

# Avaliação de modelos de densidade de probabilidade em séries de dados meteorológicos

Luana Osaka Ohashi<sup>1</sup>

José Eduardo Boffino de Almeida Monteiro<sup>2</sup>

**Resumo:** O objetivo deste estudo foi avaliar o ajuste de funções de densidade de probabilidade a séries de dados de chuva. Os modelos de distribuição testados foram os Normal, Log- Normal, Gama, Exponencial. Foram utilizadas 2400 séries de dados diários, de 1999 a 2013, de precipitação diária, provenientes do Sistema de Monitoramento Agrometeorológico (Agritempo) e da Agência Nacional de Águas (ANA). Os dados foram agrupados mês a mês, de janeiro a dezembro, compondo doze conjuntos de dados. As funções de distribuição foram ajustadas para o conjunto de dados de cada mês para cada estação meteorológica. A aderência dos dados às distribuições teóricas foi verificada pelo método de Anderson-Darling a um nível de 5% de significância. Observou-se que a distribuição Gama apresentou aderência significativa em 96% das séries estudadas, seguida pela Log-Normal com aderência significativa em 91% das séries.

**Palavras-chave:** dados pluviométricos, função densidade de probabilidade.

---

<sup>1</sup> Estudante de Estatística do Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas (IMECC/Unicamp), estagiária da Embrapa Informática Agropecuária, Campinas, SP.

<sup>2</sup> Engenheiro agrônomo, doutor em Agrometeorologia, pesquisador da Embrapa Informática Agropecuária, Campinas, SP.

## Introdução

O estudo de distribuições de probabilidade de variáveis meteorológicas ao longo do tempo é uma forma de compreender os fenômenos meteorológicos e determinar seus padrões de ocorrência. Uma análise de distribuição de frequência permite uma previsibilidade razoável da variabilidade meteorológica de uma região, sendo esta uma ferramenta de grande importância para o planejamento e gestão de diversas atividades agrícolas e humanas a curto e longo prazo.

Dourado Neto et al. (2005) realizaram um estudo para ajustar modelos de distribuição de probabilidade a uma série de 82 anos de precipitação pluvial diária em Piracicaba, SP. Os autores avaliaram os modelos Gama, Exponencial, Weibull, Normal e Log-Normal pelo teste de Qui-Quadrado a 5% de significância, nos meses de novembro, dezembro, janeiro, fevereiro e março. O melhor ajuste foi obtido com o modelo de distribuição Gama.

O objetivo deste trabalho foi verificar o ajuste univariado dos modelos de distribuição de probabilidade Normal, Log-Normal, Gama, Log-Normal a séries de dados de Precipitação. Os modelos ajustados serão, posteriormente, utilizados para a definição de um intervalo associado a diferentes níveis de probabilidade, permitindo identificar valores improváveis ou provavelmente errados, para cada local. Esses critérios servirão para compor um dos algoritmos de um sistema automatizado de controle de qualidade de dados meteorológicos.

## Materiais e Métodos

Foram utilizadas cerca de 2400 séries históricas de dados diários, de 1999 a 2013, da ANA<sup>1</sup> e do Agritempo<sup>2</sup>. Os dados analisados são originários de estações localizadas nas regiões identificadas na Figura 1.

Todas as observações foram analisadas e estratificadas por meses e estação meteorológica. Nos agrupamentos mensais, o mês de cada ano e cada estação que apresentou menos de cinco valores considerados válidos foram excluídos das análises. Foram considerados como válidos os valores de

---

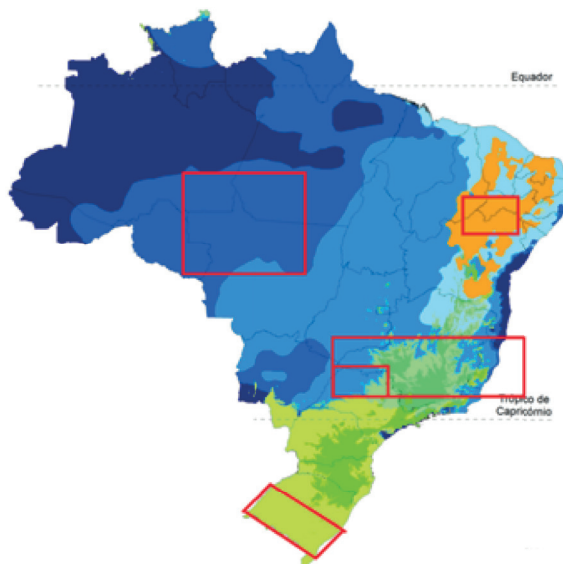
<sup>1</sup> Disponível em: <<https://lucene.apache.org/>>. Acesso em: 28 set. 2015.

<sup>2</sup> Disponível em: <<http://commons.apache.org/>>. Acesso em: 28 set. 2015.

precipitação maiores ou iguais a zero, ou seja, não negativos. Porém, para o ajuste dos modelos, foram considerados somente os dias com chuva, ou seja, valores maiores que zero (DOURADO NETO et al., 2005). Os tratamentos e análises de dados foram realizados com auxílio do software R (R DEVELOPMENT CORE TEAM, 2011).

Foram avaliadas as funções Normal, Gama, Exponencial e Log-Normal para a variável precipitação, as estimativas dos parâmetros das funções das densidades de probabilidade foram obtidas pelo método da Máxima Verossimilhança.

A escolha dos modelos de distribuição de probabilidades ajustadas a cada tipo de variável foi baseada em resultados de estudos semelhantes publicados (DAMÉ et al, 1996; DOURADO NETO et al., 2005), ou nas características das variáveis. Em razão da simplicidade, a seleção de modelos foi limitada aos que continham não mais do que 3 parâmetros.



**Figura 1.** Classificação climática de Köppen no Brasil e regiões selecionadas para o estudo. Região 1, Nordeste (-8,0: -42,5) a (-10,5: -38,0); Região 2, Sudeste (-18,0: -52,0) a (-22,0: -39,0); Região 3, Norte de São Paulo (-20,0: -52,0) a (-22,0: -49,0); Região 4, Sul do Rio Grande do Sul (-27,7: -55,5) a (-30,4: -51,8) a (-33,0: -54,0) a (-30,0: -58,0) e Região 5, Mato Grosso, Pará e Amazonas (-6,50: -61,0) a (-13,0: -52,0).

Fonte: Alvares et al. (2013).

Neste trabalho, foi usado o teste de aderência de Anderson-Darling (ANDERSON; DARLING, 1954) como critério para indicar o grau de ajuste do modelo ao conjunto de dados. Depois da qualificação dos ajustes em relação ao nível de significância de 5% ( $\alpha=0,05$ ) foi calculada a porcentagem das estações em cada mês que obtiveram um ajuste adequado estatisticamente.

## Resultados e Discussão

A Tabela 1 indica o percentual de estações que tiveram os dados observados compatíveis com os modelos de distribuição teórica, de acordo com o teste de Anderson-Darling. É possível verificar porcentagens bastante altas para as distribuições. O percentual mais elevado de estações com ajustes significativos foi maior no modelo Gama, com 96%, em média, e acima de 95% em todas as regiões avaliadas. Esses resultados corroboram o que tem sido relatado na literatura para modelos de distribuição de chuva (DAMÉ et al., 1996; DOURADO NETO et al., 2005).

O segundo maior percentual foi obtido com o modelo Log-Normal, com 91% na média das cinco regiões. Porém, neste modelo, as regiões 2 e 5 apresentaram cerca de 88% de estações com ajustes significativos. Os modelos Exponencial e Normal apresentaram percentuais de ajuste médio de 72% e 33,7%, respectivamente. A Tabela 2 indica, mês a mês, o percentual de estações que tiveram os dados observados compatíveis com o modelo de distribuição Gama, de acordo com o teste de Anderson-Darling. Apesar da estacionalidade das chuvas e diferentes padrões de precipitação dependendo do mês em cada região, obteve-se alto percentual de aderência em

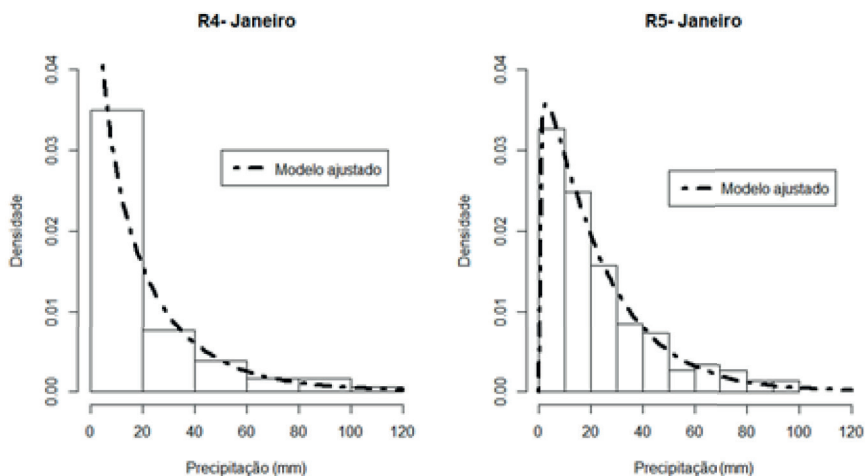
**Tabela 1.** Percentual médio de estações que foram compatíveis estatisticamente com as distribuições teóricas para a variável precipitação.

Região	Estações válidas	Gama	Normal	Exponencial	Log-Normal
R1	162	98%	50,5%	61,4%	98,9%
R2	1164	95%	29,2%	68,9%	88,6%
R3	326	98,7%	47,3%	84,7%	95,2%
R4	138	99,4%	23%	81%	93,8%
R5	108	95,2%	29,5%	72,6%	88%
<b>Média Geral Ponderada</b>		96,2%	33,7%	72%	91%

todos os meses, com valores acima de 90%. Somente na Região 5, o mês de março apresentou percentual abaixo de 90%.

**Tabela 2.** Percentual médio de estações que foram compatíveis estatisticamente com a distribuição Gama nos doze meses do ano.

Região	Jan.	Fev.	Mar.	Abr.	Mai	Jun.	Jul.	Ago.	Set.	Out.	Nov.	Dez.
R1	99%	99%	99%	97%	97%	97%	97%	97%	98%	99%	99%	99%
R2	96%	97%	96%	94%	93%	92%	94%	95%	96%	95%	97%	96%
R3	99%	99%	98%	99%	99%	98%	99%	99%	100%	98%	99%	99%
R4	99%	100%	99%	99%	99%	99%	99%	100%	99%	100%	100%	99%
R5	93%	94%	88%	94%	96%	99%	100%	100%	97%	94%	95%	92%



**Figura 1.** Curva da distribuição Gama ajustada aos dados de precipitação e o histograma de uma estação das regiões 4 e 5 no mês de janeiro.

Com base nesses resultados, é possível afirmar que seria adequado utilizar o modelo Gama para análise da probabilidade de chuva em todos os meses e regiões estudadas. Isso permite a formulação do método de controle de qualidade de dados de chuva baseado nos limites de probabilidade calculada somente pelo modelo Gama ajustado para cada ponto, sem a necessidade de considerar outros modelos para regiões ou meses distintos.

A Figura 2 ilustra dois exemplos do histograma de dados observados e a distribuição ajustada pelo modelo para o mês de janeiro de uma das estações na Região 4 e outra na Região 5. Todos os ajustes foram considerados adequados segundo o teste Anderson-Darling. Destes casos em questão, é possível inferir pelo modelo que a frequência de valores diários acima de 100 mm, por exemplo, é inferior a 0,1%.

## Considerações Finais

Apesar de apresentar alguma variação por região ou por mês, alguns modelos tiveram ajuste estatisticamente satisfatório na maior parte das séries. A distribuição Gama apresentou aderência significativa em 96% das séries estudadas, seguida pela Log-Normal com aderência significativa em 91% das séries. Além disso, o modelo de distribuição Gama foi o que apresentou o melhor desempenho em todas as regiões e meses, exceto na Região 1 que teve um resultado pouco menor que o modelo Log-normal.

## Referências

ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; GOLÇALVES, J. L. M.; SPAROVEK, G. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711-728, 2013.

ANDERSON, T. W.; DARLING, D. A. A test of goodness-of-fit. **Journal of the American Statistical Association**, v. 49, n. 268, p. 765-769, Dec.1954.

DAMÉ, R. C. F.; TEIXEIRA, C. F. A.; MOREIRA, N. M.; SOUTO, M. V. Análise de frequência hidrológica dos dados de precipitação pluvial de algumas estações agroclimatológicas da região do sul do Rio Grande do Sul. **Ciência Rural**, v. 26, n. 3, p. 351-355, 1996.

DOURADO NETO, D. ; ASSIS, J. P.; TIMM, L. C.; MANFRON, P. A.; SPAROVEK, G.; MARTIN, T. N. Ajuste de modelos de distribuição de probabilidade nas séries históricas de precipitação pluvial diária em Piracicaba- SP. **Revista Brasileira de Agrometeorologia**, v. 13, p. 273-283, 2005.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2011. Disponível em: <<http://www.R-project.org>>. Acesso em: 10 dez. 2015.