

MODELOS DE PROBABILIDADE DA FAMÍLIA TWEEDIE PARA OCORRÊNCIAS DE PRECIPITAÇÃO PLUVIAL EM ARARAS SP

Renata Alcarde SERMARINI¹
Idemauro Antonio Rodrigues de LARA¹
Sônia Maria De Stefano PIEDADE¹
Miguel Angelo MANIERO²

- RESUMO: A precipitação pluvial é uma variável meteorológica relevante e de impacto imediato sobre o clima, agricultura e economia de uma região. A caracterização dessa variável é feita por meio de um modelo probabilístico, exigindo o ajuste de diferentes distribuições, conforme o período e a localização do evento. Por outro lado, 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 e podem ser empregados para modelar eventos relacionados à precipitação pluvial. Nesse contexto, o objetivo deste trabalho é aplicar estes métodos para ajustar distribuições da família Tweedie aos dados sobre precipitação pluvial da cidade de Araras-SP, para cada um dos doze meses do ano, tomando como base os dados oficiais da estação meteorológica do Centro de Ciências Agrárias da Universidade Federal de São Carlos, durante o período de 1972 a 2012. Verificou-se que a metodologia é apropriada para a representação desta variável e que a distribuição Poisson Composta foi a de melhor ajuste para a maior parte dos meses. A análise dos dados foi feita com o auxílio do pacote *Tweedie exponential family models*, disponível no software R.
- PALAVRAS-CHAVE: Precipitação pluvial; modelos probabilísticos; modelo de dispersão exponencial Tweedie.

¹Universidade de São Paulo – USP, Escola Superior de Agricultura “Luiz de Queiroz” – ESALQ, Departamento de Ciências Exatas, CEP: 13418-900, Piracicaba, SP, Brasil. E-mail: ralcarde@usp.br; idemauro@usp.br; soniamsp@usp.br

²Universidade de São Carlos – UFSCar, Departamento de Recursos Naturais e Proteção Ambiental, CEP: 13565-905, São Carlos, SP, Brasil. E-mail: miguel@cca.ufscar.br

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). A título de ilustração, Blain et al. (2007) utilizaram o modelo gama para descrever dados de precipitação pluvial da região de Campinas no período de 1890 a 2005. Silva et al. (2007) deram destaque para os ajustes das distribuições gama e Weibull, para caracterização da chuvas em Santa Maria-RS. A distribuição beta foi eficiente para representar a precipitação pluvial em mesorregiões no estado do Pernambuco (Silva et al., 2010). Pizzato et al (2012), fez uso da distribuição gama incompleta para a modelagem das chuvas na região de Cáceres-MT. Na região Centro-Sul do Ceará a distribuição Gumbel foi utilizada para descrever o período chuvoso, enquanto que a distribuição exponencial melhor se ajustou para o período de seca (Silva et al., 2013). Na literatura da área também tem merecido destaque a modelagem de valores extremos, sejam de máximos ou de mínimos, já que a ocorrência desses eventos está relacionada com inundações e secas que, em geral, representam prejuízos a diversos setores da sociedade, em particular, para a agricultura (Cruciani et al, 2002; Sansigolo, 2008; Santos et al., 2009; Aragão et al., 2013). 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. Os modelos lineares generalizados (MLG) foram propostos por Nelder e Wedderburn (1972) com o propósito de unificar uma série de técnicas disponíveis para análise de dados, discretos ou contínuos, desde que a distribuição inerente pertença à família exponencial. Outra característica dos MLG é que a função de variância, monótona e derivável, depende da média. Essa teoria permitiu a generalização de métodos de análise para dados “não normais” por meio da técnica da análise de desvio, fundamentada na teoria da verossimilhança. Os fundamentos dos MLG foram, por sua vez, estendidos para outras classes de modelos, como os modelos de quasi-verossimilhança (Wedderburn, 1974) e os modelos de dispersão, introduzidos por Jørgensen (1997), ambos com o objetivo de estender a metodologia já conhecida dos modelos lineares generalizados, sendo fundamentalmente caracterizados pela função de variância. 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 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 uma aplicação aos dados sobre precipitação pluvial de Araras SP.

2 Material e métodos

2.1 Material

O município de Araras está localizado no interior do estado de São Paulo, na mesorregião de Piracicaba, a uma latitude $22^{\circ}21'25''$ sul e a uma longitude $47^{\circ}23'03''$ oeste, estando a uma altitude de 646 metros. Quanto ao clima, predomina o tropical de altitude, caracterizado por invernos de temperaturas amenas e secos e verões quentes e úmidos. Os dados referentes à precipitação pluvial foram obtidos 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), em uma estação convencional, cuja base de dados é de domínio público. Os dados selecionados compreendem as alturas de precipitações mensais, em mm, de uma série histórica entre os anos 1972 e 2012.

2.2 Modelagem estatística

Da base de dados climatológicos de Araras-SP, selecionaram-se as variáveis referentes ao número de dias com precipitação e total de precipitação pluvial diária. A estes dados foram verificados ajustes às distribuições pertencente à família Tweedie, considerando-se cada mês, ou seja, objetivou-se a caracterização de cada mês do período correspondente da série histórica.

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:

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

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 $\text{PD}(\mu, \sigma^2)$, e os modelos exponenciais de dispersão $\text{ED}(\mu, \sigma^2)$. Uma distribuição probabilística é chamada de modelo exponencial de dispersão (ED) se o componente do desvio da função (1) toma a forma:

$$d(y, \mu) = yf(\mu) + g(\mu) + h(y) \quad (2)$$

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

$$V(\mu) = \mu^p, \quad p \subset \{(-\infty, 0) \cup [1, +\infty)\} \quad (3)$$

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 $\text{Tw}(\mu, \sigma^2)$. As distribuições pertencentes a essa família estão resumidas na Tabela 1. Na Figura 1, observa-se a forma da distribuição para alguns valores de p .

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 (não existe)	$p = 0$ $0 < p < 1$	\mathbb{R}	\mathbb{R}
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, $i = 1, 2, \dots, N_{jk}$, 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:

$$f_{X_{ijk}}(x_{ijk}, \alpha, \gamma) = \frac{(-\gamma)^{-\alpha}}{\Gamma(-\alpha)} x_{ijk}^{-(\alpha+1)} \exp(\gamma x_{ijk}), \quad \forall x_{ijk} > 0, \alpha < 0, \gamma < 0. \quad (4)$$

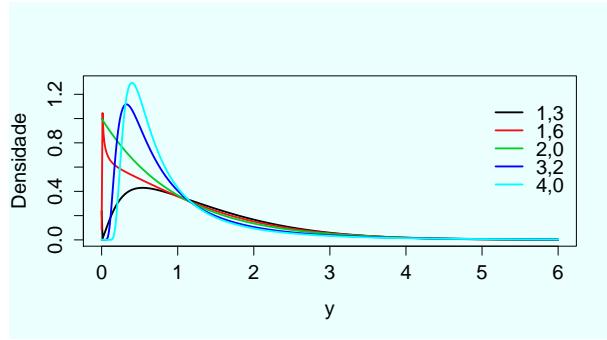


Figura 1 - Distribuições Tweedie para variações do valor de p .

Logo, a variável aleatória:

$$Y_{jk} = \sum_{i=1}^{N_{jk}} X_{ijk} \quad (5)$$

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 . Na definição da função de densidade (4) os parâmetros γ e α são considerados negativos a fim de que atendam às condições de definição da família Tweedie de distribuições. Então, a variável aleatória $Y_{jk} | N_{jk}$ também tem distribuição gama, com parâmetros $-N_{jk}\alpha$ e $-\gamma$. Adicionalmente registra-se que a expressão (5) é um processo estocástico, indexado pelo parâmetro $t = jk$, ou seja, $\{Y_t; t \geq 0\}$, denominado processo de Poisson Composto, uma vez que $\{N_t; t \geq 0\}$ é um processo de Poisson e as variáveis aleatórias, X_{ijk} , são independentes e identicamente distribuídas, cada qual com a função de densidade de probabilidade descrita em (4). Por outro lado, a função geradora de cumulantes de $Y_{jk} | N_{jk}$ é dada por:

$$\kappa_{(\alpha, \lambda)} = \frac{\alpha - 1}{\alpha} \left(\frac{\lambda}{\alpha - 1} \right)^\alpha \quad (6)$$

Se, agora, considerarmos a função geradora de cumulantes da família ED (equação 1) e, comparando-a com a função geradora de cumulantes da equação (6), tem-se as seguintes relações envolvendo os parâmetros da família Tweedie:

$$\lambda = \frac{1}{\phi} \frac{\mu^{2-p}}{2-p}, \quad \alpha = \frac{2-p}{1-p} \quad \text{e} \quad \gamma = \phi(1-p)\mu^{p-1}$$

em que λ denota o número médio de eventos com precipitação por mês e γ é o parâmetro relativo à quantidade de chuva. Portanto, os parâmetros da distribuição de $Y_{jk} | N_{jk}$ têm interpretações práticas. 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:

$$f_{Y_t N_t}(y_t, n_t) = \frac{[(\phi^{1-\alpha}) t \kappa_{(\alpha, \lambda)}(-\frac{1}{y_t})]^n}{\Gamma(-n\alpha) n! y_t} \exp(\gamma y_t - (\phi^{1-\alpha}) t \kappa_{(\alpha, \lambda)} \gamma), \quad (7)$$

e

$$P(N_t = 0) = P(Y_t = 0) = \exp[-(\phi^{1-\alpha}) t \kappa_{(\alpha, \lambda)} \gamma]. \quad (8)$$

As equações 7 e 8 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 de precipitação de Araras-SP, 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, que consiste inicialmente em estimar os parâmetros de perturbação (sem interesse imediato) pelo método da máxima verossimilhança, para valores fixos dos parâmetros de interesse. Posteriormente, os parâmetros de perturbação da função de verossimilhança são substituídos pelos valores estimados, para enfim ocorrer a estimação dos parâmetros de interesse. Trata-se, portanto, de um processo iterativo em duas etapas, que comuta entre atualizar os valores dos parâmetros de perturbação e dos parâmetros de interesse. Empiricamente, este método consiste em obter o valor que maximiza a função de verossimilhança para diversas alternativas de valores para p pela inspeção do gráfico do “perfil de verossimilhança”. Esse procedimento permite selecionar modelos com melhor ajuste nesta classe de distribuições. Maiores detalhes sobre esse procedimento podem ser encontrados em McCullagh e Tibshirani (1990). 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

Nas Tabelas 2 e 3 apresentam-se algumas estatísticas descritivas referentes ao número de dias com precipitação pluvial em Araras-SP. Devem-se destacar os “zeros” como valores mínimos para os períodos junho, julho, agosto e setembro (Tabela 3), assim como “altos” coeficientes de variação para estes meses, respectivamente (Tabela 2). Destaca-se, portanto, a grande incidência de dias com ausência de precipitação neste período.

Tabela 2 - Média, desvio padrão e coeficiente de variação para a variável número de dias com precipitação pluvial em Araras SP, no período de 1972-2012

Meses	Média	Desvio Padrão	Coeficiente de Variação (%)
janeiro	15,54	3,66	23,57
fevereiro	12,59	4,17	33,14
março	10,00	3,16	31,62
abril	5,88	2,93	49,92
maio	5,49	3,15	57,35
junho	3,71	2,71	73,19
julho	2,80	3,13	111,67
agosto	2,76	2,59	93,84
setembro	6,02	3,38	56,11
outubro	8,61	3,43	39,80
novembro	10,41	2,97	28,56
dezembro	13,78	3,86	28,04

Tabela 3 - Separatrizes para a variável número de dias com precipitação pluvial em Araras SP, no período de 1972-2012

Meses	Mínimo	Máximo	1º Quartil	Mediana	3º Quartil
janeiro	5,00	26,00	13,00	16,00	18,00
fevereiro	3,00	22,00	9,00	13,00	15,00
março	3,00	18,00	8,00	10,00	12,00
abril	1,00	13,00	4,00	6,00	8,00
maio	1,00	16,00	3,00	5,00	7,00
junho	0,00	10,00	2,00	3,00	5,00
julho	0,00	16,00	1,00	2,00	4,00
agosto	0,00	8,00	0,00	2,00	4,00
setembro	0,00	14,00	4,00	6,00	8,00
outubro	2,00	15,00	7,00	8,00	11,00
novembro	5,00	18,00	9,00	10,00	12,00
dezembro	6,00	22,00	11,00	14,00	16,00

Já com relação ao total de chuva mensal, os resumos descritivos estão apresentados nas Tabelas 4 e 5.

Pode-se observar pela Tabela 4 que as menores médias correspondem ao período de inverno, abrangendo os meses de junho a agosto. Nesse período também se verificam os maiores coeficientes de variação. Os meses com maiores índices pluviométricos correspondem ao período de primavera e verão, abrangendo os meses de outubro a fevereiro.

Na Tabela 5 novamente deve-se observar a incidência de “zeros” como valores mínimos de precipitação para os meses de junho a setembro, em acordo com os valores registrados para o número de dias, apresentado na Tabela 3. Neste período,

Tabela 4 - Média, desvio padrão e coeficiente de variação para a variável total de precipitação pluvial mensal em Araras SP, no período de 1972-2012

Meses	Média	Desvio Padrão	Coeficiente de Variação (%)
janeiro	284,35	116,38	40,93
fevereiro	194,01	104,33	53,78
março	152,91	90,73	59,34
abril	71,23	48,14	67,59
maio	69,57	54,27	78,00
junho	45,41	51,01	112,34
julho	36,29	40,84	112,56
agosto	31,24	39,97	127,96
setembro	68,00	56,49	83,08
outubro	121,39	72,41	59,65
novembro	150,37	76,27	50,72
dezembro	218,54	94,92	43,44

Tabela 5 - Separatizes para a variável total de precipitação pluvial mensal em Araras SP, no período de 1972-2012

Meses	Mínimo	Máximo	1º Quartil	Mediana	3º Quartil
janeiro	75,80	587,20	198,70	276,20	352,00
fevereiro	22,00	547,00	132,60	185,40	257,00
março	17,00	420,80	82,80	130,60	207,90
abril	6,00	181,40	38,00	53,60	95,20
maio	3,00	195,60	30,60	52,00	89,20
junho	0,00	206,40	12,70	29,20	55,60
julho	0,00	147,40	3,00	18,00	60,20
agosto	0,00	161,20	0,00	15,60	54,40
setembro	0,00	255,60	33,60	60,20	84,80
outubro	3,60	325,40	82,20	105,20	148,10
novembro	32,40	407,20	118,60	150,40	186,60
dDezembro	50,60	493,00	162,60	195,00	251,50

destaca-se o mês de agosto, em que se tem um 1º quartil igual a zero, ou seja, em ao menos 25% do período de 41 anos da série histórica estudada, não houve incidência de chuva para este mês.

Na Tabela 6 estão apresentadas as estimativas dos parâmetros do modelo Tweedie referente aos dados sobre precipitação mensal em Araras-SP. Pode-se observar que os valores estimados para p conduzem à distribuição Poisson Composta para todos os meses, exceto possivelmente para o mês de abril. Este fato, também pode ser observado pela Figura 2, em que são apresentados os gráficos referentes aos perfis de verossimilhança. Deve-se observar também que o produto das estimativas

pontuais, $\hat{\alpha}\hat{\gamma}\hat{\lambda}$, equivale à estimativa da média de precipitação mensal (ver Tabela 4), o que reforça não somente a interpretabilidade dos parâmetros do modelo Tweedie mas também a sua importância enquanto representação da ocorrência do fenômeno.

Tabela 6 - Estimativas dos parâmetros do modelo Tweedie para os dados de precipitação mensal em Araras-SP, no período de 1972-2012

	\hat{p}	$\hat{\phi}$	$\hat{\alpha}$	$\hat{\gamma}$	$\hat{\lambda}$	$\hat{\alpha}\hat{\gamma}$	$\hat{\alpha}\hat{\gamma}\hat{\lambda}$
janeiro	1,63	1,30	-0,58	-29,40	16,66	17,07	284,35
fevereiro	1,39	6,96	-1,58	-20,80	5,91	32,85	194,01
março	1,80	0,95	-0,26	-41,46	14,38	10,63	152,91
abril	2,04	0,38	0,04	-33,42	-54,35	-1,31	71,23
maio	1,80	1,44	-0,26	-33,59	8,08	8,61	69,57
junho	1,45	7,81	-1,24	-19,10	1,91	23,72	45,41
julho	1,37	10,80	-1,68	-15,35	1,40	25,83	36,29
agosto	1,41	11,31	-1,44	-18,95	1,14	27,34	31,24
setembro	1,39	7,57	-1,56	-15,41	2,83	23,99	68,00
outubro	1,47	4,01	-1,13	-17,90	6,00	20,24	121,39
novembro	1,63	1,52	-0,58	-22,87	11,32	13,28	150,37
dezembro	1,63	1,29	-0,58	-24,65	15,27	14,32	218,54

Nas Figuras 3, 4 e 5 são apresentadas as distribuições da precipitação pluvial mensal em Araras-SP por meio de histogramas, bem como as curvas das distribuições ajustadas. Verifica-se que o modelo Poisson Composto ajustou-se bem aos dados para os meses com maior incidência de chuva, ou seja, o período entre novembro e março. Para o período de seca, o modelo Poisson Composto permitiu captar o excesso de zeros e devido a isso o ajuste foi satisfatório e a modelagem adequada.

Considerações finais

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 a 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 em Araras-SP. 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.

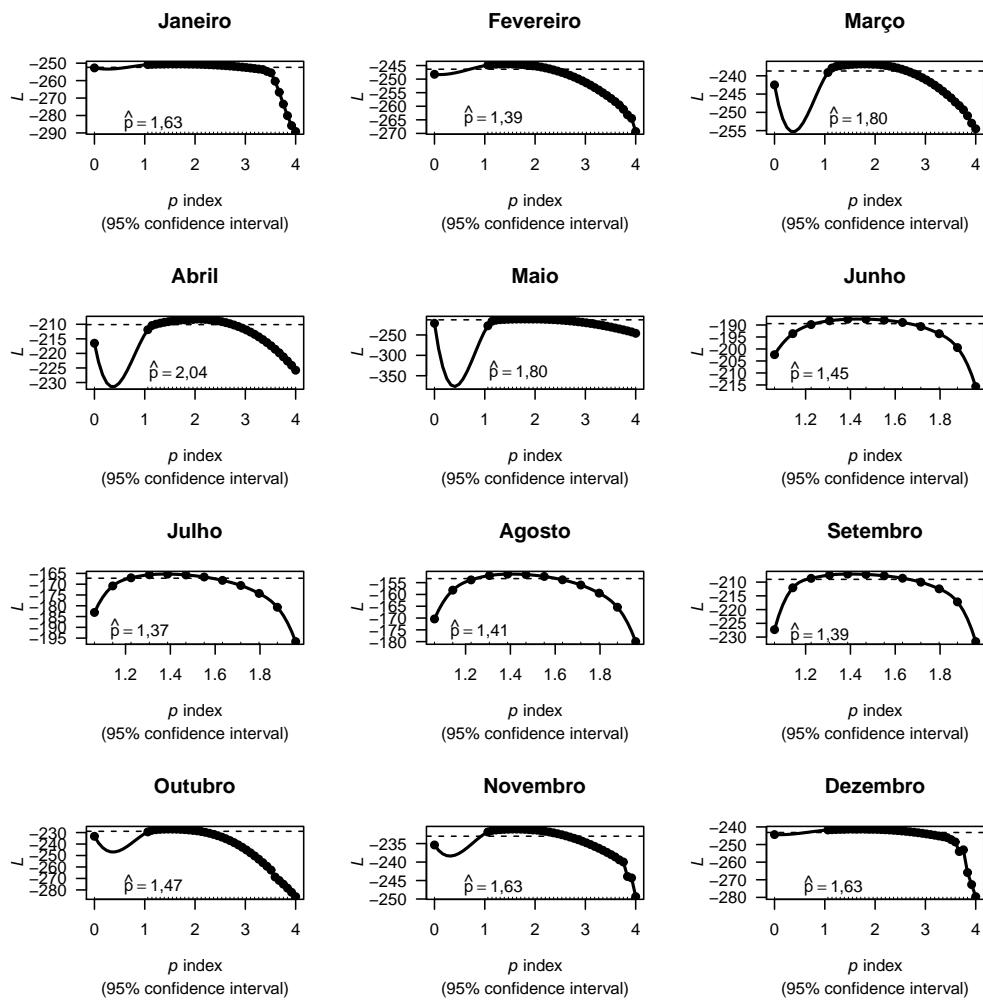


Figura 2 - Gráficos dos “perfis de verossimilhança” para cada um dos meses e respectivos valores máximos de p .

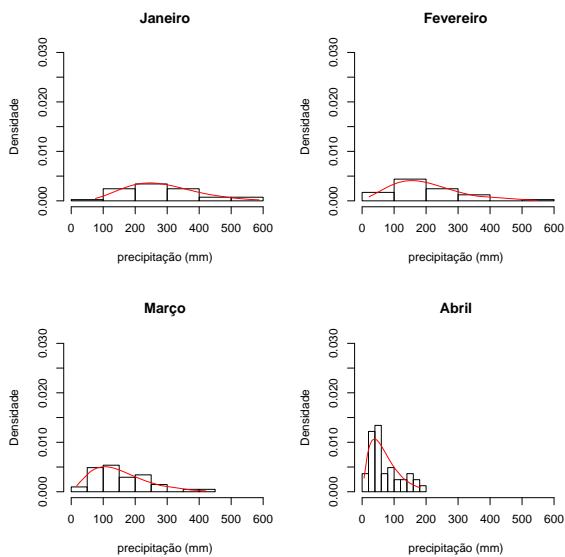


Figura 3 - Histograma e respectivas curvas de ajuste para as distribuições ajustadas referentes aos meses de Janeiro a Abril.

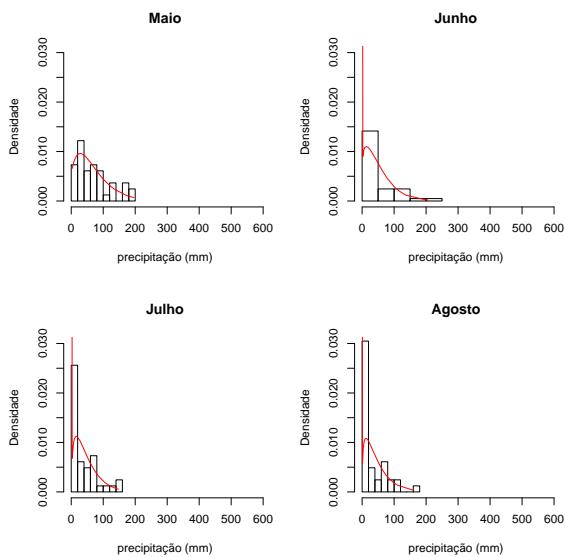


Figura 4 - Histograma e respectivas curvas de ajuste pela distribuição Poisson Composta referentes aos meses de Maio a Agosto.

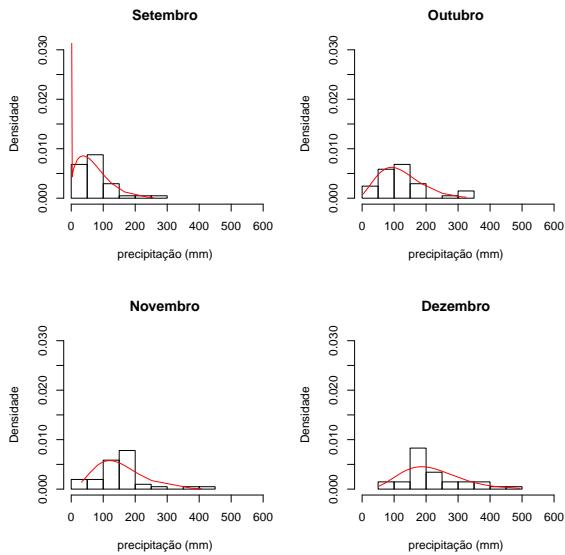


Figura 5 - Histograma e respectivas curvas de ajuste pela distribuição Poisson Composta referentes aos meses de Setembro a Dezembro.

SERMARINI, R. A.; LARA, I. A. R.; PIEDADE, S. M. S.; MANIERO, M. A. Probability models of the Tweedie family for occurrences of rainfall in Araras SP . *Rev. Bras. Biom.*, São Paulo, v.32, n.2, p.282-295, 2014.

■ **ABSTRACT:** Rainfall is an important meteorological variable that has immediate impact on the climate, agriculture and economy of a region. The characterization of this variable is done through a probabilistic model, requiring adjustment of different distributions, as the time and location of the event. On the other hand, the theory of dispersion models includes both statistically and probability and provides a collection of methods related to the distributions of exponential families, for which the concepts of leasing and scale can be generalized by the concepts of position and dispersion. The Tweedie family belongs to a class of models called exponential dispersion models and can be used to model events related to rainfall. In this context, the aim of this paper is to apply these methods to adjust Tweedie family of distributions to data on rainfall Araras city, for each of the twelve months of the year, based on official data from the Center for Applied Sciences, Federal University of São Carlos weather station during the period 1972-2012. It was found that the methodology is appropriate for representing this variable and that the Poisson Composed distribution is the best fit for most months. Data analysis was performed with the aid of Tweedie exponential family models package available in R software.

■ **KEYWORDS:** Rainfall; probabilistic models; Tweedie exponential dispersion model.

Referências

- ARAGÃO, R.; SANTANA, G. R. COSTA, C. E. F. F.; CRUZ, M. A. S.; FIGUEIREDO, E. E.; SRINIVASAN, V. Chuvas intensas para o estado de Sergipe com base em dados desagregados de chuva diária. *Revista Brasileira de Engenharia Agrícola e Ambiental*, Campina Grande, v. 17, n. 3, p. 243-252, 2013.
- BLAIN, G. C.; PIEDADE, S. M. S.; CAMARGO, M. B. P. C.; GIAROLLA, A. Distribuição temporal da precipitação pluvial mensal observada no posto meteorológico do Instituto Agronômico, em Campinas, SP, *Bragantia*, Campinas, v. 66, n. 2, p. 347-355, 2007.
- 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*, Campina Grande, v. 6, n. 1, p. 76-82, 2002.
- DUNN, P. K. *Tweedie exponential family models: The Tweedie Package*, v.1.2, June. 2006.
- DUNN, P. K.; WHITE, N. Power-variance models for modelling rainfall. In: International Workshop on Statistical Modelling, 20., 2005, Sydney. *Anais eletrônicos...* Sydney, 2005. Disponível em: [jhttps://eprints.usq.edu.au/734/1/Dunn-White-pvar.pdf](https://eprints.usq.edu.au/734/1/Dunn-White-pvar.pdf). Acesso em: 20 fev. 2013.
- JØRGENSEN, B. *The Theory of Dispersion Models*. Monographs on Statistics and Applied Probability. New York: Chapman & Hall. 1997. 237p.
- MCCULLAGH, P.; TIBSHIRANI, R. A simple method for adjustment of profile likelihood. *Journal of the Royal Statistical Society, Serie B*, London, v. 52, n. 2, p.325-344,1990.
- NELDER, J. A.; WEDDERBURN, R. W. Generalized Linear Models. *Journal of the Royal Statistical Society, Serie A*, London, v. 135, n. 3, p.370-384, 1972.
- PIZZATO, J. A.; DALLACORT, R.; TIEPPO, R. C.; MODOLÓ, A. J.; CREMOM, C.; MOREIRA, P. S. P. Distribuição e probabilidade de ocorrência de precipitação em Cáceres (MT). *Pesquisa Agropecuária Tropical*, Goiás, v. 42, n. 2, p. 137-142, 2012.
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- SANSIGOLO, C. A. Distribuições de extremos de precipitação diária, temperatura máxima e mínima e velocidade do vento em Piracicaba, SP (1917-2006). *Revista Brasileira de Meteorologia*, v. 23, n. 3, p. 341-346, 2008.
- SANTOS, G. G.; FIGUEIREDO, C. C.; OLIVEIRA, L. F. C.; GRIEBELER, N. P. Intensidade-duração-frequência de chuvas para o Estado de Mato Grosso do Sul. *Revista Brasileira de Engenharia Agrícola e Ambiental*, Campina Grande, v. 13, (Suplemento), p. 899-905, 2009.

- SILVA, B. B.; FERREIRA, M. A. F.; SILVA, V. P. R.; FERREIRA, R. C. Desempenho de modelo climático aplicado à precipitação pluvial do Estado de Pernambuco. *Revista Brasileira de Engenharia Agrícola e Ambiental*, Campina Grande, v. 14, n. 4, p. 387-395, 2010.
- 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*, Fortaleza, v. 44, n. 3, p. 481-487, 2013.
- 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*, Campina Grande, v. 11, n. 1, p. 67-72, 2007.
- WEDDERBURN, R.W.M. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, Cambridge, v. 61, n. 3, p. 439-447, 1974.
- 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. In: SIMPOSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 12., 2007, Santa Maria. *Anais...* Santa Maria, 2007. 1 CD-ROM.

Recebido em 20.02.2014.

Aprovado após revisão em 20.02.2014.