

## MODELOS DECOMPOSIÇÃO TEMPORAL E DE REGRESSÃO HARMÔNICA: UMA COMPARAÇÃO PARA A SÉRIE MENSAL DA TEMPERATURA MÍNIMA E MÁXIMA MÉDIA DE PRESIDENTE PRUDENTE (SP)

MIRIAM RODRIGUES SILVESTRE, MANOEL IVANILDO SILVESTRE BEZERRA

Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), Faculdade de Ciências e Tecnologia (FCT),  
Departamento de Estatística (Dest), Presidente Prudente, SP, Brasil

miriam@fct.unesp.br, manoel@fct.unesp.br

Recebido Setembro de 2013 - Aceito Abril de 2015

### RESUMO

Nesse trabalho são apresentados dois modelos de regressão construídos para estimar as séries mensais de temperatura mínima e máxima média (Tmin e Tmax) do município de Presidente Prudente (SP). Os modelos de decomposição temporal com tendência linear e sazonalidade, e modelo harmônico com tendência linear e componentes harmônicas, foram construídos para os períodos de 50 anos (1961-2010) e 30 anos (1981-2010). Os modelos foram comparados através de medidas de acurácia nas séries de 50 e 30 anos. As medidas de precisão, exatidão e desempenho foram avaliadas no conjunto de validação (2011-2014). Foi proposto o uso do coeficiente de correlação de Spearman no lugar do coeficiente de Pearson, no caso de falta de normalidade das observações originais e estimativas dos modelos, para avaliar a precisão e o desempenho dos modelos através do índice CS. Observou-se que o modelo harmônico apresentou melhores resultados que o modelo de decomposição temporal, principalmente quando foi construído com a série mais longa (50 anos) para Tmax e mais curta (30 anos) para Tmin.

**Palavras-chave:** temperatura máxima, temperatura mínima, modelos de regressão, modelo de decomposição temporal, modelo harmônico.

**ABSTRACT:** TEMPORAL DECOMPOSITION AND HARMONIC REGRESSION MODELS: A COMPARISON TO THE SERIES OF MONTHLY AVERAGE MINIMUM AND MAXIMUM TEMPERATURE AT PRESIDENTE PRUDENTE (SP)

In this work two regression models, constructed to estimate the monthly minimum and maximum average temperature (Tmin and Tmax) series at Presidente Prudente (SP) city are presented. The temporal decomposition model with linear trend and seasonality and the harmonic model with linear trend and harmonic components were built for periods of 50 years (1961-2010) and of 30 years (1981-2010). The models were compared by accuracy measures on the 50 and 30 years series. The precision and performance measurements were evaluated for the validation data set (2011-2014). It was proposed the Spearman correlation coefficient instead of the Pearson coefficient in the case of non-normality of the original observations and model estimates, and to assess the accuracy and the performance of the models through the CS index. It was observed that the harmonic model showed better results than the temporal decomposition model, especially when it was built using the longer Tmax (50 years) and shorter Tmin series (30 years).

**Keywords:** maximum temperature, minimum temperature, regression model, temporal decomposition model, harmonic model.

## 1. INTRODUÇÃO

Segundo dados divulgados pelo IBGE Países@ (2015), a partir de dados fornecidos pela United Nations Statistical Divisions (UNSD, 2015), a população mundial em agosto de 2014 era constituída de 7.199.722.769 habitantes. O último censo demográfico realizado pelo IBGE no Brasil, publicado na Sinopse do Censo Demográfico 2010 do IBGE (IBGE CENSO, 2013), indicou que a população residente em 2010 era de 190.755.799 milhões de pessoas, e desse total, 160.925.792 habitavam áreas urbanas, o que representa 84,36% da população como sendo urbana. Considerando somente a Região Sudeste, esse percentual sobe para 92,95%.

No município de Presidente Prudente (SP), recorte espacial desse trabalho, segundo o IBGE Cidades@ (2013), a população total residente em 2010 era de 207.610, e a população estritamente urbana era de 203.375 habitantes, ou seja, 97,96%. Essa informação mostra a relevância de estudos realizados em áreas urbanas devido ao grande contingente populacional vivendo em cidades, o qual é afetado por diversos fatores, entre eles, os fatores relacionados ao clima. Uma importante variável nesse contexto é a temperatura, a qual influencia o organismo humano, e suas atividades tanto comerciais, assim como industriais ou agrícolas. Nesse trabalho, são abordadas as séries de temperatura mínima e máxima média mensal (Tmin e Tmax) para o município de Presidente Prudente (SP), do período 1961 a 2010.

Segundo Morettin e Tolo (2004), “uma série temporal é qualquer conjunto de observações ordenadas no tempo”. Para a análise das séries temporais podem ser empregadas duas classes de modelos: os modelos paramétricos, que contém um número finito de parâmetros; e os não paramétricos, que envolvem um número infinito de parâmetros. Os modelos paramétricos englobam os modelos de erros (ou de regressão), tais como os modelos de decomposição temporal e modelo de regressão harmônico, os quais serão empregados nesse trabalho.

Para o Estado de São Paulo, Pinto et al. (1972) apresentam modelos de regressão linear múltiplo para estimar as temperaturas médias mensais para o Estado de São Paulo. O modelo proposto depende apenas da altitude (m) e latitude (minutos), e é construída uma equação para cada mês do ano. Entretanto, os autores não trabalham as medidas Tmin e Tmax para que se possa comparar.

Pedro Júnior et al. (1991) também apresentam modelos mensais de regressão envolvendo Tmin e Tmax, para o Estado de São Paulo, com exceção das regiões litorâneas e do Vale do Ribeira, e constataram que há grande influência da altitude nas equações mensais.

Nota-se que nos trabalhos citados acima, Pinto et al. (1972) e Pedro Júnior et al. (1991), são desenvolvidos modelos de regressão

múltipla mensais envolvendo as variáveis altitude e latitude.

Com relação ao modelo harmônico, Cargnelutti Filho et al. (2010) o utilizaram para modelar a temperatura do ar média mensal em 37 locais do Estado do Rio Grande do Sul, para o período de 1931 a 2000.

Antonini et al. (2009) construíram um modelo para a temperatura média diária do Estado de Goiás utilizando o modelo harmônico e as coordenadas geográficas (latitude, longitude e altitude) de 21 estações meteorológicas em Goiás, os períodos variaram conforme a estação, pois as mesmas não continham séries com períodos idênticos e nem muito longas. O modelo foi testado com os dados de três estações de altitudes diferentes: elevada (1.100 m), média (554 m) e baixa (431 m). O desempenho do modelo foi considerado mediano para altitudes baixas e elevadas e muito bom para altitudes médias.

Fante e Sant’Anna Neto (2012) apresentam estatísticas para sete estações do estado de São Paulo (Avaré, Catanduva, Franca, Presidente Prudente, Votuporanga, Piracicaba e São Simão) compreendendo o período de 1961 a 2009. Os resultados obtidos apontaram para um aumento de Tmin e Tmax nos dois últimos decênios (1991-2000 e 2001-2009), quando comparadas às médias de todo o período (1961-2009) para a maioria das estações analisadas, com exceção de Catanduva (maior aumento para Tmax no período 1961-1970). Para Presidente Prudente esse aumento foi de 0,6°C e 1°C para Tmin e de 0,2°C e 0,6°C para Tmax, para cada decênio, respectivamente. Entretanto, esses autores não desenvolvem nenhum modelo para ajustar temperatura.

Nesse trabalho pretende-se desenvolver dois modelos de regressão: o modelo de decomposição com tendência e sazonalidade determinística e o modelo com tendência linear e harmônicas periódicas para as séries Tmin e Tmax, específicos para a estação de Presidente Prudente (SP).

## 2. MATERIAL E MÉTODOS

Os dados foram obtidos junto ao Instituto Nacional de Meteorologia (INMET, 2013) e AGRITEMPO (2015), referentes à Estação Meteorológica de Presidente Prudente – SP, a qual está situada nas coordenadas de 22° 12’ de latitude sul, 51° 38’ de longitude oeste e a uma altitude de 435,55 metros acima do nível médio do mar; e tem o código 83716 como seu número de localização. Dos dados disponíveis no INMET para a estação convencional utilizou-se a série de janeiro de 1961 a maio de 2012, e devido à inexistência de dados posteriores, os meses subsequentes de junho de 2012 a dezembro de 2014 foram obtidos junto ao site do AGRITEMPO (2015), para a mesma estação meteorológica.

Os dados foram organizados em planilhas no software Microsoft Excel®, e posteriormente foi realizada a elaboração

dos modelos de regressão utilizando-se os softwares estatísticos Minitab® e R, esse último de acesso livre.

Primeiramente, foram construídos gráficos de linhas das séries mensais de Tmin e Tmax com o objetivo de verificar a existência de sazonalidade, e gráficos de autocorrelações das observações, utilizados para orientar o pesquisador na escolha do modelo mais adequado à série temporal, considerando-se as duas opções possíveis: determinístico ou estocástico. Com a análise de ambos os gráficos, detectou-se a presença de sazonalidade nas séries Tmin e Tmax, indicando que poderiam ser construídos modelos de regressão determinísticos que incorporassem a sazonalidade identificada. Os modelos estatísticos determinísticos adotados foram: modelo de regressão de decomposição temporal e o modelo de regressão harmônico. E para a construção dos modelos foram considerados dois períodos: de janeiro de 1961 a dezembro de 2010 (séries de 50 anos) e de janeiro de 1981 a dezembro de 2010 (série de 30 anos). Após, foram construídas medidas para validação dos modelos em dados não utilizados na construção dos mesmos, referentes a um único período de janeiro de 2011 a dezembro de 2014.

Nas próximas seções serão definidos os modelos de regressão empregados, e as medidas para avaliar a acurácia, precisão, exatidão e desempenho dos mesmos.

### 2.1 Modelo de Regressão de Decomposição Temporal

Para o ajuste de um modelo que representasse adequadamente os dados de Tmin e Tmax, foi aplicada a técnica estatística de análise de regressão, porém de forma a contemplar a sazonalidade mensal observada. Maiores detalhes sobre a técnica podem ser encontrados em Morettin e Tolo (2004) e Draper e Smith (1981). Foi construído um modelo matricial, definido na Equação 1:

$$Y = X\gamma + \varepsilon, \tag{1}$$

no qual Y representa o vetor de observações da variável de interesse, constituído de N=50x12=600 observações para uma série mensal de 50 anos; X = [C : D] é uma matriz composta por:

$$C = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & N \end{bmatrix} \text{ e } D = \begin{bmatrix} D_1 \\ \vdots \\ D_k \\ \vdots \\ D_{50} \end{bmatrix}, \text{ para}$$

$$D_k = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ -1 & -1 & -1 & \dots & -1 & -1 \end{bmatrix} \begin{matrix} \leftarrow \text{jan} \\ \leftarrow \text{fev} \\ \vdots \\ \leftarrow \text{nov} \\ \leftarrow \text{dez} \end{matrix},$$

$\gamma = \begin{bmatrix} \beta \\ \alpha \end{bmatrix}$  é o vetor de parâmetros do modelo e  $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix}$  é o vetor

de erros. O vetor de parâmetros  $\gamma$  é constituído de duas partes, a primeira parte representando a tendência linear  $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ , sendo que  $\beta_0$  é a constante e  $\beta_1$  a coeficiente angular; e a segunda parte a sazonalidade mensal da série dada por  $\alpha' = [\alpha_1 \dots \alpha_{11}]$ . O modelo para cada observação  $t$  será definido como:

$$Y_t = \beta_0 + \beta_1 t + \sum_{i=1}^{11} \alpha_i d_{it} + \varepsilon_t, \tag{2}$$

para  $d_i = \begin{cases} 1 & \text{para } i = \text{janeiro a novembro} \\ -1 & \text{para } i = \text{dezembro} \\ 0 & \text{caso contrário} \end{cases}$ .

O vetor de observações estimado  $\hat{Y} = X\hat{\gamma}$  é obtido através do vetor de estimativas de mínimos quadrados dos parâmetros, o qual é dado por:  $\hat{\gamma} = (X'X)^{-1} X'Y$ .

A significância estatística do modelo é verificada através do teste de hipóteses:  $H_0: \beta_1 = \alpha_1 = \dots = \alpha_{11} = 0$  versus  $H_1$ : pelo menos um parâmetro é diferente de zero. Para realizar o teste de hipóteses, calcula-se a estatística do teste F, a qual apresenta, sob  $H_0$ , uma distribuição F de Fisher-Snedecor, ou seja,  $F \sim F(gl_1=p-1, gl_2=n-p, 1-\alpha)$ . Rejeita-se  $H_0$  ao nível de significância  $\alpha$  se F observado for igual ou superior a  $F(gl_1, gl_2, 1-\alpha)$  tabelado; ou se o P-valor for suficientemente pequeno, nesse caso, menor ou igual ao nível de significância  $\alpha=0,05$  definido para esse trabalho.

### 2.2 Modelo de Regressão Harmônico

Considere o modelo de regressão matricial, como definido anteriormente na Equação 1, no qual Y representa o vetor de observações da variável de interesse. Suponha agora uma matriz de delineamento  $X = [T : H]$ , tal que:

$$T = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & t \\ \vdots & \vdots \\ 1 & N \end{bmatrix},$$

$$H = \begin{bmatrix} \cos(\omega_1) & \dots & \cos(\omega_6) & \text{sen}(\omega_1) & \dots & \text{sen}(\omega_6) \\ \cos(\omega_1,2) & \dots & \cos(\omega_6,2) & \text{sen}(\omega_1,2) & \dots & \text{sen}(\omega_6,2) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \cos(\omega_1,t) & \dots & \cos(\omega_6,t) & \text{sen}(\omega_1,t) & \dots & \text{sen}(\omega_6,t) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \cos(\omega_1,N) & \dots & \cos(\omega_6,N) & \text{sen}(\omega_1,N) & \dots & \text{sen}(\omega_6,N) \end{bmatrix},$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix}, \text{ e } \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \alpha_1 \\ \vdots \\ \alpha_6 \\ \gamma_1 \\ \vdots \\ \gamma_6 \end{bmatrix} = \begin{cases} \text{Intercepto} \\ \text{tendência linear} \\ \cos(\omega_1 t) \\ \vdots \\ \cos(\omega_6 t) \\ \text{sen}(\omega_1 t) \\ \vdots \\ \text{sen}(\omega_6 t) \end{cases},$$

$$\text{e } \omega_k = \frac{2\pi k}{12}, k = 1, \dots, 6 \text{ e } t = 1, \dots, N.$$

Considere  $\boldsymbol{\beta}$  como sendo o vetor de parâmetros do modelo, contendo  $p = 14$  parâmetros; e  $\boldsymbol{\varepsilon}$  o vetor de erros. Portanto, o modelo de regressão com harmônicas é definido como:

$$Y_t = \beta_0 + \beta_1 t + \sum \alpha_k \cos(\omega_k t) + \sum \gamma_k \text{sen}(\omega_k t) + \varepsilon_t \quad (3)$$

Considerando as  $N=600$  observações para a série de 50 anos, obtém-se o vetor de estimativas de mínimos quadrados dos parâmetros, dado por:  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ , e o vetor de observações estimado,  $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ . Foi avaliada a significância estatística do modelo através do teste de hipóteses:

$$H_0 : \beta_1 = \alpha_1 = \dots = \alpha_6 = \gamma_1 = \dots = \gamma_6 = 0$$

versus  $H_1$ : pelo menos um parâmetro é diferente de zero. Como visto na Seção 2.1, para realizar o teste de hipóteses, calcula-se a estatística do teste F, a qual apresenta sob  $H_0$  a distribuição F de Fisher-Snedecor, ou seja,  $F \sim F(g_1=p-1, g_2=n-p, 1-\alpha)$ . A regra de decisão para a rejeição ou não da hipótese  $H_0$  será a mesma adotada para o modelo descrito na seção 2.1.

### 2.3 Medidas utilizadas para avaliar a acurácia do modelo

Para a avaliação da acurácia dos modelos foram utilizadas as medidas do Erro Médio Absoluto (MAE), Erro Quadrático Médio (MSE), e raiz quadrada do MSE (RMSE), essas medidas podem ser encontradas em Wilks (2006, p. 278). As medidas MAE e MSE utilizam a estimativa  $\hat{Y}_k$  obtida com a construção do modelo e a verdadeira observação  $Y_k$ . Quanto mais próxima de zero está a medida MAE melhor é a previsão. A medida MSE é bem parecida à medida MAE, exceto pelo emprego da função quadrática do erro, ao invés da função do erro absoluto. Essa medida é mais sensível aos erros maiores que a função MAE, e também, mais sensível aos outliers. O valor da medida MSE será sempre positivo, devido à função quadrática, e quanto menor seu valor mais acurado é o modelo que gerou os erros. A medida RMSE é a raiz quadrada de MSE, e a vantagem em sua utilização é que esta medida tem a

mesma dimensão física das observações, no caso da variável temperatura, em °C, por isso sua mais interpretação é mais fácil. As expressões matemáticas das medidas MAE, MSE e RMSE serão apresentadas posteriormente na Tabela 1 da seção 3.3.

### 2.4 Medidas para avaliar a precisão, exatidão e desempenho dos modelos

Os modelos descritos nas seções 2.1 e 2.2 serão avaliados com relação a sua qualidade preditiva através da precisão, exatidão e desempenho. Para avaliar a precisão, os autores Cargnelutti Filho, et al. (2010) e Antonini et al. (2009) empregam o coeficiente de correlação de Pearson, definido na Equação 4, a seguir:

$$\frac{\sum_{k=1}^n (\hat{Y}_k - \bar{\hat{Y}})(Y_k - \bar{Y}_k)}{\sqrt{\sum_{k=1}^n (\hat{Y}_k - \bar{\hat{Y}})^2} \sqrt{\sum_{k=1}^n (Y_k - \bar{Y}_k)^2}} \quad (4)$$

Porém, para avaliar se a correlação é realmente significativa, podem ser realizados testes estatísticos como o teste paramétrico de Correlação de Pearson e o não paramétrico de Correlação de Spearman (Conover, 1999). O Teste de Correlação de Pearson utiliza a estatística (4) e é indicado quando as duas variáveis que estão sendo correlacionadas apresentam distribuição normal bivariada. Em caso contrário, o Teste de Correlação de Spearman é preferível por não necessitar da normalidade das variáveis envolvidas, sua estatística é definida na Equação 5 e envolve os postos das observações no lugar das observações originais, tal que  $R(\cdot)$  significa a ordem ou posto do valor entre parênteses para todos os valores da variável:

$$\frac{\sum_{k=1}^n R(\hat{Y}_k) - R(Y_k) - n\left(\frac{n+1}{2}\right)^2}{\sqrt{\sum_{k=1}^n R(\hat{Y}_k)^2 - n\left(\frac{n+1}{2}\right)^2} \sqrt{\sum_{k=1}^n R(Y_k)^2 - n\left(\frac{n+1}{2}\right)^2}} \quad (5)$$

A proposta que se apresenta nesse artigo é a de que se deva avaliar qual dos dois testes entre Pearson e Spearman é o mais indicado para o conjunto de dados em questão, de acordo com a confirmação ou não da suposição de normalidade das observações. Para tanto, pode ser aplicada uma generalização do Teste de Normalidade de Shapiro-Wilk (Gonzalez-Estrada e Villasenor-Alva, 2013) para normalidade multivariada. O software R apresenta um programa para realizar esse teste (Gonzalez-Estrada e Villasenor-Alva, 2013). Após a realização do teste de Shapiro-Wilk multivariado, caso a normalidade bivariada seja confirmada ( $p\text{valor} > 0,05$ ) pode-se aplicar o Teste de Correlação de Pearson, para avaliar se a correlação encontrada é estatisticamente significativa ( $p\text{valor} \leq 0,05$ ), e

utilizar o seu coeficiente de correlação, definido na Equação 4, como uma medida para avaliar a capacidade preditiva do modelo quanto à precisão. Na falta de normalidade, deverá ser aplicado o teste de Correlação de Spearman, conforme a Equação 5.

A exatidão será avaliada pelo índice de concordância ( $d$ ) de Willmott (1982) dado na Equação 6:

$$d = 1 - \frac{\sum_{k=1}^n (\hat{Y}_k - Y_k)^2}{\sum_{k=1}^n (|\hat{Y}_k - \bar{Y}| + |Y_k - \bar{Y}|)^2} \quad (6)$$

Finalmente, o desempenho do modelo será obtido através do índice de consistência  $CS$  de Camargo e Sentelhas (Cargnelutti Filho et al., 2010), definido por:

$$CS = r \times d, \quad (7)$$

onde  $r$  é o coeficiente de correlação calculado na precisão. A proposta do presente artigo é utilizar em  $r$  o coeficiente de Pearson, definido na Equação 4, para o caso de ocorrer normalidade multivariada, ou o coeficiente de Spearman, dado na Equação 5, na falta dela. O índice  $d$  é o índice de concordância já definido na Equação 6.

### 3. RESULTADOS E DISCUSSÕES

Uma análise inicial dos dados de Tmin e Tmax para os períodos de 50 anos (1961-2010) e 30 anos (1981-2010) apresentam sazonalidade mensal, conforme pode ser observado no gráfico de linhas das séries mensais apresentado na Figura 1, no qual nota-se que as medidas de Tmax dos meses de maio, junho e julho são reduzidas quando comparadas às que ocorrem para os demais meses do ano. A partir de agosto a temperatura máxima apresenta elevação com pico em fevereiro, e após é observada uma tendência de decréscimo entre março e abril. Para Tmin o comportamento é similar. Observa-se que praticamente não há diferença entre os dois períodos com relação à Tmax, porém, para Tmin o período mais recente (1981-2010) apresenta temperaturas médias do ar mais elevadas do que o período completo (1961-2010).

Nas próximas seções 3.1 e 3.2 serão apresentados os modelos construídos para o ajuste das séries Tmax, e na seção 3.3 as previsões para os 48 meses seguintes, a partir de janeiro de 2011 até dezembro de 2014. As seções 3.4, 3.5 e 3.6 serão referentes às séries Tmin.

#### 3.1 Modelo de Regressão de Decomposição Temporal para série Tmax

Foi construído o Modelo de Decomposição Temporal, denominado Modelo 1, conforme definições na Seção 2.1 para

a série Tmax referente ao período 1961-2010. Após a análise de variância inicial, retirou-se os parâmetros não significativos e o modelo final estimado é apresentado na Equação 8.

$$\begin{aligned} \hat{Y}_t = & 28,8735336 + 0,0007126t + 1,7483027d_{1t} + \\ & + 1,9635902d_{2t} + 1,6368776d_{3t} + \\ & + 0,404165d_{4t} - 2,5265474d_{5t} - 3,3552602d_{6t} \\ & - 3,0099728d_{7t} - 0,8946854d_{8t} + \\ & + 1,0838894d_{10t} + 1,4951769d_{11t}. \end{aligned} \quad (8)$$

Para avaliar a qualidade do modelo ajustado, geralmente utiliza-se o coeficiente de determinação múltiplo  $R^2$ . Nesse modelo obteve-se  $R^2=0,6463$ , isso significa que o modelo ajustado explica cerca de 64,63% da variação total da Tmax. O coeficiente  $R^2$  ajustado resultou em 63,97%. A normalidade dos resíduos foi comprovada através da aplicação do teste de Shapiro-Wilk (Estatística  $W=0,9953$  e  $p$ -valor=0,06785).

O Modelo 1 para a série Tmax referente ao período 1981-2010, a análise de significância dos componentes, também indicou a retirada de  $d_{9t}$ . A equação do modelo final é apresentada na Equação 9:

$$\begin{aligned} \hat{Y}_t = & 28,8840216 + 0,0016613t + 1,8248203d_{1t} + \\ & + 1,8764923d_{2t} + 1,4948310d_{3t} + \\ & + 0,4765030d_{4t} - 2,7884917d_{5t} - 3,4901530d_{6t} \\ & - 3,1651476d_{7t} - 0,8834756d_{8t} + \\ & + 1,2865351d_{10t} + 1,7582071d_{11t}. \end{aligned} \quad (9)$$

O coeficiente de determinação múltiplo  $R^2=68,51\%$  e o coeficiente  $R^2_{ajust}=67,51\%$ , isso significa que o modelo ajustado explica cerca de 67,51% da variação total de Tmax. A normalidade dos resíduos foi comprovada através da aplicação do teste de Shapiro-Wilk (Estatística  $W=0,9954$  e  $p$ -valor=0,3716).

#### 3.2 Modelo de Regressão Harmônico para a série Tmax

O modelo harmônico, denominado Modelo 2, para a série 1961-2010 após a retirada dos componentes harmônicos não significativos é apresentado na Equação 10. Os parâmetros significativos foram  $\hat{\beta}_0$ ,  $\hat{\beta}_1$ ,  $\hat{\alpha}_1$ ,  $\hat{\alpha}_2$ ,  $\hat{\alpha}_3$ ,  $\hat{\alpha}_6$ ,  $\hat{\gamma}_1$  e  $\hat{\gamma}_3$ .

$$\begin{aligned} \hat{Y}_t = & 29,8738397 + 0,0007116t + 2,3637389\cos(\omega_1t) - \\ & - 0,9658782\cos(\omega_2t) + \\ & + 0,2119551\cos(\omega_3t) + 0,1233109\cos(\omega_6t) + \\ & + 0,6765348\sin(\omega_1t) - 0,1779551\sin(\omega_3t). \end{aligned} \quad (10)$$

O modelo ajustado apresentou  $R^2=64,22\%$  e o  $R^2_{ajust}=63,79\%$ . O teste de normalidade de Shapiro-Wilk indicou normalidade dos resíduos ( $W=0,9951$  e  $p$ -valor=0,05626).

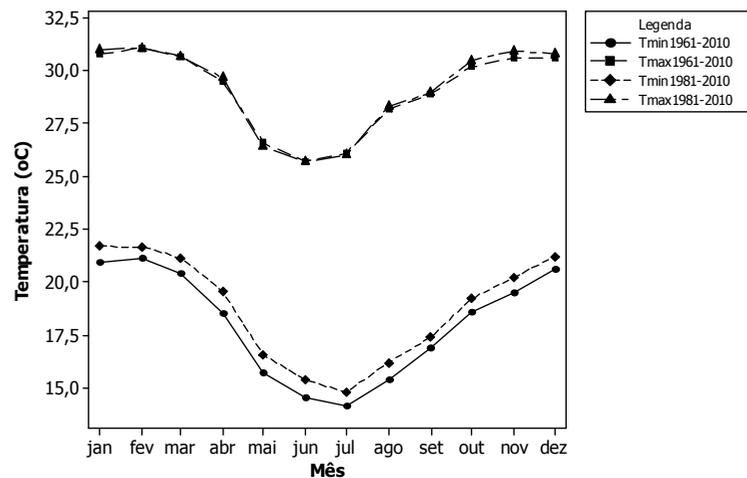


Figura 1 - Gráfico de linhas para Tmax e Tmin de Presidente Prudente (SP).

Para a série Tmax referente ao período de 1981-2010 o Modelo 2 ajustado é apresentado na Equação 11:

$$\begin{aligned} \hat{Y}_t = & 28,8810874 + 0,0016776t + 2,5239884 \cos(\omega_1 t) - \\ & - 0,9488998 \cos(\omega_2 t) + \\ & + 0,2549891 \cos(\omega_3 t) - 0,2290102 \cos(\omega_6 t) + \\ & + 0,6084434 \operatorname{sen}(\omega_1 t) - 0,2161002 \operatorname{sen}(\omega_3 t). \end{aligned} \quad (11)$$

O Modelo 2 apresentou  $R^2=67,68\%$  e o  $R^2_{\text{ajust}}=67,04\%$ . O teste de Shapiro-Wilk indicou normalidade dos resíduos ( $W=0,9945$  e  $p\text{-valor}=0,2283$ ).

### 3.3 Previsões futuras e avaliação dos modelos para Tmax

Foram realizadas previsões para os próximos 48 meses, a partir de janeiro de 2011 a dezembro de 2014, para os modelos construídos nas Seções 3.1 e 3.2 para as séries Tmax.

A seguir, é apresentada a Figura 2, com um gráfico de linhas das três séries (dados reais, previsão pelo Modelo 1 e previsão pelo Modelo 2) para uma comparação visual.

A acurácia dos dois modelos gerados foi avaliada pelas estatísticas apresentadas na Seção 2.3. Os valores estimados para as medidas de acurácia, construídos com os dados observados de Tmax para o período de validação do modelo, ou seja, de janeiro de 2011 a dezembro de 2014, estão apresentados na Tabela 1.

Analisando as medidas para a série de 50 anos (1961-2010), apresentadas na Tabela 1 para ambos os modelos, nota-se que o Modelo Harmônico (Modelo 2) apresentou menores valores para todas as medidas quando comparado ao Modelo de Decomposição Temporal (Modelo 1). Dessa forma, o Modelo 2 pode ser considerado o melhor modelo entre os dois.

Já para a série de 30 anos (1981-2010), o Modelo 2 apresentou a melhor (menor) medida de MAE, mas para a medida RMSE, o Modelo 1 apresentou um erro de aproximadamente  $0,99^\circ\text{C}$  enquanto o Modelo 2 obteve um erro de  $1,05^\circ\text{C}$ . Portanto o Modelo 1 é o mais adequado para esse período.

Finalizando ao comparar-se as duas séries e modelos construídos, conclui-se que a série de 50 anos, representada pelo Modelo 2, apresenta melhores resultados quanto à previsão de novas observações de Tmax.

De acordo com a última coluna da Tabela 2, todos os testes de Shapiro-Wilk Bivariado realizados indicaram que a suposição de normalidade bivariada dos pares de observações foram rejeitadas. Assim, para o cálculo de correlação, será utilizado o Teste de Correlação não-paramétrico de Spearman, indicado para situações onde não há normalidade bivariada.

Também foram calculadas as medidas  $r$  de Correlação de Spearman, os índices  $d$  de Willmott e CS de Camargo e Sentelhas, para os valores estimados pelos modelos ajustados e os valores reais observados, para as  $n=48$  observações. A Tabela 3 apresenta os resultados obtidos para os índices calculados.

Analisando os índices apresentados na Tabela 3 pode-se concluir que para a série 1961-2010, o Modelo 2 tem maior exatidão ( $d$ ) e de acordo com o Índice CS pode ser considerado de bom desempenho, segundo a tabela de classificação para esse Índice CS, apresentada em Antonini et al. (2010). Já para o período de 30 anos (1981-2010), o Modelo 1 é mais indicado e também apresenta bom desempenho segundo o Índice CS.

Comparando ambos os modelos e períodos conclui-se que o melhor modelo para Tmax é o Modelo 2 para a série de 50 anos (1961-2010).

Uma observação quanto ao emprego do coeficiente de correlação de Pearson é que se o mesmo é incorretamente empregado, ou seja, na falta de normalidade, a precisão dos

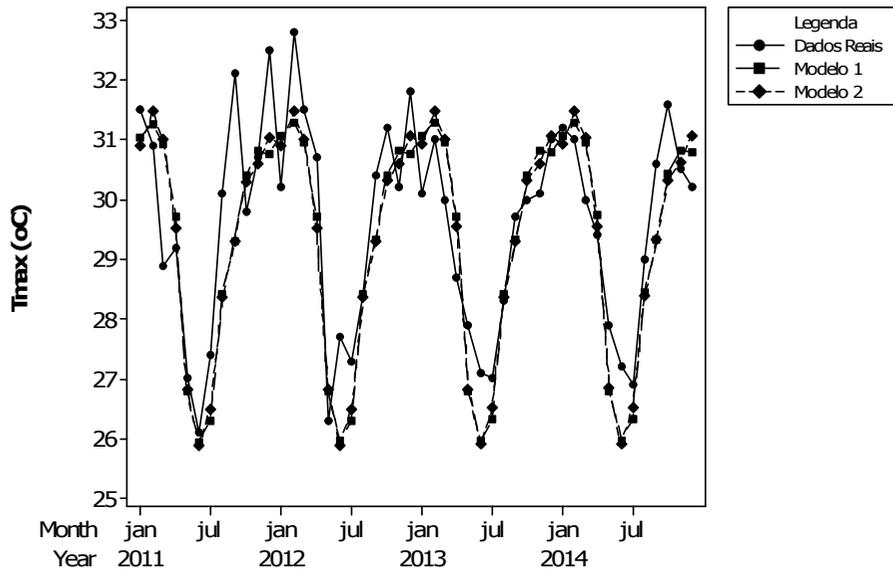


Figura 2 - Série Tmax original e previsão dos Modelos 1 e 2 para 2011-2014, geradas a partir da série de 50 anos.

Tabela 1 - Estatísticas para avaliar a acurácia das previsões para séries Tmax de 50 e 30 anos

Estatística	1961-2010		1981-2010	
	Modelo 1	Modelo 2	Modelo 1	Modelo 2
$MAE = \frac{1}{n} \sum_{k=1}^n  \hat{Y}_k - Y_k $	0,8123851	<b>0,7831448</b>	0,844487	<b>0,8198095</b>
$MSE = \frac{1}{n} \sum_{k=1}^n (\hat{Y}_k - Y_k)^2$	0,9732278	<b>0,9345541</b>	<b>0,97273</b>	1,102219
$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (\hat{Y}_k - Y_k)^2}$	0,986523	<b>0,9667234</b>	<b>0,9862707</b>	1,049866

Tabela 2 - Testes de Shapiro-Wilk bivariados para as séries Tmax e previsões de Tmax utilizando o Modelo de Decomposição Temporal (Modelo 1) e o Modelo Harmônico (Modelo 2) para série de 50 e 30 anos.

Par de séries Tmax (dados reais, previsão para 2011-2014)	Shapiro-Wilk Bivariado			
	1961-2010		1981-2010	
	Estatística	Pvalor	Estatística	Pvalor
Modelo 1: $(Y_{obs}, \hat{Y}_1)$	0,9135	9,467e-06*	0,9016	2,532e-06*
Modelo 2: $(Y_{obs}, \hat{Y}_2)$	0,9236	3,132e-05*	0,9252	3,857e-05*

\*=significativo a 0,05 (ou 5%)

Tabela 3 - Correlação (precisão), Índices d de Willmott (exatidão) e Índice CS de Camargo e Sentelhas (desempenho) avaliados sobre os dados reais e os valores previstos para série Tmax, para os Modelos 1 e 2 para o período 2011-2014.

Série Tmax	Modelo	Spearman (r)	p-valor	Índice d	Índice CS	Desempenho
1961-2010	1	0,7392947	1,959725e-09	0,9218907	0,6815489	Bom
	2	<b>0,7738345</b>	1,123277e-10	<b>0,9249693</b>	<b>0,7157732</b>	<b>Bom*</b>
1981-2010	1	<b>0,7697071</b>	1,621938e-10	<b>0,9253875</b>	<b>0,7122773</b>	<b>Bom</b>
	2	0,7029083	2,548877e-08	0,9142635	0,6426434	Mediano

\* melhor modelo

modelos é superestimada, já que o índice de Correlação de Pearson apresenta valores maiores que o de Spearman para os mesmos dados e, portanto, o índice de desempenho CS também superestima o desempenho do modelo ao realizar previsões futuras.

### 3.4 Modelo de Regressão de Decomposição Temporal para série Tmin

Foi construído o Modelo de Decomposição Temporal, conforme definições na Seção 2.1 para a série Tmin referente ao período 1961-2010. Após a análise de variância inicial do modelo completo não houve normalidade dos resíduos, foi realizada a retirada dos parâmetros não significativos, mas não foi encontrado nenhum modelo que produzisse resíduos com distribuição normal. Portanto, para a série Tmin com período de 50 anos não foi possível apresentar um Modelo de Decomposição Temporal adequado.

Para o período de 30 anos (1981-2010) foi possível ajustar um Modelo 1 adequado, conforme a Equação 12.

$$\begin{aligned} \hat{Y}_t = & 18,3671016 + 0,0021167t + 2,9558085d_{1t} + \\ & + 2,9236918d_{2t} + 2,4182418d_{3t} + \\ & + 0,8227918d_{4t} - 2,2093249d_{5t} - 3,3881083d_{6t} - \\ & - 3,960225d_{7t} - 2,5723417d_{8t} + \\ & - 1,3511251d_{9t} + 0,4534249d_{10t} + 1,4679748d_{11t}. \end{aligned} \quad (12)$$

O modelo ajustado apresentou  $R^2 = 84,76\%$  e o  $R^2_{ajust} = 84,23\%$ . O teste de normalidade de Shapiro-Wilk indicou normalidade dos resíduos ( $W=0,994$  e  $pvalor=0,1692$ ).

### 3.5 Modelo de Regressão Harmônico para a série Tmin

Foi construído o Modelo de Regressão Harmônico, conforme definições na Seção 2.1 para a série Tmin referente ao período 1961-2010. Nenhum modelo de regressão harmônico foi adequado para essa série, pois os resíduos produzidos não apresentaram distribuição normal.

Para o período de 30 anos (1981-2010) foi possível ajustar um Modelo 2 adequado, conforme a Equação 13.

$$\begin{aligned} \hat{Y}_t = & 18,3668070 + 0,0021183t + \\ & + 2,9274532 \cos(\omega_1 t) - 0,6171183 \cos(\omega_2 t) + \\ & + 0,1131075 \cos(\omega_3 t) + 1,7447260 \sin(\omega_1 t) - \\ & - 0,1714605 \sin(\omega_2 t). \end{aligned} \quad (13)$$

O modelo ajustado apresentou  $R^2 = 84,40\%$  e o  $R^2_{ajust} = 84,14\%$ . O teste de normalidade de Shapiro-Wilk indicou normalidade dos resíduos ( $W=0,9941$  e  $pvalor=0,1763$ ).

### 3.6 Previsões futuras e avaliação dos modelos para Tmin

Foram realizadas previsões para os próximos 48 meses, a partir de janeiro de 2011 a dezembro de 2014, para os modelos construídos nas Seções 3.1 e 3.2 para as séries Tmin.

A seguir é apresentada a Figura 3, com um gráfico de linhas das três séries (dados reais, previsão pelo Modelo 1 e previsão pelo Modelo 2) para uma comparação visual.

A acurácia dos dois modelos gerados foi avaliada pelas estatísticas apresentadas na Seção 2.3. Os valores estimados para as medidas de acurácia, construídos com os dados observados de Tmin para o período de validação do modelo, ou seja, de janeiro de 2011 a dezembro de 2014, estão apresentados na Tabela 4.

A Tabela 5 apresenta os resultados dos testes de Shapiro-Wilk para avaliar normalidade bivariada. Nota-se que em ambos os modelos a suposição de normalidade é rejeitada.

Também foram calculados os índices  $r$  de Correlação de Spearman,  $d$  de Willmott e CS de Camargo e Sentelhas, para os valores estimados pelos modelos ajustados e os valores reais observados, para as  $n = 48$  observações. A Tabela 6 apresenta os resultados obtidos para os índices calculados.

Analisando as medidas apresentadas na Tabela 6, nota-se que os modelos 1 e 2 apresentaram comportamento bem próximos, porém, o melhor modelo entre os dois avaliados foi o Modelo Harmônico (Modelo 2).

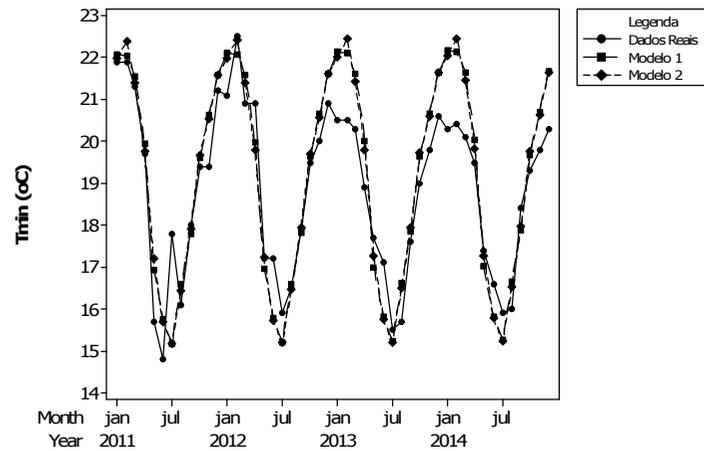
Não foi possível realizar comparação dos resultados obtidos nesse trabalho com outras referências, pelo fato de não ter sido encontrada nenhuma referência que ajustasse os modelos de regressão de decomposição temporal e o modelo harmônico propostos nesse artigo, para dados de Presidente Prudente e para as variáveis Tmin e Tmax abordadas.

## 4. CONCLUSÕES

A partir da técnica de análise de regressão foram construídos dois modelos de previsão para as séries de temperatura mínima e máxima média mensal do município de Presidente Prudente, referentes aos períodos de 50 anos (1961-2010) e 30 anos (1981-2010). O primeiro um modelo de decomposição com tendência linear e sazonalidade determinística e o segundo um modelo harmônico com tendência

**Tabela 4** - Estatísticas para avaliar a acurácia das previsões para séries Tmin para 30 anos.

Estatística	1981-2010	
	Modelo 1	Modelo 2
MAE	0,7953403	<b>0,7547436</b>
MSE	0,9350242	<b>0,9260304</b>
RMSE	0,9669665	<b>0,9623047</b>



**Figura 3** - Série Tmin original e previsão dos Modelos 1 e 2 para 2011-2014, geradas a partir da série de 30 anos (1981-2010).

**Tabela 5** - Apresenta os resultados dos testes de Shapiro-Wilk para avaliar normalidade bivariada. Nota-se que em ambos os modelos a suposição de normalidade é rejeitada.

Par de séries Tmax (dados reais, previsão para 2011-2014)	Shapiro-Wilk Bivariado (1981-2010)	
	Estatística	Pvalor
Modelo 1: $(Y_{obs}, \hat{Y}_1)$	0,9293	6.447e-05*
Modelo 2: $(Y_{obs}, \hat{Y}_2)$	0,937	0,0001749*

\*=significativo a 0,05 (ou 5%)

**Tabela 6** - Correlação (precisão), Índices d de Willmott (exatidão) e Índice CS de Camargo e Sentelhas (desempenho) avaliados sobre os dados reais e os valores previstos para série Tmin para os Modelos 1 e 2 para o período 2011-2014.

Série Tmin	Modelo	Spearman (r)	p-valor	Índice d	Índice CS	Desempenho
1981-2010	1	0,9042138	0	0,9518272	0,8606553	Ótimo
	2	<b>0,9134992</b>	0	<b>0,9521875</b>	<b>0,8698225</b>	Ótimo*

\* melhor modelo

linear. Analisando separadamente o primeiro modelo, observa-se que a tendência linear foi significativa para ambas as séries Tmin e Tmax. Esse modelo também detectou a sazonalidade mensal determinística, o que já era esperado, pois há um comportamento sazonal da temperatura em geral. No segundo modelo empregou-se harmônicas juntamente com um termo linear, e nesse modelo a tendência linear também foi significativa para ambas as séries. Quanto ao desempenho dos modelos construídos, conclui-se para Tmax, que o melhor modelo foi o modelo harmônico e utilizando a série mais longa (1961-2010), tendo obtido um desempenho classificado como “bom” segundo o índice CS. Para Tmin somente para a série mais curta de 30 anos (1981-2010) houve um ajuste adequado dos modelos apresentando desempenhos “ótimo” e bem próximos, porém, o modelo harmônico apresentou resultados levemente superiores. A proposta de incluir uma análise da existência de normalidade bivariada para decidir qual coeficiente de correlação, entre Pearson ou Spearman, deve ser utilizado, foi satisfatória, pois

a utilização do coeficiente de Pearson, quando a suposição de normalidade não é satisfeita superestima o índice CS que avalia a capacidade dos modelos.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

- AGRITEMPO. Sistema de monitoramento agrometeorológico. Disponível em: <<http://www.agritempo.gov.br>>. Acesso em: 9 mar. 2015.
- ANTONINI, J. C. A. et al. Modelo matemático para estimativa da temperatura média diária do ar no Estado de Goiás. **Pesquisa Agropecuária Brasileira**, Brasília, v. 44, n. 4, p. 331-338, abr. 2009. Disponível em: <<http://seer.sct.embrapa.br/index.php/pab/article/view/1703/5684>>. Acesso em: 19 Jul. 2013.
- ANTONINI, J. C. A. et al. Modelo de estimativa da temperatura média diária do ar no estado de Goiás. Planaltina, DF: **Embrapa Cerrados**, 2010, 27p.

- CARGNELUTTI FILHO, A. et al. Modelo harmônico para a estimativa da temperatura do ar média mensal em diferentes locais do Estado do Rio Grande do Sul. **Ciencia Rural**, Santa Maria, v. 40, n. 9, Set. 2010. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-84782010000900006&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-84782010000900006&lng=en&nrm=iso)>. Acesso em: 19 Jul. 2013.
- CONOVER, W. J. **Practical nonparametric statistics**. 3<sup>rd</sup>. ed. New York: John Wiley and Sons, 1999, 584 p.
- DRAPER, N. R.; SMITH, H. **Applied regression analysis**. 2<sup>nd</sup>. ed. New York: John Wiley & Sons. 1981. 709 p.
- FANTE, K. P.; SANT'ANNA NETO, J. L. Variabilidade da temperatura do ar em cidades interioranas do Estado de São Paulo (1961-2009). **Revista Geonorte**, Ed. Especial 2, v.2, n.5, p.675-687, 2012.
- GONZALEZ-ESTRADA E., VILLASENOR-ALVA J. A. Generalized Shapiro-Wilk test for multivariate normality. <<http://cran.r-project.org/web/packages/mvShapiroTest/mvShapiroTest.pdf>>, 2013. Acesso em: 26 ago. 2013.
- IBGE Censo: Sinopse do Censo Demográfico 2010. Disponível em: <[http://www.ibge.gov.br/home/estatistica/populacao/censo2010/tabelas\\_pdf/Brasil\\_tab\\_1\\_15.pdf](http://www.ibge.gov.br/home/estatistica/populacao/censo2010/tabelas_pdf/Brasil_tab_1_15.pdf)>. Acesso em 18 jul. 2013.
- IBGE Cidades@. Disponível em: <<http://www.ibge.gov.br/cidadesat/xtras/temas.php?codmun=354140&idtema=1&search=sao-paulo|presidente-prudente|censo-demografico-2010:-sinopse->>>. Acesso em: 18 jul. 2013.
- IBGE Países@. Disponível em: <[http://www.ibge.gov.br/paisesat/main\\_frameset.php](http://www.ibge.gov.br/paisesat/main_frameset.php)>. Acesso em: 18 fev. 2015.
- INMET. Instituto Nacional de Meteorologia. Disponível em: <<http://www.inmet.gov.br>>. Acesso em: Acesso em 9 fev. 2013.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. São Paulo: Edgard Blucher. 2004. 535 p.
- PEDRO JÚNIOR, M. J. et al. Estimativa das temperaturas médias mensais das máximas e das mínimas no estado de São Paulo. Campinas: Instituto Agrônomo de Campinas. 1991. 11p. **Boletim Técnico**, 142.
- PINTO, H. S.; ORTOLANI, A. A.; ALFONSI, R. R. Estimativa das temperaturas médias mensais do estado de São Paulo, em função de altitude e latitude. **Caderno de Ciências da Terra**. v.23, 20p. 1972.
- UNSD. United Nations Statistical Divisions – Social and Demographic and Social Statistics. Disponível em: <<http://unstats.un.org/unsd/demographic/products/socind/default.htm>>. Acesso em: 18 fev. 2015.
- WILKS, D. S. **Statistical methods in the atmospheric sciences**. 2<sup>nd</sup>. ed. Amsterdam: Academic Press, 2006. 627p.
- WILLMOTT, C.J. Some comments on the evaluation of model performance. **Bulletin of the American Meteorological Society**, v.63, p.1309-1313, 1982. apud ANTONINI, J. C. A. et al. Modelo de estimativa da temperatura média diária do ar no estado de Goiás. Planaltina, DF: **Embrapa Cerrados**, 2010, 27p.