

Modelagem de Quantidades Precipitadas em Escala Diária: Uma Análise Comparativa

Daniel Henrique Marco Detzel; Miriam Rita Moro Mine

Universidade Federal do Paraná

daniel_detzel@yahoo.com.br; mrmine.dhs@ufpr.br

Recebido: 14/07/10 - revisado: 30/11/10 - aceito: 25/03/11

RESUMO

Modelos estocásticos para geração de séries sintéticas de precipitações são ferramentas úteis a diversas aplicações práticas. Dentre os que trabalham com escalas diárias, formulações paramétricas são as mais populares, principalmente por sua grande flexibilidade. A estrutura desses modelos se divide em (i) determinação das ocorrências, com aplicação de processos estocásticos e (ii) cálculo das alturas precipitadas, com aplicação de distribuições estatísticas. O presente artigo se concentra na fase (ii) e na análise específica das distribuições probabilísticas mais utilizadas: exponencial simples, exponencial a dois parâmetros, gama a dois parâmetros e exponencial mista a três parâmetros. A intenção é investigar o comportamento dessas distribuições em onze postos pluviométricos espalhados nas bacias dos rios Paraná e Uruguai, através da aplicação do teste PPCC, análise comparativa baseada em gráficos quantil-quantil e determinação dos critérios de informação AIC e BIC para cada caso.

Palavras-Chave: Modelos estocásticos; séries sintéticas; teste PPCC.

INTRODUÇÃO

Modelos estocásticos para geração de séries de precipitações são de grande utilidade para diversas aplicações práticas. Dentre elas, destacam-se a determinação de cheias de projeto para obras hidráulicas, análises referentes a culturas agrícolas, estudos relacionados a hidroelectricidade, entre outros. Ademais, a aplicação cada vez mais habitual do Método de Monte Carlo para estimar as propriedades estatísticas das variáveis hidrológicas envolvidas agrega uma importância extra para a geração de séries consistentes.

Três são as escalas temporais comumente utilizadas para estruturação dos modelos: chuvas anuais, mensais e diárias. Diferentemente das duas primeiras, chuvas diárias ganharam ênfase dentre os pesquisadores, principalmente por sua complexidade e maior versatilidade de aplicação das séries geradas. A forma de se estruturar um modelo de geração de chuvas diárias não é única; Srikanthan e McMahon (2001) e Detzel e Mine (2009) fazem uma revisão dos diversos métodos existentes, evidenciando que diferentes soluções refletem em resultados com maior ou menor precisão.

Por trazer grande flexibilidade de ajuste a diferentes regiões de aplicação, modelos de geração

ditos paramétricos são os mais utilizados. Sua estrutura é dividida em duas fases distintas: (i) determinação das ocorrências das chuvas e (ii) cálculo das alturas precipitadas para dias considerados chuvosos. Na primeira fase, é comum o emprego de processos estocásticos markovianos ou de alternância de eventos; na segunda, diferentes distribuições estatísticas são aplicadas.

O presente artigo concentra-se na análise específica dos diferentes modelos probabilísticos utilizados na determinação das alturas precipitadas. Em geral, distribuições estatísticas que representam variáveis atmosféricas apresentam um comportamento significativamente assimétrico (Wilks, 2006). Portanto, diversos autores optaram por quatro principais distribuições com essas características, a saber: exponencial simples (Haan *et al.*, 1976; Richardson, 1981; Todorovic e Woolhiser, 1975), exponencial a dois parâmetros (Nascimento e Kelman, 1995), exponencial mista a três parâmetros (Wilks, 1998; Smith e Schreiber, 1974; Foufoula-Georgiou e Lettenmaier, 1987) e gama a dois parâmetros (Katz, 1977; Krüger *et al.*, 1998; Grondona *et al.*, 2000; Liao *et al.*, 2004; Gomes *et al.*, 2006). A qualidade de ajuste proporcionada por essas distribuições é variável, dependendo principalmente do regime pluviométrico da região na qual se deseja aplicar o modelo.

A intenção é investigar o comportamento das quatro distribuições citadas em estações pluviométricas espalhadas pelas bacias dos rios Paraná e Uruguai, componentes da porção brasileira da bacia do rio da Prata. Primeiramente, a possibilidade de uso desses modelos estatísticos para a região em estudo é averiguada formalmente, através da aplicação do teste PPCC (*Probability Plot Correlation Coefficient*) (Filliben, 1975). Na sequência, análises comparativas são feitas, com base em gráficos quantis-quantis (*Q-Q plots*) e nos coeficientes de correlação obtidos. Finalmente, são empregados os critérios indicadores de Akaike (AIC - Akaike, 1974) e de Bayes (BIC - Schwarz, 1978). Embora seja um estudo comparativo, o objetivo principal do trabalho não é apontar a distribuição de melhor desempenho geral, mas discutir as escolhas e possíveis consequências para cada modelo estatístico em questão.

MODELOS ESTATÍSTICOS CONSIDERADOS

Para o desenvolvimento das análises, as séries históricas são compostas somente por dias chuvosos. Assim, seja a sequência de variáveis aleatórias $X = (x_1, x_2, \dots, x_n)$, na qual cada x_i representa a altura precipitada no dia i , em uma amostra com n dias chuvosos. Quatro são as distribuições estatísticas consideradas, expressas abaixo em função de suas funções densidade de probabilidade (FDP):

Distribuição Exponencial simples (um parâmetro):

$$f_X(x) = \frac{1}{\beta} \exp\left(-\frac{x}{\beta}\right); x > 0 \quad (1)$$

Distribuição Exponencial a dois parâmetros:

$$f_X(x) = \frac{1}{\beta} \exp\left(-\frac{x-\alpha}{\beta}\right); x > \alpha; \beta > 0 \quad (2)$$

Distribuição Gama a dois parâmetros:

$$f_X(x) = \frac{x^{\alpha-1} \exp(-x/\beta)}{\beta^\alpha \cdot \Gamma(\alpha)}; x > 0; \alpha, \beta > 0 \quad (3)$$

na qual Γ representa a função gama.

Distribuição Exponencial Mista a três parâmetros:

$$f_X(x) = \frac{\alpha}{\beta_1} \exp\left(-\frac{x}{\beta_1}\right) + \frac{1-\alpha}{\beta_2} \exp\left(-\frac{x}{\beta_2}\right); \\ x > 0; 0 \leq \alpha \leq 1; \beta_1 > \beta_2 > 0 \quad (4)$$

A estimativa dos parâmetros de todas as distribuições foi feita através do Método da Máxima Verossimilhança, reconhecido como provedor de bons estimadores em amostras com tendências assintóticas. Vale lembrar que em estudos baseados em escalas diárias as séries utilizadas são bastante longas, justificando o emprego do método. As equações dos estimadores são dadas por:

Distribuição Exponencial simples (Woolhiser e Roldán, 1982):

$$\hat{\beta} = \frac{1}{\bar{X}} \quad (5)$$

onde \bar{X} representa a média amostral.

Distribuição Exponencial a dois parâmetros (Johnson e Kotz, 1970):

$$\hat{\alpha} = \min(x_1, x_2, \dots, x_n); \hat{\beta} = \bar{X} - \hat{\alpha} \quad (6)$$

Distribuição Gama a dois parâmetros (Botelho e Moraes, 1999):

$$\hat{\alpha} = \frac{1 + \sqrt{1 + 4 \cdot A/3}}{4 \cdot A}; \hat{\beta} = \frac{\bar{X}}{\hat{\alpha}} \quad (7)$$

Onde

$$A = \ln \bar{X} - \frac{1}{n} \sum_{i=1}^n \ln X_i \quad (8)$$

Distribuição Exponencial Mista a três parâmetros: a dedução analítica do Método da Máxima Verossimilhança não é possível para essa distribuição. Os parâmetros foram estimados com a aplicação de uma técnica iterativa conhecida por algoritmo EM (*Expectation-Maximization*), desenvolvida inicialmente por Dempster *et al.* (1977) e comentada posteriormente em Wilks (2006) como sendo uma ótima solução para problemas que envolvem distribuições estatísticas mistas. Ainda, Katz e Zheng (1999) trazem uma boa descrição do método, com detalhes de sua aplicação.

MÉTODO DE ANÁLISE

A utilização das citadas distribuições estatísticas por diversos autores em vários lugares do mundo não garante que elas sejam realmente apropriadas para modelar as quantidades precipitadas nas bacias dos rios Paraná e Uruguai. Como trabalhos relacionados à aderência destes modelos estatísticos aos dados da região em estudo são escassos, a primeira análise realizada é justamente referente a essa questão. Para tanto, o teste PPCC (*Probability Plot Correlation Coefficient*), proposto por Filliben (1975), é aplicado.

A estatística do teste consiste no cálculo do coeficiente de correlação de Pearson entre os dados observados e as respectivas funções quantis de cada distribuição. Esse coeficiente é utilizado como medida do grau de associação entre duas amostras de mesmo tamanho que, no presente caso, representam os valores observados e os quantis calculados. É determinado segundo a equação (Wilks, 2006):

$$r_{x,y} = \frac{\frac{1}{n-1} \sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})]}{s_x \cdot s_y} \quad (9)$$

onde x e y representam as duas variáveis em consideração, n é o tamanho da amostra, \bar{x} e \bar{y} são as médias de cada amostra e s_x e s_y são os desvios padrão de cada amostra.

A função quantil, por sua vez, é definida como sendo o inverso da função distribuição de probabilidades acumulada (FDA):

$$F^{-1}(p) = x(F) \quad (10)$$

Da mesma forma que uma dada observação x_i quando aplicada à FDA retorna uma probabilidade acumulada p, uma probabilidade qualquer, quando aplicada à função quantil retorna uma quantidade de chuva qualquer. No caso do teste PPCC, as probabilidades aplicadas à função quantil referem-se aos níveis probabilísticos teóricos, definidos genericamente por (Wilks, 2006):

$$p(x_i) = \frac{i-a}{n+1-2a}; i = (1, 2, \dots, n); 0 \leq a \leq 1 \quad (11)$$

onde i representa a ordem da observação e a uma constante a ser determinada de acordo com as distribuições estatísticas em teste. Especificamente para a família exponencial aqui utilizada, Stedinger *et al.* (1992) recomendam $a=0,44$. Os coeficientes de correlação calculados são comparados a valores críticos, determinados conforme formulação de Vogel e McMartin (1991) para um nível de significância de 5%. O modelo é considerado apropriado quando a correlação calculada é superior à crítica.

Ao plotar os resultados da função quantil aplicada aos níveis probabilísticos definidos em *versus* a série observada ordenada, tem-se o gráfico quantil-quantil. Essa representação é aqui utilizada para analisar, visualmente, a qualidade de ajuste proporcionada pelas quatro distribuições probabilísticas. Ambos os eixos do gráfico possuem as mesmas unidades e dimensões, fazendo com que o melhor ajuste seja aquele cujos pontos se aproximem da reta de 1:1. Os coeficientes de correlação para todos os casos são também determinados e comparados.

As últimas análises estão relacionadas com a verificação dos critérios de informação de Akaike (AIC - Akaike, 1974) e de Bayes (BIC - Schwarz, 1978). Essas técnicas não tratam especificamente de testes estatísticos, mas de indicadores que auxiliam na escolha de um modelo sobre outro, com base na função de log-verossimilhança calculada para cada caso. O equacionamento dos critérios é dado por:

$$AIC(k) = -2L_m + 2k \quad (12)$$

$$BIC(k) = -2L_m + k(\ln n) \quad (13)$$

onde k representa o número de parâmetros da distribuição, L_m a função de log-verossimilhança calculada e n o tamanho da amostra. O modelo recomendado será aquele que produzir o menor AIC ou BIC. Analisando as equações (12) e (13) fica evidente que os critérios estão calcados no princípio da parcimônia; a decisão sobre o melhor modelo é feita com base em uma equação que releva a qualidade do ajuste (através das funções de log-verossimilhança) e uma penalidade que aumenta proporcionalmente com o número de parâmetros utilizados. Nota-se também que os dois critérios são semelhantes, diferindo apenas no grau de penalidade considerado. Para amostras relativamente grandes ($n > 100$) o critério BIC mostra-se mais rigoroso.

As equações de log-verossimilhança para as quatro distribuições consideradas neste trabalho são:

Tabela 1 - Estações Pluviométricas Selecionadas

Código da Estação	Estação	Código ANA	Latitude (Oeste)	Longitude (Sul)	Comprimento da série		
					Ano Início	Ano final	Total
MAM	Monte Alegre de Minas	01848000	18°52'20"	48°52'10"	1941	2005	64
UCC	Usina Couro do Cervo	02145007	21°20'37"	45°10'13"	1942	2004	62
MM	Monte Mor	02247058	22°57'39"	47°17'45"	1954	2006	52
Ca	Caiuá	02151035	21°50'00"	51°59'00"	1942	2003	61
To	Tomazina	02349033	23°46'00"	49°57'00"	1938	2006	68
UV	União da Vitória - 396	02651000	26°13'41"	51°04'49"	1938	2006	68
Ta	Taiamã	01655003	16°43'39"	55°31'17"	1964	2007	43
Co	Caracol	02257000	22°01'51"	57°01'45"	1969	2006	37
PM	Passo Marombas	02750009	27°19'51"	50°45'03"	1959	2005	46
LC	Linha Cescon	02753004	27°48'42"	53°01'40"	1960	2006	46
Cq	Cacequi	02954001	29°52'40"	54°49'25"	1944	2005	61

Distribuição Exponencial simples:

$$L_m(x; \beta) = \sum_{i=1}^n \left(-\ln \beta - x \beta^{-1} \right) \quad (14)$$

Distribuição Exponencial a dois parâmetros:

$$L_m(x; \alpha; \beta) = \sum_{i=1}^n \left[-\ln \beta - (x - \alpha)/\beta \right] \quad (15)$$

Distribuição Gama a dois parâmetros:

$$\begin{aligned} L_m(x; \alpha; \beta) &= \\ &= \sum_{i=1}^n \left[(\alpha - 1) \ln x - (x/\beta) - \alpha \ln \beta - \ln \Gamma(\alpha) \right] \end{aligned} \quad (16)$$

Distribuição Exponencial Mista a três parâmetros:

$$\begin{aligned} L_m(x; \alpha; \beta_1; \beta_2) &= \\ &= \sum_{i=1}^n \ln \left[\frac{\alpha}{\beta_1} \exp \left(-\frac{x}{\beta_1} \right) + \frac{1-\alpha}{\beta_2} \exp \left(-\frac{x}{\beta_2} \right) \right] \end{aligned} \quad (17)$$

Ambos os critérios foram calculados para todos os postos e períodos (ver próxima seção) e os modelos indicados determinados.

ÁREA DE ESTUDO

Como dito na introdução deste artigo, os postos escolhidos estão distribuídos dentro das bacias dos rios Paraná e Uruguai, pertencentes à bacia do rio da Prata. Ao todo, foram selecionadas 11 estações pluviométricas, conforme mostram a Tabela 1 e Figura 1.

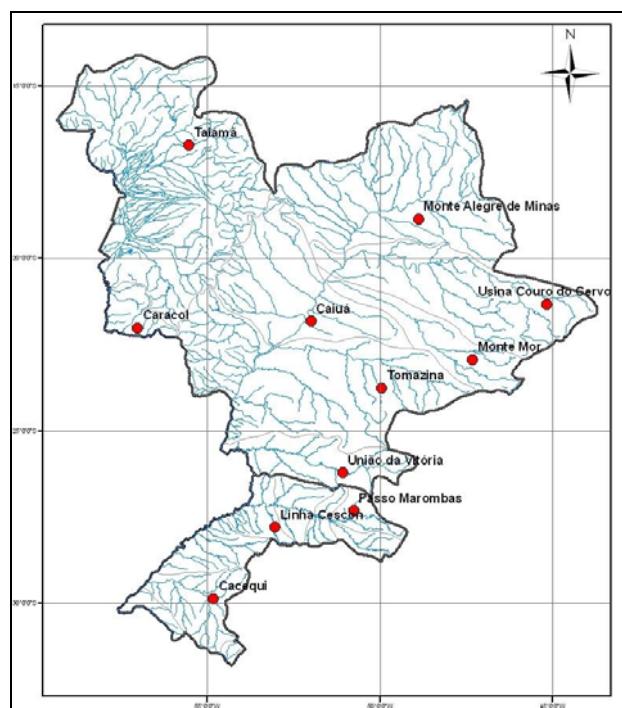


Figura 1 - Área de estudo com as estações selecionadas

Os dados necessários foram obtidos da Agência Nacional de Águas do Brasil (ANA). As eventuais falhas nos registros encontradas referem-se a alguns dias em branco sendo importante ressaltar que, devido à grande variabilidade da chuva quando se trabalha com escalas diárias, nenhum método de preenchimento foi empregado.

Segundo classificação climática de Köppen-Geider em sua forma mais recente (Peel *et al.*, 2007), ambas as bacias possuem clima predominantemente do tipo Cfa. São também encontrados os tipos Cwa e Cwb ao norte da bacia do rio Paraná; a nordeste da mesma bacia, o clima do tipo Aw faz-se presente. Por fim, pontos isolados do norte da bacia do rio Uruguai apresentam clima do tipo Cfb.

O mecanismo gerador das precipitações é fortemente influenciado pela Frente Polar Atlântica. Na porção centro-norte da bacia do rio Paraná, uma grande variedade de tipos climáticos durante o ano é encontrada. A tendência é tempo quente e úmido no verão e quente e seco no inverno, com precipitação média de 2000 mm anuais. Ao sul da bacia do rio Paraná e na bacia do rio Uruguai, por sua vez, observa-se uma discrepância entre os regimes de temperaturas e o pluvial durante o ano. Apesar de uma grande variabilidade térmica existente, a distribuição das precipitações ao longo do ano é regular e as médias ficam entre 1250 mm e 2000 mm. Em todas as regiões, os registros de neve se limitam a poucos pontos serranos fazendo com que o regime pluviométrico possa ser caracterizado somente por chuvas.

Para a aplicação dos métodos de análises, algumas considerações importantes foram adotadas. A primeira se refere ao tratamento dado à sazonalidade; em cada estação pluviométrica, 12 séries foram montadas, coincidentes com os meses do ano. É importante dizer que, dentro de cada mês, assumiram-se séries históricas estacionárias, ou seja, partiu-se do pressuposto que as características estatísticas de longo termo das séries não se alteraram com o tempo. Finalmente, na montagem das séries, o limite de chuva para que um dia fosse considerado úmido foi de 0,3 mm.

RESULTADOS

Segundo a sequência lógica mostrada na apresentação dos métodos de análise, o primeiro resultado é relativo à aderência dos quatro modelos estatísticos aos dados das estações pluviométricas consideradas. A aplicação do teste PPCC produziu

resultados positivos, a um grau de significância de 5%, como mostra a Tabela 2.

Os números de aceites na Tabela 2 referem-se ao total de períodos sazonais por estação. Conforme descrito na seção anterior, adotou-se uma sazonalidade mensal. Com o reduzido número de rejeições, pode-se dizer que as quatro distribuições estatísticas são adequadas para modelagem das precipitações diárias na região em estudo. A partir dessa constatação, as análises comparativas podem ser realizadas naturalmente. Os gráficos quantil-quantil evidenciaram situações interessantes, nas quais o desempenho das quatro distribuições variou sensivelmente. Notou-se que os fatores mais influentes para essas variações foram o tamanho da amostra e a característica das precipitações incidentes em cada posto. A Figura 2 ilustra essa situação, através de quatro exemplos de gráficos quantil-quantil elaborados.

Tabela 2 - Veredictos da aplicação do teste PPCC, com significância de 5%

Código da Estação	Número de aceites			
	E1	E2	G	EM
MAM	12	12	12	12
UCC	12	12	12	12
MM	11	11	12	12
Ca	11	11	11	12
To	11	11	12	12
UV	10	10	11	12
Ta	12	12	12	12
Co	11	11	11	11
PM	12	12	12	12
LC	12	12	11	12
Cq	12	12	12	12

Legenda: E1 – Exponencial Simples; E2 – Exponencial a 2 parâmetros; G – Gama a 2 parâmetros; EM – Exponencial Mista a 3 parâmetros.

No primeiro gráfico (MAM - dezembro) as distribuições foram ajustadas ao período mais chuvoso do estudo. Na maior parte da amostra (chuvas até 40 mm) o desempenho das quatro distribuições é semelhante, mas a partir daí as exponenciais simples e a 2 parâmetros mostram-se inferiores à gama e exponencial mista. Estas, por sua vez, assumem um comportamento oposto: enquanto a gama subestima as intensidades, a exponencial mista as superestima. No segundo gráfico (Ta - julho) a situação é oposta; o ajuste se dá sobre o período mais seco. Nota-se que o desempenho das quatro distribuições é satisfa-

tório, mas com uma leve superioridade da exponencial simples e exponencial mista sobre as demais. É importante perceber que as intensidades das chuvas nesse posto são muito inferiores às do primeiro gráfico.

O terceiro gráfico (UV - dezembro) mostra novamente um caso de amostra grande, porém com incidência de precipitações muito intensas. As duas chuvas de maior intensidade, em torno dos 150 mm, não foram bem representadas por nenhuma distribuição, embora se possa dizer que o desempenho da exponencial mista foi superior às demais. Finalmente, no quarto gráfico (Cq - fevereiro), o ajuste se deu na estação pluviométrica localizada mais ao sul da área de estudo, com resultados muito bons para as quatro distribuições, mesmo com a presença de algumas chuvas de maior intensidade.

Tabela 3 - Preferências baseadas nas correlações dos gráficos quantil-quantil

Código da Estação	E1/E2	G	EM
MAM	0	3	9
UCC	3	1	8
MM	1	6	5
Ca	2	3*	8*
To	2	3	7
UV	1	6	5
Ta	5**	1	7**
Co	2	4	6
PM	4	1	7
LC	7**	1	5**
Cq	2	3	7
Total	29	32	74

Legenda: E1 – Exponencial Simples; E2 – Exponencial a 2 parâmetros; G – Gama a 2 parâmetros; EM – Exponencial Mista a 3 parâmetros.

* distribuições gama e exponencial mista obtiveram a mesma correlação em 1 caso;

** distribuições exponenciais simples/2 parâmetros e exponencial mista obtiveram a mesma correlação em 1 caso

Entende-se, no entanto, que a análise meramente visual dos gráficos quantil-quantil é um tanto subjetiva. Na tentativa de reforçar os resultados provenientes dos ajustes, os coeficientes de correlação de Pearson para todos os casos foram calculados. Como são resultados numerosos, optou-se por elaborar duas tabelas resumidas: Tabela 3, com a indicação do número de vezes que certo modelo estatístico teve a maior correlação com os dados

observados, quando comparados aos outros modelos e Tabela 4, com a média dos coeficientes de correlação obtidos.

A primeira ocorrência a chamar a atenção é que as distribuições exponencial simples e exponencial a 2 parâmetros obtiveram o mesmo coeficiente de correlação para todos os casos. De fato, o desempenho dessas distribuições foi muito semelhante, diferenciando-se apenas para chuvas de maior intensidade, quando a exponencial a 2 parâmetros demonstrou superioridade. Como as ocorrências das chuvas intensas se limitam a poucos dias das séries, essa diferença não se refletiu nos coeficientes de correlação. Outros casos nos quais as correlações calculadas tiveram o mesmo valor ocorreram, conforme atestados na Tabela 3.

A insensibilidade do coeficiente de correlação quanto aos eventos extremos fica ainda mais evidente através dos valores médios mostrados na Tabela 4. Percebe-se claramente que a diferença entre os valores médios é praticamente insignificante. Todas as correlações se mostraram muito boas, com valores próximos à unidade. Vale lembrar que o coeficiente de correlação de Pearson se situa entre os limites $-1 \leq r \leq 1$; quanto mais perto das extremidades, mais forte é a correlação entre as amostras.

Tabela 4 – Médias dos coeficientes de correlações obtidos

Código da Estação	E1/E2	G	EM
MAM	0,9914	0,9930	0,9942
UCC	0,9896	0,9880	0,9921
MM	0,9926	0,9938	0,9944
Ca	0,9912	0,9929	0,9938
To	0,9956	0,9960	0,9968
UV	0,9940	0,9962	0,9962
Ta	0,9932	0,9921	0,9955
Co	0,9891	0,9902	0,9917
PM	0,9957	0,9950	0,9966
LC	0,9949	0,9923	0,9951
Cq	0,9953	0,9953	0,9968

Legenda: E1 – Exponencial Simples; E2 – Exponencial a 2 parâmetros; G – Gama a 2 parâmetros; EM – Exponencial Mista a 3 parâmetros.

Ainda assim, de um modo geral, pode-se destacar a distribuição exponencial mista por mostrar ampla superioridade sobre as demais, produzindo as maiores correlações em 74 dos 132 casos estudados (11 estações pluviométricas e 12 períodos sazonais). Para as demais, os resultados ficaram divididos; a distribuição gama teve melhor desempenho

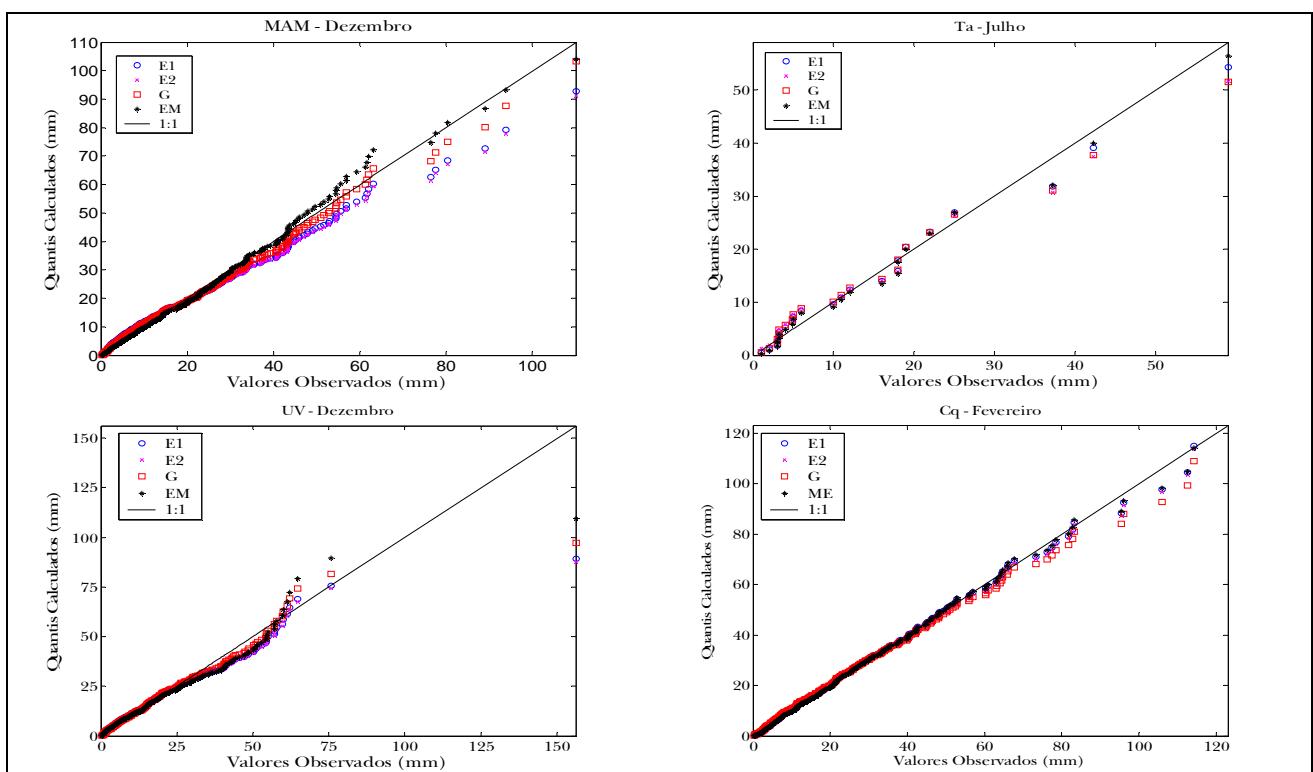


Figura 2 - Exemplos de gráficos quantil-quantil - E1: Exponencial Simples; E2: Exponencial a 2 parâmetros; G: Gama a 2 parâmetros; EM: Exponencial Mista a 3 parâmetros.

em estações de clima bem definido, localizadas na porção centro-norte da região de estudo, enquanto que as exponenciais simples e a 2 parâmetros apresentaram uma boa performance em regiões de climas menos definidos, nos quais as precipitações são bem distribuídas anualmente. Mesmo assim, é possível dar um destaque à distribuição gama por apresentar a capacidade de reprodução de chuvas de maior intensidade que eventualmente incidiram sobre os postos.

Resultados diferentes foram observados na aplicação dos critérios AIC e BIC. De acordo com os resultados apresentados na Tabela 5, a distribuição exponencial a 2 parâmetros teve a grande maioria das indicações em ambos os critérios.

Notou-se que o fator decisivo para esse resultado foi a penalidade que cada critério considera em relação ao número de parâmetros de cada modelo. Foi exatamente por esse fator que a distribuição exponencial a 2 parâmetros teve uma performance superior à exponencial mista, a segunda melhor opção. Entretanto, mesmo com a forte influência da penalidade, a distribuição gama não teve

um bom desempenho, inclusive sobre um modelo estatístico que conta com um parâmetro a mais.

Tabela 5 - Modelos indicados pelos critérios de informação AIC e BIC

Código da Estação	AIC				BIC			
	E1	E2	G	EM	E1	E2	G	EM
MAM	0	9	0	3	0	10	0	2
UCC	0	12	0	0	0	12	0	0
MM	0	8	2	2	0	9	2	1
Ca	0	9	0	3	0	9	0	3
To	0	12	0	0	0	12	0	0
UV	0	4	0	8	0	4	0	8
Ta	0	12	0	0	1	11	0	0
Co	0	9	0	3	6	5	0	1
PM	0	12	0	0	0	12	0	0
LC	0	8	4	0	0	8	4	0
Cq	0	12	0	0	0	12	0	0
Total	0	107	6	19	7	104	6	15

Legenda: E1 – Exponencial Simples; E2 – Exponencial a 2 parâmetros; G – Gama a 2 parâmetros; EM – Exponencial Mista a 3 parâmetros.

DISCUSSÃO

Após a exposição dos resultados das verificações individuais realizadas, é desejável fazer alguns comentários sobre as análises. O modelo de pior desempenho em todos os casos foi a distribuição exponencial simples. É provável que a causa desta constatação esteja relacionada com a simplicidade da formulação de seu equacionamento e, de fato, estudos anteriores mostraram deficiências dessa distribuição na modelagem de chuvas diárias (Todorovic e Woolhiser, 1975). Richardson (1981) fez uma ressalva no tocante ao uso dessa distribuição, alertando que outras distribuições, como a gama e a exponencial mista, ofereciam melhores ajustes. Mesmo assim, a empregou como hipótese simplificadora, pois sua intenção foi simular não somente chuvas diárias, mas temperaturas e radiação solar. A exceção ficou por conta de Haan *et al.* (1976) que obtiveram resultados considerados satisfatórios. A técnica foi dividir as intensidades das chuvas em classes, utilizando a distribuição exponencial simples somente no último grupo, que não contava com limite superior.

Woolhiser e Roldán (1982) elaboraram um estudo semelhante ao presente trabalho, confrontando as distribuições exponencial simples, gama e exponencial mista para a modelagem de chuvas diárias. Porém algumas diferenças são notadas: em primeiro lugar, na consideração sobre sazonalidade, os parâmetros das distribuições foram calculados com base em harmônicos de séries de Fourier. Segundo, além das distribuições estatísticas individuais, foram considerados casos de dependência entre dias chuvosos ou secos próximos, caracterizando uma análise em cadeias. Os ensaios foram baseados somente no critério de informação AIC, aplicados a cinco estações pluviométricas norte americanas. Como resultados, a distribuição exponencial mista se destacou ao obter a quase unanimidade de indicações do critério utilizado. Vale lembrar que, nas considerações em cadeias, o número de parâmetros em cada caso aumenta, refletindo diretamente na penalidade incidente na formulação do AIC.

Embora a distribuição gama tenha mostrado bons resultados em diversos estudos, adquirindo assim a preferência de diversos autores (ver referências na introdução), modelos mistos têm ganhado espaço na modelagem diária de precipitações. Em um dos estudos pioneiros na utilização da exponencial mista, Smith e Schreiber (1974), após a elaboração do gráfico da distribuição empírica de probabilidades para a precipitação em três postos pluviomé-

tricos, notaram que todas as curvas tinham um ponto de inflexão. Assim, o modelo adotado no estudo foi composto pela soma de duas distribuições exponenciais simples e um parâmetro complementar de probabilidade da incidência da primeira ou segunda função. Nessa linha de pensamento, Wilks (2006) ressalta que alguns mecanismos físicos apresentam mais de um processo gerador, fazendo com que sua representação por uma distribuição estatística simples seja incompleta. Por isso, distribuições mistas trazem um bom grau de flexibilidade aos modelos, refletindo diretamente nos resultados a serem obtidos.

A última consideração refere-se à distribuição exponencial a 2 parâmetros. Modelos de precipitação diária que utilizam essa distribuição estatística são raros; Nascimento e Kelman (1995) foram a única referência encontrada. Não obstante, os autores apontaram estudos anteriores (ELETROBRÁS-CEPEL, 1987 *apud* Nascimento e Kelman, 1995) que indicaram plena preferência por distribuições limitadas a 2 parâmetros, em detrimento aos modelos com elevado número de parâmetros. Em concordância com suas considerações, o modelo de geração, estruturado com a distribuição exponencial a 2 parâmetros e aplicado à região sul do Brasil (bacia do rio Uruguai), obteve resultados excelentes. Nota-se que a região de aplicação desse modelo faz parte da área de estudo do presente artigo.

CONCLUSÕES E RECOMENDAÇÕES

Um estudo comparativo entre distribuições estatísticas foi apresentado, na intenção de mostrar os pontos fortes e fracos desses modelos quando aplicados à geração de séries sintéticas de precipitações diárias. Após a aplicação do teste PPCC (Filliben, 1975), as distribuições exponencial simples (a 1 parâmetro), exponencial a 2 parâmetros, gama e exponencial mista se mostraram adequadas para tal objetivo.

A comparação foi feita com base em gráficos quantil-quantil e seus respectivos coeficientes de correlação, além da aplicação dos critérios de informação AIC e BIC. A região à qual a análise foi aplicada refere-se à porção brasileira do rio da Prata, ou seja, as bacias dos rios Paraná e Uruguai. Como notado, os resultados indicados por essas técnicas não foram coincidentes. Nos métodos baseados na função quantil, a preferência foi dada à distribuição exponencial mista, enquanto nos critérios de infor-

mação, o modelo exponencial a 2 parâmetros foi indicado como mais apropriado.

Esses resultados contraditórios não devem ser considerados de forma individual. No caso da elaboração de um modelo de geração, é necessária uma avaliação sobre quais são as ambições do estudo. O anseio de resultados mais ou menos rigorosos, a intenção de simulações voltadas a eventos extremos, ou ainda o nível de complexidade final da estrutura do modelo são exemplos de fatores que podem influenciar sobre a escolha de determinada distribuição estatística sobre outra.

No caso da elaboração de um estudo multi-variado, a análise sobre o desempenho dos modelos estatísticos torna-se de extrema importância. A escolha ótima recairia sobre a distribuição com melhor desempenho global, haja vista que um modelo estruturado com diversas funções probabilísticas elevaria sua complexidade enormemente, tornando-o inviável.

Dessa forma, recomenda-se a utilização de critérios parcimoniosos para a escolha do modelo estatístico mais apropriado a cada caso.

AGRADECIMENTOS

A pesquisa que levou a esses resultados recebeu fundos do Sétimo Programa de Estudos da Comunidade Europeia (FP7/2007-2013) sob acordo de concessão nº 212492 (CLARIS LPB – Um Grupo de Trabalho Europeu – Sul-americano para Avaliação de Mudanças Climáticas e Estudos de Impactos na Bacia do Rio da Prata).

A segunda autora contou também com o apoio da Bolsa Produtividade em pesquisa do CNPq.

REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, Brookline, AC-19, 6, p. 716-723, 1974.
- BOTELHO, V. A. V. A.; MORAIS, A. R. de. Estimativas dos parâmetros da distribuição gama de dados pluviométricos do município de Lavras, Estado de Minas Gerais. *Ciências Agrotécnicas*, Lavras, 23, 3, p. 697-706, 1999.
- DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via EM algorithm.

Journal of the Royal Statistical Society, Gainesville, B-39, p. 1-38, 1977.

- DETZEL, D. H. M.; MINE, M. R. M. Geração estocástica de séries sintéticas de precipitações em escala diária: revisão dos principais modelos e métodos. In.: Simpósio Brasileiro de recursos Hídricos, 18, 2009, Campo Grande, p. 1-18, CD-Rom.
- FILLIBEN, J. J. The probability plot correlation coefficient test for normality. *Technometrics*, 17, 1, p. 111-117, 1975.
- FOUFOULA-GEORGIOU, E.; LETTENMAIER, D. P. A Markov renewal model for rainfall occurrences. *Water Resources Research*, Washington, DC, 23, 5, p. 875-884, 1987.
- GOMES, A. C. et al. Modelagem da Precipitação Diária na Região de Santiago, RS. In: Congresso Brasileiro de Meteorologia, 14, 2006, Florianópolis. Disponível em: <<http://www.creatividadecoletiva.net/cbm-files/14-8180626eb34c9e5c9de98...>> a7a77af621c.pdf>. Acesso em: 12/08/2009.
- GRONDONA, M. O. et al. A Stochastic Precipitation Generator Conditioned on ENSO Phase: A Case Study in Southeastern South America. *Journal of Climate*, Boston, 13, 16, p. 2973-2986, 2000.
- HAAN, C. T.; ALLEN, D. M.; STREET, J. O. A Markov Chain Model of Daily Rainfall. *Water Resources Research*, Washington, DC, 12, 3, p. 443-449, 1976.
- JOHNSON, N. L.; KOTZ, S. Exponential Distributions. In: _____. *Distributions in Statistics: continuous univariate distributions I*. New York: Wiley-Interscience Publication 1970. p. 207-227.
- KATZ, W. R. Precipitation as a chain-dependent process. *Journal of Applied Meteorology*, Boston, 16, 7, p. 671-676, 1977.
- KATZ, W. R.; ZHENG, X. Mixture model for overdispersion of precipitation. *Journal of Climate*, Boston, 12, p. 2528-2537, 1999.
- KRÜGER, C. M.; KAVISKI, E.; MÜLLER, I. I. Modelagem estocástica da precipitação diária no Estado do Paraná. *Revista Brasileira de Recursos Hídricos*, Porto Alegre, 3, 2, p. 87-94, 1998.
- LIAO, Y.; ZHANG, Q.; CHEN, D. Stochastic modeling of daily precipitation in China. *Journal of Geographical Sciences*, Pequim, 14, 4, p. 417-426, 2004.
- NASCIMENTO, C. E. S.; KELMAN, J. Um modelo para geração estocástica de chuvas diárias. *RBE – Caderno de Recursos Hídricos*, Rio de Janeiro, 13, 2, p. 19-36, 1995.
- PEEL, M. C.; FINLAYSON, B. L.; McMAHON, T. A. Updated world map of the Köppen-Geider climate classification. *Hydrology and Earth System Sciences*, Delft, 11, p. 1633-1644, 2007.
- RICHARDSON, C. W. Stochastic Simulation of Daily Precipitation, Temperature, and Solar Radiation. *Water Re-*

- sources Research, Washington, DC, 17, 1, p. 182-190, 1981.
- SCHWARZ, G. Estimating the Dimension of a Model. *The Annals of Mathematical Statistics*, Bethesda, 6, 2, p. 461-464, 1978.
- SMITH, R. E.; SCHREIBER, H. A. Point process of seasonal thunderstorm rainfall – 2. Rainfall depth probabilities. *Water Resources Research*, Washington, DC, 10, 3, p. 418-423, 1974.
- SRIKANTHAN, R.; McMAHON, T. A. Stochastic generation of annual, monthly and daily climate data: A review. *Hydrology and Earth System Sciences*, Delft, 5, 4, p. 653-670, 2001.
- STEDINGER, J. R.; VOGEL, R. M.; FOUFOULA-GEORGIOU, E. Frequency Analysis of Extreme Events. In: _____. *Handbook of Hydrology*. New York: McGraw-Hill, Inc. 1992, cap. 18.
- TODOROVIC, P.; WOOLHISER, D. A. A Stochastic Model of n-Day Precipitation. *Journal of Applied Meteorology*, Boston, 14, p. 17-24, 1975.
- VOGEL, R. M. e McMARTIN, D. E. Probability plot goodness-of-fit and skewness estimation procedures for the Pearson type III distribution, *Water Resources Research*, Washington, DC, 27, 12, p. 3149-3158, 1991.
- WILKS, D S. Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology*, Amsterdam, 210, p. 178-191, 1998.
- WILKS, D.S. *Statistical methods in the atmospheric sciences*. 2nd. ed. San Diego: Academic Press, 2006.
- WOOLHISER, D. A.; ROLDÁN, J. Stochastic Daily Precipitation Models - 2. A Comparison of Distribution Amounts. *Water Resources Research*, Washington, DC, 18, 5, p. 1461-1468, 1982.

Modeling Daily Precipitations: Comparative Analysis

ABSTRACT

Stochastic models to generate synthetic rainfall series are useful tools for several practical applications. Parametric formulations are most popular among those that work with daily scales, mainly due to their considerable flexibility. Their structure is divided into: (i) determination of events, using stochastic processes and (ii) determination of amounts, using statistical distributions. This paper focuses on phase (ii) and on the specific analysis of the most used probabilistic distributions: simple exponential, two-parameter exponential, two parameter gamma and three parameter mixed exponential. The aim is to investigate the behavior of these distributions at eleven rain gauging stations spread throughout the Paraná and Uruguay rivers basins, using PPCC statistical test application, comparative analysis based on Q-Q plots and AIC and BIC information criteria determination for each case.

Key-words: Stochastic models; synthetic series; test PPCC.