ISSN: 2594-6153

Análise exploratória e ajuste de modelo linear geral aos dados históricos de temperatura e precipitação para a região agrícola ao sul do Parque Nacional da Chapada Diamantina, Bahia, Brasil

Rodrigo de Oliveira Campos[1]; Ana Maria Heuminski de Ávila[2]; Thiago Santos Silva[2]; Miguel Petrere Jr.[1,3]

[1]Universidade Santa Cecília (UNISANTA), Santos-SP, Brasil
[2]Universidade Estadual de Campinas (CEPAGRI/UNICAMP) - Campinas-SP, Brasil
[3]Universidade Federal do Pará (UFPA) - Belém-PA, Brasil
E-mail: rodrigocam@gmail.com

Resumo: As mudanças climáticas estão em evidência ao redor todo o mundo nos últimos anos. Variações nos padrões de chuva e temperatura influenciam diretamente nas práticas de produção agrícola. Os impactos podem depender do local, escala de produção e das práticas usuais adaptativas. Séries históricas de dados, referentes ao período entre 1961 e 2022, foram analisadas para quatro locais agrícolas distintos da região sul da Chapada Diamantina, Bahia. A precipitação média por localidade variou entre 1.97 mm/dia e 2.39 mm/dia (n=22645). A temperatura máxima diária observada por localidade no período foi entre 13.70°C e 37.49 °C (n=21762), enquanto que a mínima diária foi entre 4.23°C e 23.72°C (n=21762). A aplicação da análise de covariância aos dados de temperatura máxima média anual indicou que, ajustado o efeito do passar dos anos, houve diferenças significativas entre os locais investigados. A taxa de aumento da temperatura máxima com o passar dos anos (0.0261°C/ano) foi positiva e altamente significativa, indicando um aquecimento local.
Palavras-chave: precipitação; temperatura; agricultura; Chapada Diamantina, ANCOVA.

Exploratory data analysis and general linear model fit to historical temperature and precipitation data for the southern agricultural region of Chapada Diamantina National Park, Bahia, Brazil

Abstract: Climate change has been in evidence around the world in recent years. Variations in rainfall and temperature patterns directly influence agricultural production practices. Impacts may depend on location, scale of production and adaptive customary practices. Historical series of data, referring to the period between 1961 and 2022, were analyzed for four different agricultural sites in the southern region of Chapada Diamantina, Bahia. The average precipitation per locality varied between 1.97 mm/day and 2.39 mm/day (n=22645). The daily maximum temperature observed by location in the period was between 13.70°C and 37.49 °C (n=21762), while the daily minimum was between 4.23°C and 23.72°C (n=21762). The application of the analysis of covariance to the annual mean maximum temperature data indicated that, adjusted for the effect of passing years, there were significant differences between the investigated locations. The maximum temperature increase rate over the years (0.0261°C/year) was positive and highly significant, indicating local warming.
Keywords: precipitation; temperature; agriculture; Chapada Diamantina; ANCOVA.

Introdução

As alterações climáticas têm se apresentado cada vez mais evidentes ao redor de todo o mundo nos últimos anos. Em julho último, o planeta Terra apresentou a maior temperatura registrada deste o início da medição em 1880. Segundo o *Goddard Institute for Space Studies*

ISSN: 2594-6153

(GISS/NASA), a Terra em julho de 2023 estava 1,18°C mais quente que a média do mês mais quente que qualquer outro mês no registro de 143 anos [1,2].

Os impactos do clima futuro são hoje previstos, ou de certa forma esperados, em diversos setores econômicos, comprometedores do bem estar humano no planeta, dentre os quais a agricultura representa um dos mais preocupantes. A oferta de alimentos constitui um dos componentes de maior ameaça sob as mudanças climáticas pois as variações nos regimes de chuva e/ou temperatura sabidamente influenciam nas práticas de produção agrícola. Seus impactos podem depender da região, escala de produção, bem como das práticas corriqueiras no que tange aos processos adaptativos.

Objetivos

O presente trabalho visa analisar as tendências temporais quanto aos padrões de precipitação e temperatura para a região sul e sudoeste do entorno do Parque Nacional da Chapada Diamantina, Bahia, Brasil.

Material e Métodos

Foram acessadas as séries temporais de dados de (i) precipitação diária (mm/dia), (ii) temperatura máxima e (iii) mínima (°C) disponíveis para a região. Assim, dados provenientes de processos de interpolação, baseados em métodos de ponderação de distância inversa e distância angular [3], foram obtidos para as seguintes quatro localidades distintas aqui denominadas de: “Brejão” (Lat. -13.35° Long. -41.19° Alt. 853 m) e “Cascavel” (Lat. -13.32° Log. -41.38° Alt. 1141 m) no município de Ibicoara-BA; “Iramaia” (Lat. -13.35° Long. -41.0° Alt. 677) no município de mesmo nome; e “Piatã” (Lat. -13.23° Long. -41.7° Alt. 832 m) entre os municípios de Piatã e Abaíra-BA. Os dados disponíveis de precipitação são referentes aos anos de 1961 a 2022 (n=22645), enquanto que os dados de temperatura são de 1961 a 2020 (n=21762) [3].

Modelos de regressão linear simples e análise de covariância (ANCOVA) foram ajustados, em caráter exploratório, aos dados médios anuais de precipitação (mm/dia), temperatura máxima e mínima (°C) como opções de variável resposta, sendo o tempo, em anos, utilizado como covariável e as localidades do “Brejão”, “Cascavel”, “Iramaia” e “Piatã” como fatores. O seguinte modelo ANCOVA foi então ajustado:

$$Y_i = \mu + \varphi_i + \beta_1(X_1 - \overline{X_1}) + Interação + \varepsilon$$

Onde, Y_I = precipitação (mm/dia) ou temperatura (°C) no local i; μ = média populacional; φ_i = efeito do i-ésimo nível do fator “local”; β_I = coeficiente angular da covariável “anos”;

ISSN: 2594-6153

X_1 = valor da covariável "anos" no local *i*; $\overline{X_1}$ = média da covariável "anos" para todas as observações; *Interação* = refere-se à interação entre o fator "local" e a covariável "anos" (para testar a homogeneidade de inclinação das linhas de regressão); e ε = erro aleatório suposto $N(0,\sigma^2)$.

A análise dos resíduos ordinários foi executada através da aplicação do teste de normalidade de Shapiro-Wilk sobre a distribuição dos valores obtidos pelo modelo final. Além disso foram ainda calculados os resíduos studentizados para uma avaliação adicional validar os modelos. A homogeneidade de variância foi avaliada através do teste de Levene [4].

Resultados

Os valores de precipitação diária (mm/dia) observados na área de estudo nos últimos 62 anos (1961-2022) atingiram o máximo de 92.2 mm/dia. O intervalo interquartil foi entre 0 mm/dia e 2.1 mm/dia. A precipitação média por localidade foi de 1.97 mm/dia (n=22645, var=30.92) em Iramaia; 2.19 mm/dia (n=22645, var=33.02) no Brejão; 2.25 mm/dia (n=22645, var=30.61) em Cascavel; e 2.39 mm/dia (n=22645, var=30.97) em Piatã.

A variável temperatura máxima apresentou valores entre 13.70°C e 37.49 °C ao longo do período analisado (1961-2020). O intervalo interquartil foi entre 23.72°C e 30.55°C; sendo que os valores médios por localidade foram: 25.57°C (n=21762, var=7.53) em Cascavel; 27.95°C (n=21762, var=7.42) em Piatã; 28.34°C (n=21762, var=7.69) em Iramaia; e 28.58°C (n=21762, var=7.68) no Brejão. Quanto à temperatura mínima observada para os mesmos 60 anos, os valores ocorreram entre 4.23°C e 23.72°C. O intervalo interquartil foi entre 13.92°C e 19.72°C. Os valores de temperatura mínima média por localidade foram: 15.23°C (n=21762, var=4.17) em Cascavel; 17.81°C (n=21762, var=4.11) em Piatã; 18.06°C (n=21762, var=3.89) em Iramaia; e 18.21°C (n=21762, var=7.68) no Brejão.

A ANCOVA foi executada para os dados de precipitação, temperatura máxima e mínima. A ANCOVA aplicada aos dados de precipitação média anual ao longo de 62 anos se mostrou altamente significativa, porem com um R^2 ajustado = 7.8% com pouca utilidade prática, e assim foi descartada.

A ANCOVA aplicada aos dados de temperatura máxima média anual (Figura 1) apontou que houve diferenças significativas entre os locais investigados (R^2 ajustado=74.2%; *F-statistic*: 172.7; 4 e 235 G.L; p-valor: < 2.2e-16) (Tabela 1). As médias ajustadas obtidas por localidade, e respectivos intervalos de confiança, foram: 25.6°C (25.4-25.8) em Cascavel; 27.9°C (27.8-28.1) em Piatã; 28.3°C (28.1-28.5) em Iramaia; e 28.6°C (28.4-28.8) no Brejão. A distribuição dos resíduos ordinários não apresentou normalidade com base no teste de

ISSN: 2594-6153

Shapiro-Wilk (W = 0.98684, p-*value* = 0.02636). Uma análise posterior sobre os resíduos studentizados permitiu detectar a ocorrência de dois outliers (Figura 2). Análises posteriores serão buscadas para remediar esse efeito.

A ANCOVA aplicada aos dados de temperatura mínima média anual apresentou não paralelismos entre as linhas de regressão; apontando a interação significativa entre a covariável "anos" e o fator "local Piatã" (p=0.0359), R^2 ajustado = 86.9%. Os gráficos dessa análise não foram aqui apresentados por limitação de espaço.

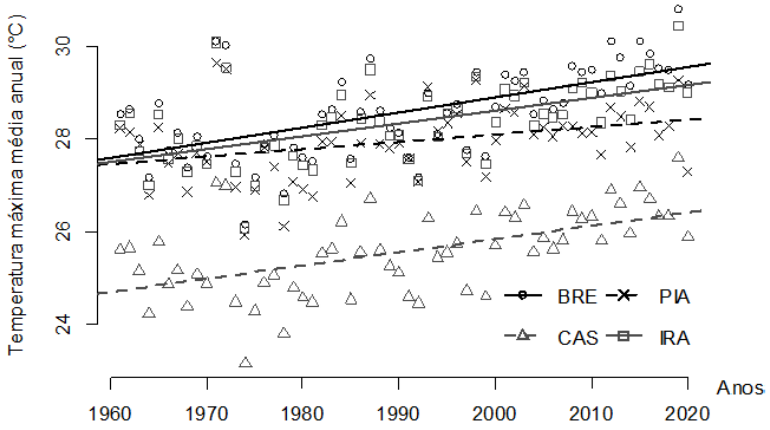


Figura 1: Diagrama de dispersão entre os valores médios anuais de temperatura máxima (°C) e os anos (1961-2020) por local da região sul do entorno do Parque Nacional da Chapada Diamantina (BRE=Brejão; CAS=Cascavel; PIA=Piatã; IRA=Iramaia).

Tabela 1: ANCOVA para o ajuste do modelo dos dados médios anuais de temperatura máxima (°C) por local (Brejão; Cascavel; Piatã; e Iramaia) com os anos (1961-2020) como covariável. R^2 ajustado=74.2%.

Fonte de variação	G.L.	SQ	QM	F	P
Anos	1	49.16	49.16	86.45	< 2e-16 ***
Local	3	343.5	114.51	201.39	< 2e-16 ***
Resíduo	235	133.6	0.57		

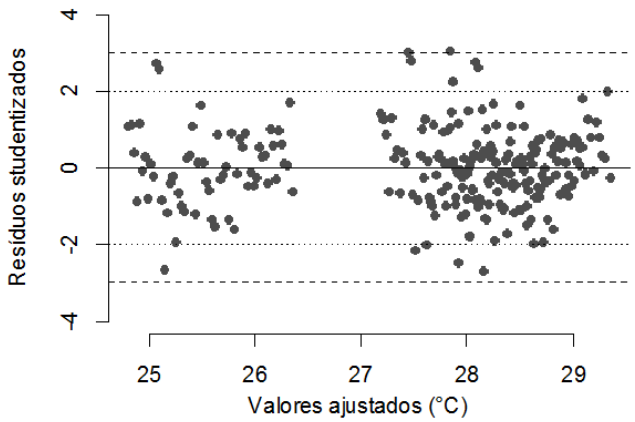


Figura 2: Diagrama de dispersão entre os resíduos studentizados e o valores ajustados do modelo de análise de covariância para os dados médios anuais de temperatura máxima (°C).

Discussão

ISSN: 2594-6153

A análise exploratória dos dados de precipitação, temperatura máxima e mínima levou à apresentação de resumos estatísticos e diagramas indispensáveis para o adequado direcionamento de abordagens inferenciais (e.g. modelos lineares gerais). Além disso, em se tratando de dados das ciências atmosféricas, nas quais somos frequentemente confrontados com amostras extremamente grandes [5], o benefício da aplicação da análise exploratória de dados é ainda mais evidente com a consequente redução destes volumes através do cálculo das médias anuais, reduzindo o ruído ambiental. A busca por contrastes e/ou padrões de distribuição dos dados nos permite melhor desenvolver hipóteses plausíveis sobre os fenômenos de interesse. Neste sentido, a aplicação da ANCOVA é de grande valia. A linearidade observada na relação entre a precipitação (y) e os anos (x), embora altamente significativa, tem uma reduzida explicabilidade devido ao baixo valor do R^2 ajustado = 7.8%. A aplicação de outra abordagem inferencial deve então ser considerada para esta variável, uma vez que não houve paralelismo entre as linhas de regressão dos locais investigados, suposição exigida para o correto emprego da ANCOVA [4,6].

Os resultados da aplicação da ANCOVA para as variáveis respostas temperatura máxima e mínima, separadamente, indicaram o uso oportuno desta análise inferencial. Para o modelo ajustado aos dados de temperatura máxima, o paralelismo entre as linhas de regressão possibilitou obtermos um modelo final onde o R^2 ajustado = 74%. Assim, a menor temperatura máxima média anual observada no período de 60 anos foi apresentada para Cascavel (25.6°C) e a maior para o Brejão (28°C). Embora os resíduos ordinários não tenham apresentado distribuição normal, a utilidade da ANCOVA para melhor compreensão dos padrões pode ser considerada sem grandes restrições. Outro aspecto importante acerca da distribuição dos resíduos (Figura 2), é o seu caráter disjunto, resultando em dois “aglomerados” distintos. Isto sugere que há indícios de que um “fator desconhecido”, não foi incluído no modelo para a explicação das variações de temperatura entre os locais. Neste sentido, considerando as características físicas de cada local, é plausível considerarmos, por exemplo, a inserção de dados de altitude local em modelos futuros [7,8]. O presente estudo considerou três locais abaixo de 1000 m (BRE, IRA, PIA) e um único local acima (CAS). Para melhor investigar essa questão, novos ajustes serão explorados ao se considerar um maior número de locais com diferentes altitudes.

Quanto à ANCOVA com os dados de temperatura mínima como variável resposta (não apresentada), a interrupção do ajuste causada pela interação significativa entre a covariável

ISSN: 2594-6153

"anos" e o fator "local" (i.e., Piatã), não diminui o valor deste modelo. Seu resultado indicou que a linha de regressão para o local Piatã foi diferente das demais localidades, sugerindo, através do coeficiente de inclinação (i.e., β_1), que o incremento na temperatura mínima foi maior para este local em relação aos demais com o passar dos anos entre 1961 e 2020. Assim, em casos de heterogeneidade de linhas de regressão, uma alternativa viável de continuidade de investigação é recomendada por Huitema [5], aplicando a técnica original de Johnson-Neyman [9].

Conclusões

Foi possível verificar os efeitos do fator "local" sobre as variações de precipitação, temperatura máxima e mínima, controlando os efeitos do passar dos "anos". A taxa de aumento da temperatura máxima com o passar dos anos foi positiva e altamente significativa, indicando um aquecimento local de 0.0261°C/ano. A aplicação de modelo de séries temporais de dados e modelos lineares generalizados representam alternativas adequadas para a continuidade das análises mais robustas sobre contrastes e padrões existentes nos dados de temperatura e precipitação para a região sul do Parque Nacional da Chapada Diamantina nas ultimas seis décadas.

Agradecimentos: Os autores agradecem o apoio fornecido pela UNISANTA, pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e pela bolsa outorgada ao primeiro autor.

Referências

1. GISTEMP Team, 2023: GISS Surface Temperature Analysis (GISTEMP), version 4. NASA Goddard Institute for Space Studies. Dataset accessed 2023-08-20 at https://data.giss.nasa.gov/gistemp/.
2. Lenssen, N., G. Schmidt, J. Hansen, M. Menne, A. Persin, R. Ruedy, and D. Zyss, 2019: Improvements in the GISTEMP uncertainty model. J. Geophys. Res. Atmos., 124, no. 12, 6307-6326, doi:10.1029/2018JD029522.
3. Xavier, A. C., Scanlon, B. R., King, C. W., & Alves, A. I. 2022. New improved Brazilian daily weather gridded data (1961–2020). International Journal of Climatology, 42 (16), 8390– 8404. https://doi.org/10.1002/joc.7731.
4. Quinn, G.; Keough, M. 2002. Experimental Design and Data Analysis for Biologist. 10.1016/S0022-0981(02)00278-2.
5. Wilks, D.S. 2011. Statistical Methods in the Atmospheric Sciences, 3rd ed. Academic Press: Oxford, UK and Waltham, MA.
6. Huitema, B. E. 1980. The Analysis of Covariance and Its Alternatives. Wiley, New York.
7. Lancaster, I.N. 1980. Relationships Between Altitude and Temperature in Malawi, South African Geographical Journal, 62:1, 89-97, DOI: 10.1080/03736245.1980.10559624.

ISSN: 2594-6153

8. Dong, D., Huang, G., Qu, X., Tao, W., & Fan, G. (2014). Temperature trend–altitude relationship in China during 1963–2012. Theoretical and Applied Climatology, 122(1-2), 285–294. doi:10.1007/s00704-014-1286-9.
9. Wilcox, R.R. 1987. Pairwise comparisons of J independent regression lines over a finite interval, simultaneous pairwise comparisons of their parameter, and the Johnson–Neyman procedure. British Journal of Mathematical and Statistical Psychology 40: 80–93.