Revista da Estatística UFOP, Vol III(3), 2014, ISSN 2237-8111
Edição especial: 59ª Reunião Anual da Regional Brasileira da Sociedade Internacional de Biometria - RBRAS
Trabalhos apresentados na versão poster.

Ajustes de modelos probabilísticos para ocorrências de precipitação pluvial em Araras e Piracicaba, SP: um estudo comparativo

Idemauro Antonio Rodrigues de Lara [1]
Renata Alcarde Sermarini [2]
Sônia Maria De Stefano Piedade [3]
Miguel Angelo Maniero [4]

1 Introdução

A precipitação pluvial é uma variável agrometeorológica relevante e de impacto imediato sobre o clima, agricultura e economia de uma região. Em particular, a produtividade agrícola de larga escala depende muito da distribuição temporal das chuvas, o que tem motivado especialistas na área a buscarem alguns modelos teóricos e empíricos para descreverem ao menos aproximadamente a distribuição destes dados em determinados intervalos de tempo. Do ponto de vista estatístico, dados sobre precipitação pluvial podem ser representados por meio de uma variável aleatória contínua que tem uma função de densidade de probabilidade. Algumas funções de densidade de probabilidade clássicas como a gama, exponencial, log-normal são frequentemente empregadas para modelar eventos desta natureza (Cruciani et al., 2002).

Frequentemente tem-se que ajustar mais de uma distribuição, que pode variar conforme a estação do ano ou a região (Silva et al., 2007; Silva et al., 2013). Sendo assim, o uso de uma família de funções de densidade pode representar de forma simplificada a ocorrência do fenômeno, independente da época. A família Tweedie pertence a uma classe de modelos denominada modelos exponenciais de dispersão. Tais modelos, bem como os modelos próprios de dispersão, compõem os modelos de dispersão, introduzidos por Jørgensen (1997), com o objetivo de estender a metodologia já conhecida para os modelos lineares generalizados.

A teoria dos modelos de dispersão engloba tanto estatística quanto probabilidade e dispõe de uma coleção de métodos relacionados às distribuições das famílias exponenciais, para as quais os conceitos de locação e escala podem ser generalizados pelos conceitos de posição e dispersão. A família Tweedie pertence a uma classe de modelos denominada modelos exponenciais de dispersão. Zocchi et al (2007) utilizaram a família Tweedie de distribuições para tratar a modelagem do número de eventos de precipitações (distribuição Poisson) e do volume de chuvas mensais em Piracicaba (distribuição gama), durante o período de 1917 a 2006, excluindo-se os anos de 1971, 1972 e 1973. Os autores tomaram como base a metodologia já apresentada por Dunn e White (2005) para dados pluviais, evidenciando, assim, a eficiência desta classe de modelos para descrição de eventos desta

[1]Departamento de Ciências Exatas, ESALQ/USP, e-mail: *idemauro@usp.br*
[2]Departamento de Ciências Exatas, ESALQ/USP, e-mail: *ralcarde@usp.br*
[3]Departamento de Ciências Exatas, ESALQ/USP, e-mail: *soniamsp@usp.br*
[4]Departamento de Recursos Naturais, CCA/UFSCar, e-mail: *miguel@cca.ufscar.br*

natureza. Apesar da importância desta classe de modelos, tanto do ponto de vista conceitual, quanto do ponto de vista prático, ainda a sua utilização é pequena nas diversas áreas do saber, e, em particular, em estudos relacionados às variáveis climatológicas e ambientais. Neste contexto, o objetivo deste trabalho é apresentar e reforçar o uso a família Tweedie de distribuições, bem como selecionar modelos probabilísticos pertinentes, por meio da maximização do perfil de verossimilhança, apresentando um estudo comparativo com os dados sobre precipitação pluvial das cidades Araras e Piracicaba.

2 Material e Métodos

2.1 Material

Os dados foram obtidos nas estações do Posto Agrometeorológico, do Departamento de Biossistemas, da Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ/USP), em Piracicaba SP, e na base de dados climatológicos do Departamento de Recursos Naturais e Proteção Ambiental, do Centro de Ciências Agrárias da Universidade Federal de São Carlos (UFSCar/Araras), correspondendo a uma série histórica entre os anos de 1972 e 2012. Destas bases de dados climatológicos, selecionaram-se as variáveis referentes ao número de dias com precipitação e total de precipitação pluvial diária.

2.2 Métodos

A seguir, faz-se uma breve descrição dos modelos de dispersão, da família Tweedie de distribuições e extensões. Um modelo de dispersão, MD (μ, σ^2), com parâmetro de locação μ e parâmetro de dispersão σ^2 pertence à família de distribuições cuja função densidade de probabilidade pode ser escrita na forma:

$$p(y, \mu, \sigma^2) = a(y; \sigma^2) \exp\left[-\frac{1}{2\sigma^2} d(y, \mu) \right], \quad y \in C \tag{1}$$

em que $\Omega \subseteq C \subseteq \mathbb{R}$, $a(y; \sigma^2)$ é uma função monótona apropriada, d é a componente do desvio, $\mu \in \Omega$ e $\sigma^2 > 0$. Os modelos de dispersão são classificados em dois grupos principais, que não são mutuamente exclusivos, os modelos próprios de dispersão PD(μ, σ^2), e os modelos exponenciais de dispersão ED(μ, σ^2). Uma distribuição probabilística é chamada de modelo exponencial de dispersão (ED) se o componente do desvio da função (1) toma a forma:

$$d(y, \mu) = yf(\mu) + g(\mu) + h(y)$$

para funções $f(.)$, $g(.)$ e $h(.)$ apropriadas. Uma classe importante de modelos exponenciais de dispersão é aquela em que as funções de variância são da forma:

$$V(\mu) = \mu^p, \quad p \subset \{(-\infty, 0) \cup [1, +\infty)\}$$

ou seja, o parâmetro p corresponde a uma potência da média, caracterizando a função de variância, a qual, por sua vez, identifica a distribuição. Essa classe de modelos é chamada família Tweedie de distribuições, denotada por Tw(μ, σ^2). As distribuições pertencentes a essa família estão resumidas na Tabela 1.

Tabela 1: Distribuições da família exponencial Tweedie e respectivos valores de p.

Distribuição	p	$\mathcal{D}$	Ω
Extrema Estável	$p < 0$	$\mathbb{R}$	$\mathbb{R}_+$
Normal	$p = 0$	$\mathbb{R}$	$\mathbb{R}$
(não existe)	$0 < p < 1$		
Poisson	$p = 1$	$\mathbb{N}_0$	$\mathbb{R}_+$
Poisson Composta	$1 < p < 2$	$\mathbb{R}_0$	$\mathbb{R}_+$
Gama	$p = 2$	$\mathbb{R}_+$	$\mathbb{R}_+$
Positiva Estável	$2 < p < 3$	$\mathbb{R}_+$	$\mathbb{R}_+$
Normal Inversa	$p = 3$	$\mathbb{R}_+$	$\mathbb{R}_+$
Positiva Estável	$p > 3$	$\mathbb{R}_+$	$\mathbb{R}_+$
Extrema Estável	$p = \infty$	$\mathbb{R}$	$\mathbb{R}$

A fim de ilustrar a importância desta classe de modelos para dados climatológicos, mostra-se a construção do modelo Poisson Composto. Considere que a variável aleatória X_{ijk} representa a precipitação pluviométrica no i-ésimo dia, $i = 1, 2, \ldots, N_{jk}$, do j-ésimo mês no k-ésimo ano. Assumindo que cada variável X_{ijk} tem distribuição gama com parâmetros de forma e de escala, $-\alpha$ e $-\gamma$, respectivamente, cuja função de densidade de probabilidade é dada por:

$$f_{X_{ijk}}(x_{ijk}, \alpha, \gamma) = \frac{(-\gamma)^{-\alpha}}{\Gamma(-\alpha)} x_{ijk}^{-(\alpha+1)} \exp(\gamma\, x_{ijk}), \ \ \forall \ x_{ijk} > 0, \ \alpha < 0, \ \gamma < 0.$$

Logo, a variável aleatória $Y_{jk} = \sum_{i=1}^{N_{jk}} X_{ijk}$ denota a precipitação total mensal do j-ésimo mês do k-ésimo ano, sendo N_{jk} uma variável aleatória com distribuição Poisson de parâmetro λ. Então, a variável aleatória $Y_{jk} \mid N_{jk}$ também tem distribuição gama, com parâmetros $-N_{jk}\alpha$ e $-\gamma$. Registra-se que os parâmetros da distribuição de $Y_{jk} \mid N_{jk}$ têm interpretações práticas, em que λ denota o número médio de eventos com precipitação por mês e γ é o parâmetro relativo à quantidade de chuva. Por fim, se, agora, denotarmos por simplicidade o parâmetro de dispersão da equação 1 por $\phi = \sigma^2 > 0$, a distribuição conjunta (Y_t, N_t) é dada por:

$$f_{Y_t N_t}(y_t, n_t) = \frac{[(\phi^{1-\alpha}) t \kappa_{(\alpha,\lambda)}(-\frac{1}{y_t})]^n}{\Gamma(-n\alpha) n! y_t} \exp(\gamma y_t - (\phi^{1-\alpha}) t \kappa_{(\alpha,\lambda)} \gamma), \tag{2}$$

e

$$P(N_t = 0) = P(Y_t = 0) = \exp[-(\phi^{1-\alpha}) t \kappa_{(\alpha,\lambda)} \gamma]. \tag{3}$$

As equações 2 e 3 demonstram que distribuição Poisson Composta tem uma importância singular na modelagem da precipitação pluvial, sobretudo em regiões e períodos, em que os dados de natureza contínua apresentam "excessos" de zeros. Essa característica é usual na mesorregião de Piracicaba, sobretudo no período de inverno (meses de maio a agosto), em que há uma tendência natural para a não ocorrências de chuvas em diversos dias, fazendo-se, então, necessária uma distribuição de probabilidade adequada para acomodar esta inflação de zeros. Para a seleção de modelos da família Tweedie de distribuições adequados para descrever os dados mensais das duas localidades, observaram-se os valores estimados para o parâmetro p, conforme apresentado na Tabela 1. Para a sua estimação foi utilizado o método da máxima verossimilhança perfilada. A análise dos dados foi feita com o auxílio do pacote Tweedie exponential family models: The Tweedie Package, v. 1.2 para o programa R (Dunn, 2006).

3 Resultados e Discussão

Na Figura 1 é apresentado o volume de chuva observado para os meses de maio a agosto, com relação às cidades de Araras e Piracicaba, no período de 1972-2012, bem como os modelos ajustados, sendo estes o modelo Poisson Composto. Verifica-se que o modelo ajustou-se bem aos dados nas duas localidades.

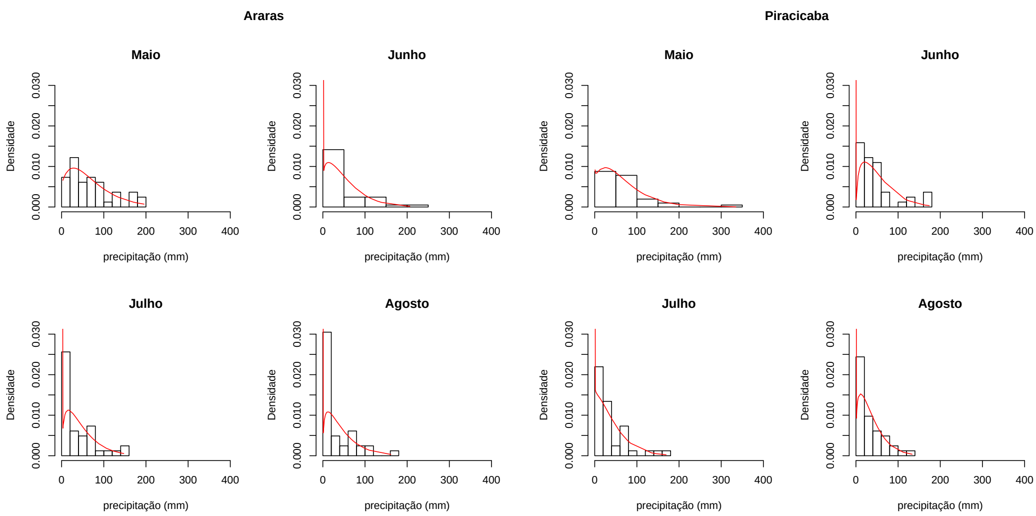

Figura 1: Volume de chuva observado e curvas ajustadas para os meses de maio a agosto, para as duas localidades

Tabela 2: Estimativas dos parâmetros do modelo Tweedie para os dados de precipitação mensal nas duas localidades, no período de 1972-2012.

	Araras			Piracicaba		
	$\hat{p}$	$\hat{\lambda}$	$\hat{\gamma}$	$\hat{p}$	$\hat{\lambda}$	$\hat{\gamma}$
janeiro	1,63	-29,40	16,66	1,80	32,21	-28,16
fevereiro	1,39	-20,80	5,91	1,55	7,64	-28,03
março	1,80	-41,46	14,38	2,04	-0,84	-46,61
abril	2,04	-33,42	-54,35	1,55	5,03	-18,58
maio	1,80	-33,59	8,08	1,71	5,46	-30,23
junho	1,45	-19,10	1,91	1,35	1,94	-12,90
julho	1,37	-15,35	1,40	1,52	2,06	-18,92
agosto	1,41	-18,95	1,14	1,41	1,71	-12,40
setembro	1,39	-15,41	2,83	1,55	3,96	-19,60
outubro	1,47	-17,90	6,00	1,55	7,17	-20,02
novembro	1,63	-22,87	11,32	1,06	5,04	-1,82
dezembro	1,63	-24,65	15,27	0,00	4,61	20,87

Verifica-se por meio da Tabela 2, com base nos calores estimados para o parâmetro p, que o modelo Poisson Composto é o mais indicado para descrever o volume de chuva nas duas localidades, na maior parte dos meses. Nos mês de abril (Araras) e março (Piracicaba) o modelo sugerido é, possivelmente, gama. Em Piracicaba, para o mês de dezembro, o modelo sugerido é o modelo normal. Apesar na similaridade na escolha dos modelos nas duas regiões, registra-se, também pela Tabela 2, diferenças nas estimativas dos demais parâmetros relacionados ao número de dias com ocorrência de chuva ($\hat{\lambda}$) e ao volume de chuva ($\hat{\gamma}$), devido ao fato histórico de na cidade de Araras haver uma maior incidência de chuvas do que na cidade de Piracicaba.

4 Conclusões

Apesar dos dados deste trabalho não se tratarem de uma série histórica grande, verificou-se que nos meses mais úmidos há distribuições com menor grau de assimetria, já nos meses mais secos (abril, maio, junho, julho e agosto) ocorreu uma maior assimetria. Nos meses mais secos, ainda, verificou-se uma alta dispersão relativa (coeficiente de variação) e incidência de zeros. No tocante à aplicação da metodologia proposta, ressalta-se a eficiência para a modelagem simultânea do número de dias com chuva e do total de precipitação para cada um dos meses, o que permitiu caracterizar o comportamento histórico da precipitação pluvial nas duas cidades. A distribuição Poisson Composta foi a de "melhor ajuste", de acordo com o procedimento de seleção baseado no perfil de verossimilhança, mostrando-se eficiente na modelagem de dados contínuos com a presença de zeros.

Referências

[1] CRUCIANI, D.E.; MACHADO, R.E.; SENTELHAS, P.C. **Modelos da distribuição temporal de chuvas intensas em Piracicaba, SP**. Revista Brasileira de Engenharia Agrícola e Ambiental, v.6, n.1, p.76-82, 2002.

[2] DUNN, Peter K. **Tweedie exponential family models: The Tweedie Package**. V.1.2, June. 2006.

[3] DUNN, P. K.; WHITE, N. **Power-variance models for modelling rainfall**. In: 20th International Workshop on Statistical Modelling, 2005.

[4] JØRGENSEN, B. **The Theory of Dispersion Models**. Monographs on Statistics and Applied Probability 76. Chapman & Hall. 1997.

[5] SILVA, Í. N.; OLIVEIRA, J. B.; FONTES, L. O.; ARRAES, F. D. D. Distribuição de frequência da chuva para região Centro-Sul do Ceará, Brasil. **Revista Ciência Agronômica**, v. 44, n. 3, p. 481-487, jul-set, 2013.

[6] SILVA, J. C.; HELDWEIN, A. B.; 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.

[7] R Core Team (2012). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

[8] ZOCCHI, S. S.; JANEIRO, V.; DEMÉTRIO, C. G. B.; **Modelagem simultânea do número de precipitações e do volume de precipitações mensais em Piracicaba, SP, usando a distribuição Tweedie**. Anais do 12^o SEAGRO, 2007.