normal: A distribuição de probabilidade contínua mais utilizada é a distribuição normal (HASTINGS & PEACOCK, 1975), geralmente citada como curva normal ou curva de Gauss. Sua importância em análise matemática resulta do fato de que muitas técnicas estatísticas, como análise de variância, de regressão e alguns testes de hipótese, assumem ou exigem a normalidade dos dados.

A distribuição normal é uma distribuição de dois parâmetros. Sua função densidade de probabilidade tem a seguinte forma:

$$f(x) = \frac{1}{s\sqrt{2p}}\exp\left(-\frac{(x-m)^2}{2s^2}\right); \qquad (11)$$

$$para\ -\infty < x < +\infty$$

em que μ e a media e σ o desvio-padrao da variável aleatória.

A probabilidade de que uma variável x assuma valores menores ou iguais a x quando ela tem $N(\mu,\sigma^2)$, distribuição normal com média μ e variância σ^2, é estimada por

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx \qquad (12)$$

Mas essa equação não pode ser resolvida analiticamente sem o uso de métodos de integração aproximada. Por esse motivo, usa-se a transformação $Z=(x-\mu)/\sigma$, a variável Z tem distribuição normal com média zero e variância um [N(0,1)]. A variável Z é chamada variável reduzida, e a Equação 12 pode ser reescrita na seguinte forma:

$$F(Z) = \frac{1}{\sqrt{2p}} \int_{-\infty}^{z} \exp\left(-\frac{Z^2}{2}\right) dz \qquad (13)$$

$$para\ -\infty \leq Z \leq \infty$$

PACITTI (1974) argumenta que, para evitar grande trabalho de computação no cálculo da função da distribuição normal padrão, usa-se um polinômio que aproxima a função com boa precisão. Este polinômio é o apresentado a seguir:

$$F(Z) \approx 1 - \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right)\left(a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 + a_5 t^5\right) \qquad (14)$$

para Z ≥ 0, já que para Z < 0 tem-se *F(Z) = 1-F(-Z).*

Em que t = 1/ (1+kZ) ; k = +0,2316419; a_1 = +0,31938153; a_2 = -0,356563782; a_3 = +1,781477937; a_4 = -1,821255978; a_5 = +1,330274429; e Z é a variavel reduzida ou transformada de x.

e) Distribuição Weibull: É utilizada em análise hidrológica para eventos extremos; contudo, sua utilização em séries climáticas é pouco conhecida. Foi proposta primeiramente por Fisher e Tippet em 1928, tendo sido desenvolvida independentemente por Walodi Weibull, em 1939, o trabalho de Weibull se destacou e a distribuição passou a ser chamada pelo seu nome (JOHNSON & KOTZ, 1970).

Sua função de densidade de probabilidade é apresentada de diversas formas, sendo comum em alguns trabalhos apresentá-la como

$$f(x) = \frac{g}{b}\left(\frac{x-a}{b}\right)^{g-1} \exp\left[-\left(\frac{x-a}{b}\right)^{g}\right] \qquad (15)$$

$$para\ \ x \geq a$$

em que f(x) = 0 para outros intervalos, x é a variável aleatória e α ≥ 0, β > 0 e γ> 0 os parâmetros da distribuição. Nessa forma, a distribuição é normalmente chamada de Weibull a três parâmetros. A sua função de distribuição acumulada é

$$F(x) = \int_0^{\infty} f(x)dx = 1 - \exp\left[-\left(\frac{x-\alpha}{\beta}\right)^{\gamma}\right] \qquad (16)$$

O parâmetro α é chamado de parâmetro de posição, pois controla a posição da curva sobre o eixo das abscissas. O parâmetro de escala β controla as dimensões que a curva assume, dada uma forma constante. O parâmetro γ, é chamado parâmetro de forma. Algumas propriedades da Weibull a dois parâmetros são apresentadas por JOHNSON & KOTZ (1970).

Muitas maneiras de estimar os parâmetros da distribuição Weibull foram desenvolvidas. O principal método de ajuste da distribuição Weibull (COUTO, 1980) e o da máxima verossimilhança, que consiste em determinar os valores de γ e β pela suas equações fundamentais. Nota-se, porém, que β é função de γ. Normalmente, a solução desse tipo de sistema (Equação 17) é obtido resolvendo a equação por um processo interativo, utilizou-se o método da bissecção, sendo β encontrado, a partir de γ aproximado, pela substituição na Equação 18

$$\frac{\sum_{i=1}^{n} x_i^{\hat{\gamma}} \ln(x_i)}{\sum_{i-1}^{n} x_i} - \frac{\sum_{i=1}^{n} \ln(x_i)}{n} - \frac{1}{\hat{\gamma}} = 0 \qquad (17)$$

$$\hat{\beta} = \left(\frac{\sum_{i-1}^{n} \ln(x_i^{\hat{\gamma}})}{n}\right)^{\frac{1}{\hat{\gamma}}} \qquad (18)$$

Ao se ajustar uma distribuição de probabilidade a um conjunto de dados, trabalha-se com a hipótese de que a distribuição pode representar adequadamente aquele conjunto de informações. Uma maneira de comprovar esta hipótese é através de alguns testes não paramêtricos, como os mencionados a seguir:

a) Teste de aderência por χ^2 (qui-quadrado): A hipótese de nulidade admite que a distribuição seja a especificada (gama, Weibull etc...), com os seus parâmetros estimados com base nos dados amostrais. A hipótese é testada fazendo-se a comparação entre as freqüências observadas e as freqüências teóricas, em cada classe de freqüência da amostra, com a variável aleatória χ^2 dada por (CAMPOS, 1979):

$$\chi^2 = \sum_{i=1}^{k}\left(\frac{(Fo_i - Fe_i)^2}{Fe_i}\right) \qquad (19)$$

em que, k é o número de classes, Fo_i a freqüência observada e Fe_i a freqüência esperada, de acordo com a distribuição que está sendo testada. Os valores críticos de χ^2 para alguns níveis de significância são descritos por tabelas próprias.

b) Teste de aderência por Kolmogorov-Smirnov: Este teste foi introduzido por Kolmogorov e Smirnov (1933), citados por ASSIS (1996), como metodologia para sua aplicação, pode-se considerar F(x) a proporção dos valores esperados menores ou iguais a x e S(x) a proporção dos valores observados menores ou iguais a x, em que *Dobs* é módulo do desvio máximo observado:

$$Dobs = Max \mid F(x) - S(x) \mid \qquad (20)$$

Para isto compara-se *Dobs* com *Dtab* (*Dtab* é o desvio máximo tabelado, encontrado em tabelas adequadas); se *Dobs* for menor, existe concordância entre as freqüências observadas e esperadas, a amostra provêm de uma população que segue a distribuição de probabilidade sob teste.

O teste de qui-quadrado é específico para dados agrupados, em que as classes que possuírem valores menores que três ou cinco devem ser agrupadas em outras classes, sendo um fator limitante para uso em dados com poucas classes. Este teste é baseado na soma dos erros absolutos das freqüências, que é comparada com um valor tabelado de acordo com o nível de significância desejado e os graus de liberdade da distribuição. Isto favorece o aspecto cumulativo dos erros pela somatória.

O teste de Kolmogorov-Smirnov pode ser usado tanto para dados agrupados quanto para dados individuais. Nos dados agrupados não há restrição quanto ao número nem ao valor das classes. É baseado no módulo da maior diferença entre a probabilidade observada e a estimada, que é comparada com um valor tabelado de acordo com o número de observações da série sob teste. Isto evita o aspecto cumulativo dos erros.

Foram utilizadas aproximadamente 243 estações meteorológicas situadas no Estado de Minas Gerais, fornecidos pela Agência Nacional de Águas e Energia Elétrica (ANEEL) e pelo Instituto Nacional de Meteorologia (INMET). A *World Meteorological Organization* (WMO) em sua nota técnica nº 82 (THOM, 1966) preconiza que o número mínimo de anos de dados climáticos para análise é de 30 anos, bem como recomenda o *run test*, para análise de homogeneidade dos dados. Tais considerações foram observadas. Foram considerados, para fins de análise a precipitação diária e total para os períodos decendiais e mensais de janeiro a dezembro. Considerou-se somente dias com chuva, ou seja, aqueles com precipitação maior ou igual a 0,1mm Gates e Tong (1976), citados por ALMEIDA (1995); onde dias com precipitação igual a zero foram desconsiderados, devido ao uso de função logarítmica para cálculo dos parâmetros de algumas distribuições; os anos bissextos foram desconsiderados. Para verificar o nível de aderência entre a probabilidade observada e a estimada para as distribuições exponencial, gama, log-normal, normal e Weibull aplicadas em cada um dos períodos descritos, foram utilizados os testes de qui-quadrado e Kolmogorov-Smirnov, com níveis de significância de 5% e 20%, respectivamente. Ressalta-se que as funções foram escolhidas dentre algumas comumente utilizadas para estas análises (ASSIS et. all, 1996; HASTINGS & PEACOCK, 1975).

Resultados e discussão

A Tabelas 1 mostra a distribuição à qual os dados melhor se ajustaram pelos testes de qui-quadrado e Kolmogorov-Smirnov, para períodos considerados, enquanto nas Tabelas 2 e 3 estão apresentados os números de aderências observadas pelos mesmos testes

A seguir é apresentada uma análise das distribuições e dos testes, procurando destacar algumas características importantes, independentemente de terem tido ou não bom desempenho na análise geral.

a) Distribuição exponencial: Nesta distribuição os dados se concentram nas classes iniciais e diminuem a concentração nas finais. Pode-se questionar o fato de que, mesmo tendo naturalmente os dados, esta distribuição de freqüência (distribuição exponencial) não foi a que melhor estimou os valores observados para alguns períodos. Isto se deve à capacidade do modelo matemático em estimar cada classe de freqüência individualmente, quando ocorrem picos intermediários de freqüência.

Ao observar os dados das Tabelas 2 e 3, para a distribuição exponencial, verifica-se boa aderência nos meses com chuvas e diminuição nos meses secos, para os períodos decendiais totais e diários. Por outro lado, no período mensal total há uma inversão, tendo a melhor aderência os meses secos.

Considerando o erro relativo entre as probabilidades observadas e as estimadas, pode-se observar

Tabela 1. Resumo do ajustamento das funções densidade de probabilidade estudadas, com base no teste de qui-quadrado e Kolmogorov-Smirnov para mês e períodos considerados (W=Weibull; E=exponencial; G=gama; N=normal; Ln=log-normal)

Mês	Diário do Decêndio			Diário Mensa	Total do Decêndio			Total Mensal
	1ª	2ª	3ª		1ª	2ª	3ª	
				Teste de Qui-quadrado				
Janeiro	W	W	W	W	G	W	W	G
Fevereiro	W	W	W	W	E	E	E	G
Março	W	W	W	W	W	E	E	G
Abril	*W*	*W*	*W*	*W*	*E*	*E*	*E*	*G*
Maio	*E*	*E*	*E*	*W*	*E*	*E*	*E*	*E*
Junho	*E*	*E*	*E*	*W*	*E*	*E*	*E*	*E*
Julho	*E*	*E*	*E*	*W*	*E*	*E*	*E*	*E*
Agosto	*E*	*E*	*E*	*W*	*E*	*E*	*E*	*E*
Setembro	*E*	*E*	*E*	*W*	*E*	*E*	*E*	*E*
Outubro	W	W	W	W	E	E	W	N
Novembro	W	W	W	W	W	W	W	N
Dezembro	W	W	W	W	G	G	W	G
				Teste de Kolmogorov-Smirnov				
Janeiro	G	G	G	W	G	G	G	G
Fevereiro	G	G	G	G	W	G	G	G
Março	G	G	G	W	G	G	G	N
Abril	*G*	*G*	*W*	*W*	*G*	*G*	*G*	*G*
Maio	*G*	*G*	*G*	*G*	*G*	*G*	*G*	*G*
Junho	*G*	*Ln*	*G*	*G*	*G*	*G*	*G*	*G*
Julho	*G*	*G*	*G*	*G*	*G*	*G*	*G*	*G*
Agosto	*G*	*G*	*G*	*G*	*G*	*G*	*G*	*G*
Setembro	*G*	*G*	*G*	*G*	*G*	*G*	*G*	*G*
Outubro	G	W	G	W	G	G	G	G
Novembro	G	G	G	W	G	G	G	G
Dezembro	G	G	G	G	G	G	G	G

que as probabilidades estimadas no período com chuva são em grande parte superestimadas. Uma característica da distribuição exponencial é a de que partindo de um valor da classe inicial, estima-se outros valores proporcionalmente menores, formando uma curva em forma de "J" invertido. Ao analisar a distribuição de classes de freqüência de algumas estações isoladamente, nota-se que as classes intermediárias têm valores maiores que as classes iniciais ou bem próximo desta, não formando uma curva em "J" invertido, mas sim algo próximo a um "M". A capacidade da distribuição exponencial é limitada em estimar estes "vales" intermediários, superestimando-os, pois seu único parâmetro é a média; ao se efetuarem os testes de aderência, estas classes são reprovadas. O desvio-padrão entre os três decêndios totais é pequeno, se comparado com o desvio-padrão do período mensal total. Com as classes de freqüências das estações no período seco, esses picos intermediários não ocorrem, tendo estas a forma de um "J" invertido suave, o que reduz o erro relativo na estimação, aumentando o número de aderências.

b) Distribuição gama: Os parâmetros da distribuição gama tiveram valores mínimo, médio e máximo de 0,32449, 1,16340 e 11,33372, respectivamente, para alfa, e de 0,43699, 26,82905 e 345,71672, respectivamente, para beta. Trata-se de valores importantes, pois, quando utilizados em outras análises, no Estado de Minas Gerais, estes parâmetros terão uma faixa de abrangência próxima destes valores, norteando o pesquisador em seus resultados. Nas condições em questão para o teste de Kolmogorov-Smirnov a aderência desta distribuição é predominante, conforme Tabela 1, em contrapartida para o

Rev. Bras. Agrometeorologia, v. 10, n. 1, p. 153-162, 2002

Tabela 2. Número de aderências ao teste de qui-quadrado .

Distribuição/Mês	Jan.	Fev.	Mar.	Abr.	Mai.	Jun.	Jul.	Ago.	Set.	Out.	Nov.	Dez.
						Diário - 1º decêndio						
Exponencial	76	72	64	56	61	55	48	31	71	71	69	75
Gama	84	80	79	74	55	41	29	17	60	69	71	76
LogNormal	29	40	35	52	30	22	11	8	27	43	41	25
LogNormal3	19	24	24	25	7	1	2	0	6	21	22	18
Normal	0	0	1	2	8	7	5	8	14	1	0	0
Weibull	86	82	87	80	53	40	27	19	60	72	77	78
						Diário - 2º decêndio						
Exponencial	77	70	65	64	64	51	51	57	61	67	75	73
Gama	76	78	72	70	55	33	28	35	60	74	81	75
LogNormal	26	33	35	43	26	12	13	18	31	43	33	23
LogNormal3	17	19	19	14	6	2	1	1	9	25	25	16
Normal	0	0	0	3	9	10	7	12	9	1	0	0
Weibull	79	79	77	75	51	32	25	36	58	80	83	76
						Diário - 3ª decêndio						
Exponencial	74	73	56	62	54	53	47	55	75	64	69	69
Gama	78	80	72	60	45	29	32	43	75	70	78	74
LogNormal	26	44	40	44	37	13	18	23	33	39	29	24
LogNormal3	21	25	27	8	6	0	2	2	11	28	21	15
Normal	0	1	0	5	4	8	10	6	7	0	0	0
Weibull	85	82	77	64	50	28	28	44	75	76	81	77
						Diário -Mensal						
Exponencial	50	47	35	28	49	47	54	52	57	38	46	54
Gama	67	67	60	53	52	56	58	52	69	57	60	67
LogNormal	4	15	17	39	43	37	41	37	31	19	10	5
LogNormal3	3	10	13	25	17	14	13	12	17	14	7	4
Normal	0	0	0	0	2	1	1	5	1	0	0	0
Weibull	75	74	73	70	65	61	59	56	73	70	67	71

teste de qui-quadrado ela surgiu apenas em alguns totais (decêndiais ou mensais) do período chuvoso. Considerando a análise dos testes de aderência, podemos observar que a somatória dos erros de estimação é grande mas sua amplitude é pequena em relação as outras distribuições, sobretudo Weibull, o que explica sua grande aderência pelo teste de Kolmogorov-Smirnov e reprovação por qui-quadrado.

c) Distribuição log-normal: Embora esta distribuição seja muito empregada em outras áreas da análise climática e hidrológica, não apresentou desempenho satisfatório na estimação das probabilidades, não sendo em média, em nenhum dos períodos considerados (Tabela 1), uma boa distribuição para estimação de dados nas condições e períodos estudados.

d) Distribuição normal: Ao observar os dados das Tabelas 2 e 3, nota-se que, para as estimativas diárias (decendiais e mensais) a aderência da normal é muito baixa; já para as estimativas totais (decendiais e mensais) ela é alta. Ao analisar a distribuição de classes de freqüência de algumas estações isoladamente, observa-se que para os valores diários há maior freqüência nas classes iniciais, reduzindo bruscamente a partir da segunda ou terceira classe. Essa queda não é acompanhada pelo modelo matemático da distribuição normal (que é mais simétrica), como o são os modelos da distribuição gama e Weibull (bem assimétricas), superestimando os valores das classes seguintes. Quando da aplicação dos testes de aderência, estas classes são reprovadas. Para os valores to-

Tabela 3. Número de aderências ao teste de Qui-quadrado.

Distribuição/Mês	Jan.	Fev.	Mar.	Abr.	Mai.	Jun.	Jul.	Ago.	Set.	Out.	Nov.	Dez.
Total - 1ª decêndio												
Exponencial	58	82	72	83	55	32	22	5	39	75	61	60
Gama	67	63	76	61	39	11	9	2	26	61	75	83
LogNormal	33	27	31	26	12	4	3	0	5	21	33	45
LogNormal3	7	3	8	2	0	0	0	0	0	3	11	15
Normal	60	51	64	34	25	8	4	1	20	47	72	73
Weibull	65	67	76	64	39	10	8	1	26	63	79	82
Total - 2º decêndio												
Exponencial	63	78	75	73	53	11	11	11	41	78	63	40
Gama	76	71	69	58	26	3	5	1	24	74	72	76
LogNormal	31	27	24	23	9	1	1	0	6	27	28	41
LogNormal3	10	8	5	1	0	0	0	0	0	5	8	16
Normal	74	66	61	38	18	1	3	2	12	55	75	75
Weibull	80	77	74	58	27	3	5	1	23	74	79	74
Total - 3º decêndio												
Exponencial	71	71	79	66	49	13	18	17	63	67	54	58
Gama	73	62	69	47	24	5	9	6	56	71	76	77
LogNormal	25	20	25	19	6	0	3	1	19	25	45	39
LogNormal3	7	4	6	1	0	0	0	0	2	5	12	14
Normal	70	47	48	28	12	3	7	5	46	70	71	75
Weibull	75	63	67	45	24	5	9	4	56	75	78	78
Total - Mensal												
Exponencial	31	33	28	61	80	68	58	56	79	29	4	4
Gama	82	84	82	83	68	48	40	33	72	77	80	86
LogNormal	54	42	46	57	36	18	13	11	30	47	58	71
LogNormal3	12	6	9	12	7	1	1	0	2	7	8	7
Normal	82	82	82	71	51	25	28	16	56	82	83	83
Weibull	59	74	62	78	71	47	42	33	75	66	45	31

tais (decendiais e mensais), não ocorreu variação brusca entre as classes iniciais, o que torna mais possível ao modelo sua estimação, aumentando sua aderência nesse período.

e) Distribuição Weibull: Verifica-se que, para os valores diários de precipitação no período com chuva, esta distribuição se mostra um modelo dos mais precisos, superando a gama, que é uma distribuição muito empregada nestes casos.

No caso das estimativas diárias de probabilidade, detectou-se a superioridade do desempenho da distribuição Weibull, com exceção dos decêndios do período seco, em que predomina a distribuição exponencial.

Nas estimativas totais de probabilidade para o período seco, é predominante a utilização da distribuição exponencial. Já no período com chuva há variação entre as distribuições Weibull, exponencial, gama e normal, nesta mesma ordem; a última aparece somente em dois meses. Em alguns casos, se fosse escolhida a segunda melhor distribuição para o período, sem distanciar muito da melhor, ter-se-ia uma menor variação no tipo de distribuição para o contexto dos valores mensais, predominando uniformemente a Weibull. Esta predominância sobre outras distribuições, estudadas em condições semelhantes, também foi obervado por DUAN et. al. (1998) no noroeste do pacífico nos Estados Unidos.

Os parâmetros da distribuição Weibull tiveram valores mínimo, médio e máximo de 0,10000, 2,57042 e 160,80000, respectivamente, para alfa, de 0,55191, 37,56826 e 482,87709, respectivamente, para beta, e de 0,44177, 1,04472 e 4,37132, respectivamente, para gama.

Comparando os resultados da Tabela 1, nota-se que, para o teste de Kolmogorov-Smirnov, a distribuição gama teve grande aderência em todos os períodos considerados, o que não ocorreu para o teste de qui-quadrado, sendo a distribuição Exponencial e Weibull as de melhor desempenho.

f) Testes de aderência : O teste de Kolmogorov-Smirnov é bastante utilizado para análise de aderências de distribuições em estudos climáticos; contudo, o seu nível de aprovação de uma distribuição sob teste é muito elevado, como pode ser visto na Tabela 1, o que gera certa insegurança em relação aos critérios do teste, conforme exposto anteriormente. Tendo em vista as características apresentadas pelo teste de qui-quadrado, optou-se pela sua escolha como referência em análises realizadas neste trabalho.

O teste de qui-quadrado é considerado mais rigoroso do que o de Kolmogorov-Smirnov, o que se deve no presente trabalho, entre outros, aos seguintes aspectos: a) Considerando que uma distribuição sob teste tenha duas ou mais classes com probabilidades observadas diferentes das estimadas e, conseqüentemente, freqüências da mesma forma, quando se aplicam esses valores à equação de definição do teste qui-quadrado (Equação 19), tem-se uma somatória dos erros absolutos; aplicando a equação de definição do teste de Kolmogorov-Smirnov (Equação 20), tem-se um único valor, o módulo da diferença.

Isto mostra que os erros, no teste de qui-quadrado, são considerados de forma cumulativa e em todas as classes e que, no teste de Kolmogorov-Smirnov, eles são considerados somente na classe em que foi maior. b) O quadro de freqüências de uma série pluviométrica apresenta maiores valores nas classes iniciais e menores nas finais; a definição do teste de qui-quadrado determina que devem ser reunidas em uma única classe as classes com freqüências estimadas inferiores a três ou cinco. Os modelos testados geralmente superestimam as classes iniciais e subestimam as finais, com algumas exceções.

Para atender as definições do teste de qui-quadrado, as classes estimadas com freqüência inferior a três ou cinco devem ser somadas a outra classe mais próxima. Decorre daí que a somatória ocorrerá também nas classes de freqüência observadas, gerando um erro absoluto grande, que, somado aos anteriores, resulta em valores de qui-quadrado maiores que os tabelados, não aprovando a distribuição sob teste, quando a estimação não for boa. c) Para o teste de Kolmogorov-Smirnov, com base na situação apresentada na letra anterior, não surgirá problema algum, pois, independentemente da distribuição de classes, o que interessa é o módulo da maior diferença, permitindo que o teste aprove a maioria das distribuições, com muitos erros, mas de pequena proporção. d) Os valores de qui-quadrado calculados são comparados com valores críticos tabelados.

Estes valores críticos são obtidos de tabelas referenciadas pelo nível de significância e pelo grau de liberdade, no caso do qui-quadrado, e pelo nível de significância e pelo número de observações, no teste de Kolmogorov-Smirnov. Observa-se que, neste teste, independentemente da capacidade da distribuição em estimar as freqüências observadas e do número de classes, o valor crítico tabelado depende unicamente do número de observações, algo que não varia de distribuição para distribuição, dependendo apenas da série sob teste. Considerando agora o teste de qui-quadrado, vê-se que o grau de liberdade depende dos parâmetros da distribuição, em torno de dois ou três, e do número de classes (característica dos dados). Analisando a observação feita na letra “a”, este número reduz quando a distribuição subestima as classes finais, devido ao agrupamento de algumas classes em outras, e o grau de liberdade fica menor, reduzindo o valor crítico tabelado; isto mostra que o valor crítico para o qui-quadrado depende da capacidade da distribuição em estimar as freqüências observadas, o que não ocorre no Kolmogorov-Smirnov.

Conclusões

Considerando os resultados obtidos, concluiu-se, para as condições estudadas, que:

a) Para as estimativas diárias (decendiais e mensais) da probabilidade, destaca-se o desempenho da distribuição Weibull, com exceção dos decêndios do período seco, em que predominou a distribuição exponencial. Portanto, para o estado de Minas Gerais, não se recomenda a distribuição gama;
b) Nas estimativas totais (decendiais e mensais) da probabilidade para o período seco, é predominante a utilização da distribuição exponencial, e, para o período com chuva, há variação entre as distribuições Weibull, exponencial, gama e normal, nesta ordem, com esta última aparecendo somente em dois meses.
c) O teste de qui-quadrado apresentou melhores ca-

racterísticas para verificar o ajustamento de uma distribuição de probabilidade estimada a dados observados, para os propositos do presente trabalho.

Referências bibliográficas

ALMEIDA, R.M.B. **Características climatológicas do regime de chuva em Minas Gerais**, Viçosa: Universidade Federal de Viçosa, 1995. 64 p. Tese (Mestrado em Metorologia Agrícola), UFV, 1995.

ASSIS, F.N. de, ARRUDA, H.V., PEREIRA, A. R. **Aplicações de estatísticas a climatologia; Teoria e prática**, Pelotas: Ed. Universitária/UFPel, 1996. 161 p.

CAMPOS, H. **Estatística experimental não paramétrica**, Piracicaba: Escola Superior de Agricultura Luiz de Queiroz/ USP, 1979. 3. ed, 200 p.

COUTO, H.T.Z. **Distribuições de diâmetro em plantações de *Pinus caribaea morelet***, Piracicaba: Escola Superior de Agricultura Luiz de Queiroz/ USP, 1980, 79 p. Tese (Livre Docência), ESALQ, 1980.

DUAN, JINFAN et. al. Evaluation of probability density function in precipitation models for the pacific northwest**, Journal of the American Water Resources Association**, v. 34, n. 3, p. 617-627, 1998.

FARIA, R.A. **Demanda de irrigação suplementar no Estado de Minas Gerais**, Viçosa: UFV, 1998, 75 p.

HAAN, C.T. **Statistical methods in hydrology**, Ames: Iowa State University Press, 1977. 378 p, il.

HASTINGS, N.A.J., PEACOCK, J.B. **Statistical distributions: A handbook for students and practitioners**, Longon Butterworths, England, 1975. 129 p.

HUF, F.A. NEIL, J.C. Comparison of several methods for rainfall frequency analysis. Illinois State Water Survey Urbana, **Journal of Geophysical Research**, Illinois, v. 64, n. 5, p. 541-547, 1959.

JOHNSON, N.L, KOTZ, S. **Distribution in statistics, continue univariate distribution**, New York: Hougton Mifflin, v. 2, 1970. 328 p.

KITE, G.W. Frequency and risk analysis in hydrology**, Water Resources Publications**, Fort Collins v. 3, 395 p., 1978.

MIRSHAWKA, V. **Estatística**, Vol II, São Paulo Nobel, 1971. 367 p.

PACITTI, T. **Fortran**. Rio de Janeiro: Livro Técnico e Científico, 1974. 377 p.

THOM, H.C.S. **A note on the gamma distribution. Monthly Weather Review**, Washington, v. 86, p. 117-122. 1958.

THOM, H.C.S. **Some methods of climatological analysis. Genebra:** Word Meteorological Organizatios. 1966. 103 p. (Nota técnica n. 81).