



XIX Congresso Brasileiro de Agrometeorologia

23 a 28 de agosto de 2015

Lavras – MG – Brasil

Agrometeorologia no século 21:



O desafio do uso sustentável dos biomas brasileiros

Ajuste de séries históricas de temperatura média do ar e radiação solar global diárias às funções densidade de probabilidade Normal, Gamma e Weibull em Piracicaba - SP

Simone Toni Ruiz Corrêa²; Durval Dourado Neto³; Fábio Vale Scarpere⁴

¹ Trabalho apresentado no XIX Congresso Brasileiro de Agrometeorologia, 23 a 28 de agosto de 2015.

² Agrônoma. Pós-doutoranda, Depto. Sustentabilidade, CTBE, Campinas - SP, Fone (19) 3512-3514, simone.correa@bioetanol.org.br

³ Agrônomo, Prof. Doutor, Depto. de Produção Vegetal, ESALQ/USP, Piracicaba-SP, d.dourado@gmail.com

⁴ Agrônomo, Pesquisador, Depto. Sustentabilidade, CTBE, Campinas, fabio.scarpere@bioetanol.org.br

RESUMO: Objetivou-se com o este trabalho verificar o ajuste das séries históricas de temperatura média do ar e radiação solar global diárias às funções densidade de probabilidade Normal, Gamma e Weibull em Piracicaba, SP. Para tanto, utilizou-se o teste de aderência de Kolmogorov-Smirnov a um nível de significância (α) de 5%. Verificou-se para a variável temperatura média do ar que as três distribuições analisadas obtiveram ajustes bastante satisfatórios aos dados diários das séries históricas, ou seja, são adequadas para representar o padrão de ocorrência desta variável em Piracicaba, SP. Para a variável radiação solar global, as distribuições que obtiveram melhor desempenho quanto aos ajustes aos dados diários das séries históricas foram a normal e a Gamma. Destaca-se que, para ambas as variáveis, a distribuição normal foi a mais homogênea dentre as três analisadas (ajuste superior a 99,4% em ambos os casos). Tal informação é valiosa no que diz respeito às inferências estatísticas nesta localidade, propiciando adequada previsibilidade do clima.

PALAVRAS-CHAVE: Teste de aderência, Kolmogorv-Smirnov, Distribuições de probabilidade

Historical series adjustment of the daily average air temperature and solar radiation to Normal, Gamma and Weibull probability density functions, in Piracicaba – SP

ABSTRACT: The objective of this work was to verify the fit of the historical series of daily average air temperature and solar radiation to Normal, Gamma and Weibull probability density functions in Piracicaba, SP. For this, we used the Kolmogorov-Smirnov test at a significance level (α) of 5%. It was found for the variable average air temperature that the three distributions analyzed obtained satisfactory adjustments to daily data of time series, i.e. they are suitable to represent the pattern of occurrence of this variable in Piracicaba, SP. For the solar radiation, distributions that performed better regarding adjustments to daily data of the historical series were normal and Gamma. It is noteworthy that, for both variables, normal distribution was the most homogeneous of the three analyzed (set higher than 99.4% in both cases). This information is valuable with regard to statistical inferences at this location, providing adequate predictability of climate.

KEY WORDS: Goodness of fit test, Kolmogorv-Smirnov, Probability distributions

INTRODUÇÃO

O tratamento estatístico dos elementos do clima observados ao longo do tempo, determinando seus padrões de ocorrência e propiciando uma adequada previsibilidade do padrão do tempo e clima de uma região, é um instrumento de grande valia no planejamento e na gestão de inúmeras atividades agrícolas, agropecuárias e humanas. Estes modelos, quando bem ajustados, resultam em maior embasamento estatístico da descrição climática (SANSIGOLO; NERY, 2000), e nos permitem realizar estimativas sem precisar da totalidade de informações.

O desafio do uso sustentável dos biomas brasileiros

O uso de funções densidade de probabilidade está diretamente ligado à natureza dos dados a que elas se relacionam. Algumas possuem 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. Desde que respeitado o aspecto da representatividade dos dados, as estimações dos seus parâmetros, para uma determinada região, podem ser estabelecidas como de uso geral, sem prejuízo da precisão na estimação da probabilidade (CATALUNHA et al., 2002).

O ajuste de modelos probabilísticos aos dados diários de atributos do clima fornece um resumo sucinto desses dados, além de representar uma técnica eficiente para a análise dessas informações. A forma com que a distribuição de frequências é apresentada pode ser aproximada por meio 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 por meio de testes de aderência.

Muitas são as distribuições de probabilidades conhecidas para ajuste de variáveis aleatórias contínuas, dentre as quais se destacam as distribuições: Uniforme; Normal; Log-normal; Exponencial; Gama e Weibull (MAGALHÃES; LIMA, 2010).

Este estudo abordou as distribuições normal, gama e Weibull, tendo em vista ser a primeira a distribuição de probabilidade contínua mais importante sob a ótica do conhecimento agrônomo, considerando a quantidade de fenômenos que podem ser estudados por seu intermédio e serem as distribuições gama e Weibull extremamente versáteis, função da flexibilidade para assumir diferentes formas e simetrias, condição oportuna para o propósito da simulação dos elementos do clima.

MATERIAIS E MÉTODOS

As informações diárias referentes aos dados meteorológicos em Piracicaba foram obtidos no Departamento de Engenharia de Biosistemas da Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ), da Universidade de São Paulo (USP). A Estação Meteorológica Convencional de Piracicaba está localizada nas seguintes coordenadas: latitude: 22°42’30”S; longitude de 47°38’30”W e altitude: 546 metros. Segundo a classificação climática de Köppen, o clima da região é do tipo Cwa, caracterizado por um total de chuvas no mês mais seco (agosto) de 28,4 mm e do mês mais chuvoso (janeiro) de 229,7 mm, sendo a precipitação total anual de 1328 mm. A temperatura média do mês mais quente é de 25°C, e a do mês mais frio de 17°C, sendo a temperatura média anual de 21,6°C. Segundo Villa Nova (2003), a radiação solar global média anual é de 435 cal cm⁻² d⁻¹.

As observações utilizadas neste estudo se referem à temperatura média diária, em graus Celsius (°C) e à radiação solar global diária, em cal cm⁻² d⁻¹. A série histórica de temperatura abrange o período de 1° de janeiro de 1917 a 31 de dezembro de 2011, num total de 94 anos; já a série histórica de radiação solar global diária compreende o período de 1° de janeiro de 1981 a 31 de dezembro de 2011, totalizando 31 anos.

Muitos são os possíveis métodos para ajustar uma distribuição de probabilidade a um conjunto de dados, para um valor particular de significância (α), verificando se a hipótese da distribuição do modelo teórico qualquer pode representar adequadamente um conjunto de informações. *Softwares* como o R possuem a facilidade de, simultaneamente, testar o ajuste do modelo teórico por meio de diferentes métodos, dentre os quais se destacam o teste de aderência, ou bondade do ajuste, de Kolmogorov-Smirnov, sendo:

$$D = |F(X) - S(X)| \quad (1)$$

O desafio do uso sustentável dos biomas brasileiros

em que, a máxima diferença absoluta (D) entre a função de distribuição acumulada assumida para os dados $F(X)$ e a função de distribuição empírica dos dados $S(X)$.

Como critério, compara-se D com d (valor crítico ou desvio máximo tabelado para dado nível de significância) apresentados em Campos (1983), ou calculado para um número de amostras superior a 50, conforme segue:

$$d = \frac{1,36}{\sqrt{n}} \quad (2)$$

em que n é o número de anos da série histórica (amostra).

Se D for menor que d , existe concordância entre as frequências observadas e esperadas, portanto, a amostra provém de uma população que segue a distribuição de probabilidade sob teste. Ou seja, para verificar o ajuste da distribuição aos dados, a um nível de significância (α) de 5%, a hipótese a ser testada é que $H_0: F \equiv F_0$ (os dados seguem uma dada distribuição) *versus* $H_a: F \neq F_0$, (os dados não seguem a distribuição em questão) para pelo menos um valor da variável aleatória X , rejeitando H_0 se $D \geq d$.

O “valor p ” é um nível descritivo muito utilizado para sintetizar o resultado de um teste de hipóteses e significa a probabilidade de que a estatística do teste (como variável aleatória) tenha valor extremo em relação ao valor observado (estatística), quando a hipótese H_0 é verdadeira. Neste teste, o nível de significância é a probabilidade máxima de rejeitar acidentalmente a hipótese base verdadeira (H_0) (erro do tipo I). Portanto, quando o valor p é menor que o nível de significância proposto (α), o valor observado está na região crítica e, portanto, rejeita-se a hipótese nula H_0 ; por outro lado, se o valor p é maior que o nível de significância, não se rejeita H_0 . Para fins de análise e simulação, adotou-se como critério de não rejeição de H_0 um valor p maior ou igual a 0,05 pelo teste de Kolmogorov-Smirnov tido, estatisticamente, como nível significativo.

A verificação do ajuste das distribuições de probabilidade (normal, gama e Weibull) foi realizada para todos os dias do ano para a temperatura média do ar e radiação solar global, em Piracicaba, por meio do *software* R, pacote *MASS*.

RESULTADOS E DISCUSSÃO

Como resultados, tem-se para a variável temperatura média do ar que as três distribuições apresentaram aderência satisfatória aos dados diários, com melhor desempenho da distribuição normal (99,45% de aderência), seguida pela distribuição Weibull (98,63%) e gama (97,53%). Observa-se, também, que os dados apresentaram uma forte simetria e homogeneidade no decorrer do ano, dado o nível de significância mencionado, como observado na Figura 1.

Por meio do valor p verifica-se a qualidade dos ajustes sendo que, para a distribuição normal, apenas sete dias do tiveram valores abaixo de 0,1. Não houve aderência da distribuição normal aos dados dos dias 27/8 e 11/11 enquanto a distribuição Weibull não apresentou ajuste aos dados dos dias 15/1, 2/2, 3/3, 13/4 e 27/8. Observou-se, para a distribuição gama, que o ajuste não foi positivo nos meses mais secos do ano.

Quanto aos valores dos parâmetros, têm-se para a distribuição normal que as menores médias diárias de temperatura média do ar (abaixo de 17°C) ocorreram entre os dias 10 e 13/7, enquanto que as maiores (acima de 24°C), nos meses de janeiro, fevereiro e maior parte de março. O desvio padrão oscilou entre 1,35°C e 3,31°C, com média de 2,29°C. Os parâmetros da distribuição gama têm valores mínimo, médio e máximo de 26,63, 105,51 e 326,10, respectivamente, para forma (α), e de 1,43, 4,65 e 13,25, respectivamente, para taxa (λ). Na distribuição Weibull, observa-se para o parâmetro de forma (α), 6,17, 11,60 e 20,53, respectivamente, valores mínimo, médio e máximo, e para a escala (β), 17,66, 22,59 e 25,74.

De maneira geral, observa-se que os valores destes parâmetros são menores nos meses mais frios do ano (junho, julho e agosto) e maiores nos meses mais quentes (dezembro a março). Maiores informações sobre este estudo, bem como as estimativas diárias de cada parâmetro, e valores p podem ser obtidas em Corrêa (2013).

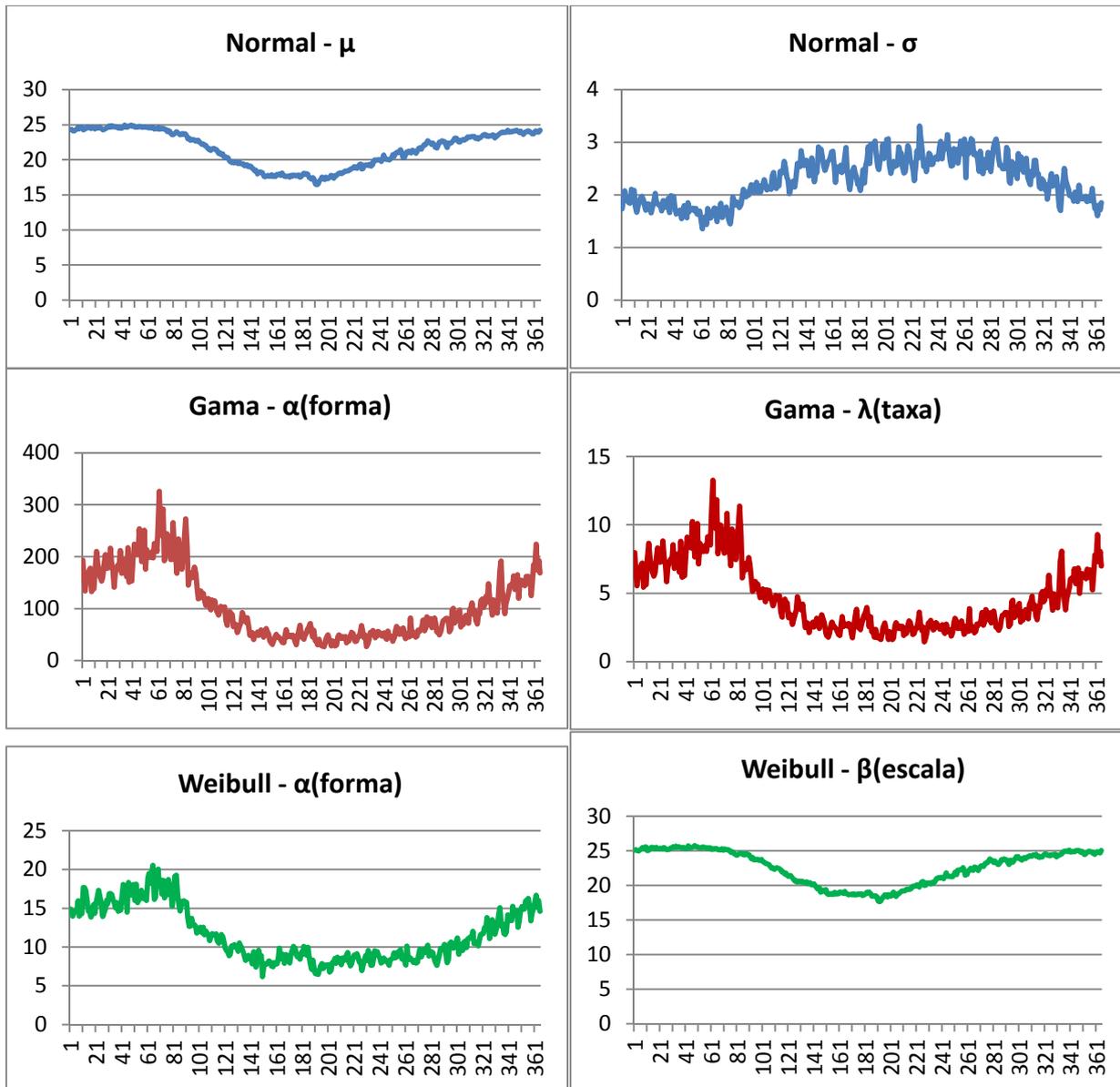


Figura 1. Ilustração dos valores das estimativas dos parâmetros média (μ) e desvio padrão (σ) para a distribuição normal, forma (α) e taxa (λ) para a distribuição gama e forma (α) e escala (β) para a distribuição Weibull, do teste de aderência (Kolmogorov-Smirnov). Valores referentes à análise estatística para a série histórica diária (94 anos) de temperatura média do ar ($^{\circ}\text{C}$) em Piracicaba (SP)

Quanto à radiação solar global, nota-se que a distribuição normal obteve ajuste em 99,73% dos casos, sendo que os valores p foram relativamente superiores às demais distribuições. Esta distribuição não aderiu aos dados em apenas um dia do ano (14/6). A distribuição gama teve aderência satisfatória em 338 dias do ano (92,60%), com decréscimo em períodos mais secos. Já a distribuição Weibull aderiu aos dados somente em 252 dias, ou 69,04% dos casos, sendo que os meses com menores aderências foram setembro, outubro e novembro.

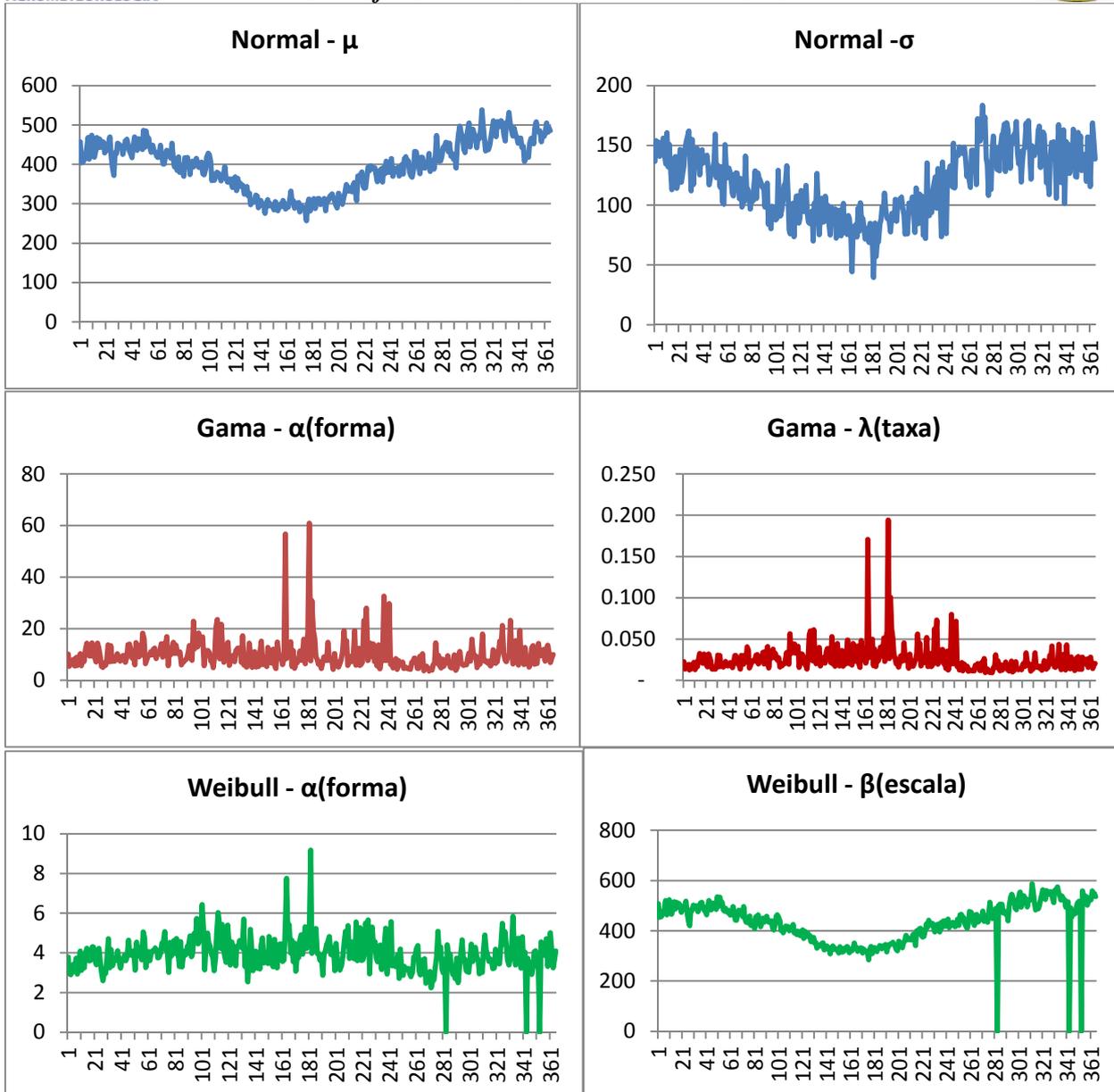


Figura 2. Ilustração dos valores das estimativas dos parâmetros média (μ) e desvio padrão (σ) para a distribuição normal, forma (α) e taxa (λ) para a distribuição gama e forma (α) e escala (β) para a distribuição Weibull, do teste de aderência (Kolmogorov-Smirnov). Valores referentes à análise estatística para a série histórica diária (31 anos) de radiação solar global ($\text{cal cm}^{-2} \text{d}^{-1}$) em Piracicaba (SP).

O processo de estimação dos parâmetros de algumas funções de distribuição, em alguns casos, não pode ser resolvido de maneira analítica devido à forma não linear apresentada por suas funções, o que leva a utilização de recursos numéricos computacionais para que se consiga estimar os parâmetros relacionados ao modelo escolhido. Um desses recursos é a otimização não linear, a qual, mediante alguns métodos, proporciona a estimação desejada e que seria de difícil obtenção, de forma analítica.

À título de demonstração dos valores dos parâmetros α e β da distribuição Weibull, dos 113 dias onde não houve ajuste, 110 não apresentaram convergência dos dados para o cálculo dos parâmetros pelo método da máxima verossimilhança (padrão), ou seja, o estimador de máxima verossimilhança

O desafio do uso sustentável dos biomas brasileiros

não conseguiu escolher aquele par de parâmetros que melhor explica a amostra observada. Foi, então, usada uma função de otimização com base em Byrd et al. (1995), pelo do método "*L-BFGS-B*", disponível no *software* R, e que permite que cada variável seja calculada com base em um limite inferior e/ou superior. Os valores dos dias 10/10, 9/12 e 19/12 também não apresentaram convergência de dados por este método.

Quanto à análise dos parâmetros, têm-se as menores médias diárias de radiação solar nos meses de maio a julho e as maiores nos meses de outubro a março. O desvio padrão oscilou entre 39,49 e 183,54, com média de 118,80.

Os parâmetros da distribuição gama apresentaram valores mínimo, máximo e médio de 3,58, 60,88 e 9,94, respectivamente, para forma (α), e de 0,009, 0,194 e 0,026, respectivamente, para taxa (λ). Para a distribuição Weibull, têm-se para o parâmetro forma (α), 2,24, 9,19 e 3,97, respectivamente, valores mínimo, máximo e médio, e para a escala (β), 284,92, 587,06 e 437,91.

CONCLUSÕES

Para a variável temperatura média do ar tem-se, portanto, que as três distribuições analisadas obtiveram ajustes bastante satisfatórios aos dados diários das séries históricas, ou seja, são adequadas para representar o padrão de ocorrência desta variável em Piracicaba, SP. Para a variável radiação solar global, as distribuições que obtiveram melhor desempenho quanto aos ajustes aos dados diários das séries históricas foram a normal e a gama.

Destaca-se que, tanto para a radiação solar global quanto para a temperatura média do ar, a distribuição normal foi a mais homogênea dentre as três analisadas (ajuste superior a 99,4% em ambos os casos). Tal informação é valiosa no que diz respeito às inferências estatísticas nesta localidade, propiciando adequada previsibilidade do clima.

REFERÊNCIAS BIBLIOGRÁFICAS

BYRD, R.H.; LU, P.; NOCEDAL, J.; ZHU, C. A limited memory algorithm for bound constrained optimization. **SIAM Journal on Scientific Computing**, Canada, n. 16, p. 1190-1208, 1995.

CAMPOS, H. **Estatística não paramétrica**. 4. ed. Piracicaba: ESALQ, 1983. 349 p.

CATALUNHA, M.J.; SEDIYAMA, G.C.; LEAL, B.G.; SOARES, C.P.B.; RIBEIRO, A. Aplicação de cinco funções densidade de probabilidade a séries de precipitação pluvial no Estado de Minas Gerais. **Revista Brasileira de Agrometeorologia**, Santa Maria, v. 10, n. 1, p. 153-162, 2002.

MAGALHÃES, M.N.; LIMA, A.C.P. de. **Noções de probabilidade e estatística**. 7. ed. São Paulo: EDUSP, 2010. 408 p.

CORRÊA, S.T.R. **Simulação estocástica de elementos do clima para estimação da produtividade de cana planta**. 2013. 177 p. Tese (Doutorado em Fitotecnia) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2013.

SANSIGOLO, C.S.; NERY, J.T. Distribuição de extremos de temperatura mínima no Estado do Paraná, **Revista Brasileira de Agrometeorologia**, Santa Maria, v. 8, n. 2, p. 247-253, 2000.

VILLA NOVA, N.A. **Dados meteorológicos do município de Piracicaba**. Piracicaba: ESALQ, Departamento de Ciência Exatas, 2003. 2 p.