

**Pontifícia Universidade Católica de Goiás**  
**Programa de Mestrado em Engenharia de Produção e Sistemas**

**INFLUÊNCIA DAS MUDANÇAS CLIMÁTICAS NA  
PRODUTIVIDADE DE GRÃOS**

**LEONARDO RODRIGUES DE OLIVEIRA MERELLES**

2018

# **INFLUÊNCIA DAS MUDANÇAS CLIMÁTICAS NA PRODUTIVIDADE DE GRÃOS**

**LEONARDO RODRIGUES DE OLIVEIRA MERELLES**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Pontifícia Universidade Católica de Goiás, como parte dos requisitos para obtenção do título de Mestre em Engenharia de Produção e Sistemas.

Orientador: Prof. José Elmo de Menezes, *Dr.*

Goiânia,

Março de 2018

M559i Merelles, Leonardo Rodrigues de Oliveira  
Influência das mudanças climáticas na produtividade de grãos  
[ recurso eletrônico] Leonardo Rodrigues de Oliveira Merelles.-  
- 2018.  
97 f.; il.

Texto em português com resumo em inglês  
Dissertação (mestrado) - Pontifícia Universidade Católica  
de Goiás, Programa de Pós-Graduação Stricto Sensu em  
Engenharia de Produção e Sistemas , Goiânia, 2018

Inclui referências f. 71-81

1. Mudanças climáticas. 2. Milho - Produção. 3. Soja - Produção.  
4. Análise de séries temporais. I.Menezes, José Elmo de. II.Pontifícia  
Universidade Católica de Goiás. III. Título.

CDU: Ed. 2007 -- 551.583(043)

# INFLUÊNCIA DAS MUDANÇAS CLIMÁTICAS NA PRODUTIVIDADE DE GRÃOS

**Leonardo Rodrigues de Oliveira Merelles**

Esta Dissertação foi julgada adequada para obtenção do título de Mestre em Engenharia de Produção e Sistemas, e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Pontifícia Universidade Católica de Goiás em Março de 2018.



---

Prof. Marcos Lajovic Carneiro, *Dr.*  
Coordenador do Programa de Pós-Graduação  
em Engenharia de Produção e Sistemas

Banca Examinadora:



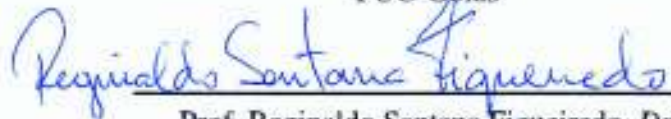
---

Prof. José Elmo de Menezes, *Dr.*  
Orientador



---

Prof. Clarimar José Coelho, *Dr.*  
PUC-Goiás



---

Prof. Reginaldo Santana Figueiredo, *Dr.*  
UFG

Goiânia-Goiás  
Março de 2018

Dedico este trabalho a meus pais, fonte de meus conhecimentos e saber. Graças a eles, tornei-me uma pessoa capaz de lutar para que meus sonhos e objetivos fossem sempre alcançados, sem jamais desanimar. Considero-me forte porque eles me ensinaram a ser forte.

## AGRADECIMENTOS

Agradeço a Deus por todas as oportunidades. Por me abençoar e conduzir por caminhos que superaram minhas dificuldades. Também, por cada pessoa escolhida por Ele para conviver comigo.

A meu pai Eli Silva Merelles e minha mãe Idelma Rodrigues de Oliveira Merelles pelo apoio, orientação e todo carinho.

A meu orientador, professor Dr. José Elmo de Menezes, pela dedicação, confiança, paciência e aprendizados compartilhados.

A FAPEG pela concessão da bolsa de estudos que fomentou minha pesquisa.

Aos professores Clarimar José Coelho, Marajá João Alves de Mendonça Filho, Maria José Pereira Dantas, Marta Pereira da Luz, Ricardo Luiz Machado e Solange da Silva agradeço pelas colaborações, contribuições, motivações e exemplo.

A minha família e amigos pelo carinho, amor incondicional e compreensão de minha ausência.

Aos colegas de mestrado pela parceria, motivações, manifestações, aprendizados e desabafos.

E, claro, a Viviane de Souza Dias por todo brilhantismo, alegria e todo companheirismo acrescentado em minha vida.

**Obrigado a todos!**

“De tanto ver triunfar as nulidades; de tanto ver prosperar a desonra; de tanto ver crescer a injustiça; de tanto ver agigantar-se os poderes nas mãos dos maus, o homem chega a desanimar-se da virtude, a rir-se da honra, e a ter vergonha de ser honesto”

*(Rui Barbosa)*

“If I have seen further, it is by standing on the shoulders of giants”

*(Isaac Newton)*

Resumo da Dissertação apresentada ao MEPROS/ PUC Goiás como parte dos requisitos necessários para a obtenção do grau de Mestre em Engenharia de Produção e Sistemas (M.Sc.)

## INFLUÊNCIA DAS MUDANÇAS CLIMÁTICAS NA PRODUTIVIDADE DE GRÃOS

Leonardo Rodrigues de Oliveira Merelles

Março / 2018

Orientador: Prof. José Elmo de Menezes, *Dr.*

### **RESUMO:**

Compreender como as mudanças climáticas influenciam na produtividade agrícola contribui com a previsibilidade de suas consequências e auxilia na gestão do agronegócio e da segurança alimentar. Neste estudo, foi utilizada a regressão por mínimos quadrados parciais (PLSR) para quantificar as contribuições das mudanças climáticas na produtividade da soja e do milho. Para verificar a qualidade de ajuste do modelo foi aplicada a regressão de séries temporais. As regiões Centro-Oeste e Sul do Brasil foram consideradas para a condução da modelagem da produtividade de grãos, utilizando como variáveis preditoras a precipitação, temperatura, umidade, evapotranspiração e os níveis de dióxido de carbono, entre os anos de 1980 a 2016. Nas duas regiões, o principal contribuinte para o incremento da produtividade de grãos foi o aumento nos níveis de dióxido de carbono e o principal contribuinte para a redução da produtividade foi a umidade do ar, sugerindo que as mudanças climáticas estão influenciando na produtividade de grãos. Os resultados confirmam as mudanças climáticas, indicando que agricultores e tomadores de decisões deverão atuar em outras variáveis que controlam a produtividade, como melhoramento genético e práticas de gestão agrícola.

**Palavras-chave:** Produtividade de grãos. Mínimos quadrados parciais. Séries temporais. Dados faltantes.



Resumo da Dissertação apresentada ao MEPROS/ PUC Goiás como parte dos requisitos necessários para a obtenção do grau de Mestre em Engenharia de Produção e Sistemas (M.Sc.)

Influence of climate change on grain yield

Leonardo Rodrigues de Oliveira Merelles

Março / 2018

Orientador: Prof. José Elmo de Menezes, *Dr.*

**RESUMO:**

Understanding how climate change influences crop yield contributes to the forecasting of its consequences and assists in the management of agribusiness and food security. In this study, partial least squares regression (PLSR) was used to quantify the contribution of climate change to soybean and corn yield. To verify the accuracy of the model, the time series regression was applied. The Central-West and South regions of Brazil were considered for the modeling of grain yield using precipitation, temperature, humidity, evapotranspiration and carbon dioxide levels between 1980 and 2016 as predictors. In both regions, the main contributor to increased grain yield was the increase in carbon dioxide levels, and the main contributor to the reduction of productivity was air humidity, suggesting that climate change is influencing grain yield. The results confirm climate change, indicating that farmers and decision makers should work on other variables that control productivity, such as genetic improvement and agricultural management practices.

**Key-words:** Grains yield. Partial least squares. Time series. Missing datas.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Classificação temática dos artigos preliminares incluídos na revisão . . . .	22
Figura 2 – Descrição das publicações sobre produtividade de grãos e mudanças climáticas . . . . .	23
Figura 3 – Mapa conceitual da produtividade de grãos no tempo e escala global . . .	24
Figura 4 – Diagrama de Pareto da produtividade de grãos no Brasil em 2016 . . . . .	31
Figura 5 – Agrupamento das estações meteorológicas para a variável temperatura no estado de Goiás . . . . .	33
Figura 6 – <i>Boxplot</i> com informações do clima de 1980 a 2016. (a) variáveis originais da região Centro-Oeste. (b) variáveis originais da região Sul. (c) variáveis transformadas da região Centro-Oeste. (d) variáveis transformadas da região Sul . . . . .	34
Figura 7 – Estações meteorológicas utilizadas no Brasil. (a) Localização dos dados do INMET. (b) Localização dos dados do NOAA . . . . .	35
Figura 8 – Número de vizinhos com as distâncias equivalentes . . . . .	52
Figura 9 – Série histórica da produtividade de grãos de 1980 a 2016. (a) produtividade da região Centro-Oeste. (b) produtividade da região Sul . . . . .	58
Figura 10 – Tendências climáticas na região Centro-Oeste do Brasil de 1980 a 2016 . .	61
Figura 11 – Tendências climáticas na região Sul do Brasil de 1980 a 2016 . . . . .	62
Figura 12 – Série histórica real da produtividade de grãos de 1980 a 2016 e a produtividade prevista por PLSR e séries temporais. (a) produtividade da região Centro-Oeste. (b) produtividade da região Sul . . . . .	69

## LISTA DE TABELAS

Tabela 1 – Análise do número de estações meteorológicas com dados faltantes das variáveis climáticas . . . . .	36
Tabela 2 – Métodos para estimar a evapotranspiração de referência . . . . .	39
Tabela 3 – Distância máxima aceitável pelos vizinhos . . . . .	53
Tabela 4 – Estatística RMSE da análise dos métodos para completar dados climáticos com 10% de dados . . . . .	53
Tabela 5 – Melhores métodos para completar os dados faltantes . . . . .	54
Tabela 6 – Análise do número de estações meteorológicas com dados faltantes após as séries serem completadas com informações dos vizinhos mais próximos	55
Tabela 7 – Métricas para estimar o melhor método de evapotranspiração de referência	57
Tabela 8 – Medidas descritivas da produtividade de Soja e Milho na região Centro-Oeste e Sul do Brasil . . . . .	57
Tabela 9 – Correlação entre as variáveis preditoras e resposta . . . . .	59
Tabela 10 – Teste sequência de Dickey e Fuller para produtividade de grãos na região Centro-Oeste e Sul do Brasil . . . . .	63
Tabela 11 – Parâmetros ajustados pelo modelo ARIMA para produtividade de grãos .	64
Tabela 12 – P-valor dos testes de normalidade e autocorrelação para os resíduos da região Centro-Oeste e Sul . . . . .	64
Tabela 13 – Resumo da modelagem da produtividade de grãos na região Centro-Oeste e Sul pela regressão PLS . . . . .	65
Tabela 14 – Valores da influência da variável na projeção ( <b>VIP</b> ) e regressão por mínimos quadrados parciais (PLSR) para região Centro-Oeste e Sul do Brasil .	66
Tabela 15 – Lista dos coeficientes da regressão para produtividade de grãos <sup>a</sup> com a regressão PLS para região Centro-Oeste e Sul . . . . .	67
Tabela 16 – Critério para examinar acurácia dos modelos na região Centro-Oeste e Sul	69
Tabela 17 – Valores de p-valor do teste de Tukey para a diferença entre o real e o previsto nos pontos extremos e a frequência dos dados climáticos . . . . .	83
Tabela 18 – Teste de aderência para os dados climáticos por estação do ano para região Centro-Oeste . . . . .	85

Tabela 19 – Teste de aderência para os dados climáticos por estação do ano para região

Sul . . . . . 86

## LISTA DE ABREVIATURAS E SIGLAS

AIC	Critério de Informação de Akaike
ANOVA	Análise de Variância
BIC	Critério de Informação de Bayesian
CV	<i>Cross-Validation</i>
CO <sub>2</sub>	Dióxido de Carbono
CONAB	Companhia Nacional de Abastecimento
EMs	Estações Meteorológicas
ET	Evapotranspiração
ET <sub>o</sub>	Evapotranspiração de Referência
FAO	<i>Food and Agriculture organization</i>
INMET	Instituto Nacional de Meteorologia
IPCC	Painel Internacional de Mudanças Climáticas
MAE	<i>Mean Absolute Error</i>
MAPE	<i>Mean Absolute Percentage Error</i>
MT	Milhões de toneladas
NIPALS	<i>Non-linear Iterative Partial Least Squares</i>
NOAA	<i>National Oceanic and Atmospheric Administration</i>
NSE	<i>Nash-Sutcliffe efficiency</i>
PBIAS	<i>Percent Bias</i>
PLS	<i>Partial Least Squares</i>
PLSC	<i>Partial Least Squares Correlation</i>
PLS-PM	<i>Partial Least Squares Path Modeling</i>
PLSR	<i>Partial Least Squares Regression</i>
P <sub>ad</sub>	Precipitação Acumulada
PRESS	<i>Prediction Error Sum of Squares</i>
RMSE	<i>Root Mean Squared Error</i>
RSS	<i>Residual Sum of Squares</i>
T <sub>cal</sub>	Temperatura Média Calculada
T <sub>max</sub>	Temperatura Máxima
T <sub>med</sub>	Temperatura Média
T <sub>min</sub>	Temperatura Mínima

$T_{\text{dtr}}$	Intervalo diurno de Temperatura
$UR_{\text{max}}$	Umidade Máxima
$UR_{\text{med}}$	Umidade Média
$UR_{\text{min}}$	Umidade Mínima
$UR_{\text{dtr}}$	Intervalo diurno da umidade relativa do ar
$U_z$	Velocidade dos Ventos com altura z
VIP	<i>Variable Influence on Projection</i>

## LISTA DE SÍMBOLOS

$\mathbf{A}$	é uma matriz
$a_{ij}$	é uma matriz genérica $\mathbf{A}$ , de $i$ -ésima linha e $j$ -ésima coluna
$\mathbf{A}'$	é uma matriz transposta de $\mathbf{A}$
$\mathbf{A}^d$	é uma matriz quadrada diagonal
$\mathbf{A}^{ts}$	é uma matriz quadrada triangular superior
$\mathbf{a}^c$	é um vetor de colunas (tamanho $i \times 1$ )
$\mathbf{a}^l$	é um vetor de linhas (tamanho $1 \times j$ )
$\ \mathbf{a}^c\ $	é a normalização de um vetor para um
$\langle \mathbf{a}^c, \mathbf{a}^c \rangle$	é o produto escalar entre dois vetores (onde o primeiro vetor é transposto)
$\hat{a}$	é o valor preditivo
$\bar{a}$	é o valor da média
$n$	é o número de amostras no conjunto de dados
$m$	é o número de variáveis ( $\mathbf{y}^c$ ) dependentes
$k$	é o número de variáveis ( $\mathbf{x}^c$ ) independentes
$h$	é um elemento genérico para contar as componentes
$H$	número de componentes totais no modelo
$\mathbf{E}$	é a matriz de deflação de $\mathbf{X}$ (resíduos ou erros de ordem $n \times k$ )
$\mathbf{F}$	é a matriz de deflação de $\mathbf{Y}$ (resíduos ou erros de ordem $n \times k$ )
$\mathbf{Z}$	é a matriz com dados do clima (ordem $n \times k$ )
$\mathbf{Z}^*$	é a matriz $\mathbf{Z}$ com dados centrados e escalados para variância unitária
$\mathbf{Y}$	é a matriz com dados da produtividade (ordem $n \times m$ )
$\mathbf{Y}^*$	é a matriz $\mathbf{Y}$ com dados centrados e escalados para variância unitária
$\hat{\mathbf{Y}}$	é a matriz preditiva de $\mathbf{Y}$
$\mathbf{X}^k$	são $k$ matrizes com os dados brutos das estações meteorológicas (ordem $t \times v$ )
$\mathbf{S}^k$	são $k$ matrizes com as sazonalidades das séries (ordem $t \times v$ )
$\mathbf{TD}^k$	são $k$ matrizes com as tendências das series (ordem $t \times v$ )
$\mathbf{A}^k$	são $k$ matrizes com os ruídos aleatórios das series (ordem $t \times v$ )
$\mathbf{B}$	é a matriz de coeficiente que modela $\mathbf{Y}^*$ (ordem $k \times m$ )

$\mathbf{C}'$	é a matriz de carga ( <i>loads</i> ) de $\mathbf{Y}$ (ordem $h \times m$ )
$\mathbf{P}'$	é a matriz de carga ( <i>loads</i> ) de $\mathbf{X}$ (ordem $h \times k$ )
$\mathbf{U}$	é a matriz de pontuação ( <i>scores</i> ) de $\mathbf{Y}$ (ordem $n \times h$ )
$\mathbf{T}$	é a matriz de pontuação ( <i>scores</i> ) de $\mathbf{X}$ (ordem $n \times h$ )
$\mathbf{W}'$	é a matriz de pesos ( <i>weights</i> ) de $\mathbf{X}$ (ordem $h \times k$ )
$\mathbf{R}'$	é uma matriz com os pesos $\mathbf{W}'$ atualizados (ordem $h \times k$ )
$\mathbf{I}_n$	é uma matriz identidade com uns (ordem $n \times n$ )
$\mathbf{c}_h^l$	é um vetor de linhas com cargas de $\mathbf{Y}$ , para componente $h$ (tamanho $1 \times m$ )
$\mathbf{p}_h^l$	é um vetor de linhas com cargas de $\mathbf{X}$ , para componente $h$ (tamanho $1 \times k$ )
$\mathbf{u}_h^c$	é um vetor de colunas com pontuações de $\mathbf{Y}$ , para componente $h$ (tamanho $n \times 1$ )
$\mathbf{t}_h^c$	é um vetor de colunas com pontuações de $\mathbf{X}$ , para componente $h$ (tamanho $n \times 1$ )
$\mathbf{w}_h^l$	é um vetor de linhas com pesos de $\mathbf{X}$ , para componente $h$ (tamanho $1 \times k$ )
$\alpha^l$	é um vetor de intercepto da regressão PLS (ordem $1 \times m$ )
$Q^2$	é a explicação da validação cruzada
$R^2$	é o coeficiente de determinação
$t$	é o tempo
$v$	é o número de estações meteorológicas
$r$	são as macrorregiões
$\epsilon$	é um valor de erro
$\theta$	é um parâmetro de média-móvel
$\phi$	é um parâmetro auto-regressivo



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>17</b>
<b>2</b>	<b>MUDANÇAS CLIMÁTICAS E PRODUTIVIDADE DE GRÃOS</b>	<b>20</b>
2.1	Mudanças climáticas	20
2.2	Método sistemático	21
2.3	Mapa conceitual da produtividade de grãos	23
2.4	Produtividade de grãos	25
<b>3</b>	<b>METODOLOGIA</b>	<b>29</b>
3.1	Caracterização da pesquisa	29
3.2	Área de estudos e base de dados	30
3.3	Variáveis climáticas	33
3.3.1	Método de preenchimento com informações de estações vizinhas	35
3.3.2	Qualidade dos dados	37
3.4	Cálculo da evapotranspiração de referência	37
3.5	Ajuste da produtividade de grãos por séries temporais	41
3.6	Ajuste da produtividade de grãos pela regressão PLS	44
3.7	Métricas de avaliação dos ajustes	49
<b>4</b>	<b>RESULTADOS E DISCUSSÕES</b>	<b>52</b>
4.1	Dados climáticos	52
4.2	Evapotranspiração	56
4.3	Ajuste de produtividade de grãos	57
4.3.1	Modelagem por séries temporais	63
4.3.2	Modelagem pela regressão PLSR	64
4.3.3	Análise comparativa e previsão	68
<b>5</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>70</b>
5.1	Métodos para completar os dados climáticos	70
5.2	Predição da produtividade de grãos	70
5.3	Trabalho futuro	71

---

<b>REFERÊNCIAS . . . . .</b>	<b>72</b>
<b>APÊNDICES</b>	<b>82</b>
<b>APÊNDICE A – TESTE DO MELHOR MÉTODO PARA COMPLE- TAR OS DADOS . . . . .</b>	<b>83</b>
<b>APÊNDICE B – TESTE DE ADERÊNCIA . . . . .</b>	<b>85</b>
<b>APÊNDICE C – ARTIGOS PUBLICADOS . . . . .</b>	<b>87</b>
<b>APÊNDICE D – <i>SCRIPT</i> EM R . . . . .</b>	<b>88</b>

## 1 INTRODUÇÃO

As mudanças climáticas tem sido confirmadas (BAJAT *et al.*, 2015; IPCC, 2014; RAYNER *et al.*, 2003) e podem estar relacionadas a processos naturais do planeta, também a mudanças antropogênicas persistentes na composição da atmosfera ou no uso do solo (IPCC, 2001). As mudanças climáticas estão influenciando na evapotranspiração (WANG *et al.*, 2015a), eficiência do uso hídrico (LIU *et al.*, 2016) e outras variáveis climáticas (TORGOVITSKI, 2015). Essas alterações podem influenciar o fluxo dos rios, o gerenciamento de usinas hidrelétricas (AHMADI *et al.*, 2015; LÓPEZ-MORENO *et al.*, 2014), a produtividade de grãos e em casos onde há desastres ambientais, podem adicionar custo as seguradoras, alterar a logística e afetar a economia local (YANG; XU, 2015).

Na agricultura, as mudanças climáticas e a variabilidade climática afetam significativamente o crescimento e o desenvolvimento de culturas (DING *et al.*, 2016; JUNK *et al.*, 2016), alterando os padrões de cultivo (KOS *et al.*, 2017) e as regiões de plantio (ASSENG *et al.*, 2013; ASSENG; PANNELL, 2013). Também, causaram reduções consideráveis nos rendimentos de várias culturas (VERÓN *et al.*, 2015) e aumentaram as incidências de doenças (OLIVEIRA *et al.*, 2017) devido às alterações na temperatura, precipitação e umidade (MÜLLER; KORN, 2013).

Nos últimos 50 anos a modelagem de cultura vem se desenvolvendo com auxílio de cientistas, economistas e engenheiros (JONES *et al.*, 2017a), sendo classificado em três grupos. Isso ocorreu devido (JONES *et al.*, 2017b) ao uso de: (1) abordagens analíticas, onde podem ser observadas em sistemas de intensificação de culturas (BERKHOUT *et al.*, 2015), aplicação da técnica Mini-FACE (é utilização de uma câmara que permite saturar o CO<sub>2</sub>) (HAO *et al.*, 2014) e outras; (2) ferramentas de tecnologia da informação, onde são aplicadas simulações para determinar a produção de culturas e medir o impacto das mudanças climáticas (WANG *et al.*, 2015a; CALZADILLA *et al.*, 2013); e (3) modelagem de sistemas, onde podem ser observados estudos de co-integração (VERÓN *et al.*, 2015), modelos agro-meteorológicos (FERREIRA, 2010) e séries temporais.

Dentro dessa classificação poderia ser incluída a regressão por mínimos quadrados parciais (do inglês partial least square regression – PLSR) na modelagem de sistemas. O PLSR é uma ferramenta robusta multivariada (WEHRENS; MEVIK, 2007; WOLD *et al.*, 2001) e vem sendo aplicada para fazer previsão de chuva no verão (estação mais desafi-

adora) (XING *et al.*, 2016), previsões de demanda para o transporte de energia elétrica (ZHANG *et al.*, 2009) e análise de resposta hidrológica para mudanças no uso e cobertura do solo (WOLDESENBET *et al.*, 2017). A aplicação dessa modelagem na produtividade de grãos ainda é limitada. Carkner e Entz (2017) realizaram a modelagem da produtividade dos grãos de soja utilizando o nível de fertilização do solo, a biomassa de ervas daninhas e a biomassa de cultivares. Entretanto, os valores de precipitação e temperatura foram utilizados apenas de forma descritiva. Assim, como as variáveis climáticas estariam influenciando na produtividade de grãos nas principais regiões produtoras no Brasil?

Devido ao desempenho apresentado pelo PLSR na literatura, o objetivo geral deste estudo é medir/quantificar os impactos das mudanças climáticas sobre a produtividade de grãos, na região Centro-Oeste e Sul do Brasil, usando a regressão de mínimos quadrados parciais.

Entretanto, estudos dessa natureza são complexos devido as falhas nos dados históricos em estações meteorológicas (BIER; FERRAZ, 2017). Então, para obter informações de dados faltantes podem ser aplicados vários métodos (EISCHEID *et al.*, 1995), como: interpolação multivariada (AKIMA, 1978); regressão linear sem o intercepto (ALLEN; DEGAETANO, 2001); distância ponderada inversa; e substituição pelo dado mais próximo (FERRARI; OZAKI, 2014; VICENTE-SERRANO *et al.*, 2010).

Assim, os objetivos específicos foram: (1) obter os dados históricos das variáveis climáticas no Brasil de 1980 a 2016; (2) determinar o melhor método para completar as variáveis (temperatura, precipitação e umidade relativa) na ausência de dados, com informações dos vizinhos mais próximos; (3) determinar o melhor método para calcular a evapotranspiração de referência (considerando os mais aplicados na literatura) quando existem falhas nas séries; e (4) ajustar a produtividade de grãos por séries temporais e PLSR.

As contribuições deste trabalho se encontram em (1) gerar uma simples abordagem para determinar como completar dados faltantes para as principais variáveis climáticas e o melhor método para calcular a evapotranspiração, que podem ser aplicados em estudos de diferentes áreas, como a geografia, geração de energia em hidrelétricas e na agricultura. (2) Também é proposto como inovação a modelagem e avaliação dos impactos das mudanças climáticas na produtividade de grãos, utilizando o PLSR.

O mapa conceitual da produtividade de grãos e suas áreas de pesquisas são apresentados no capítulo 2 (referencial teórico). O levantamento dos dados climáticos e o ajuste dos modelos à produtividade de grãos estão no capítulo 3 (metodologia). Os resultados e discussões do levantamento dos dados e ajuste da produtividade de grãos são apresentados no capítulo 4. Por último são apresentadas as considerações finais.

## 2 MUDANÇAS CLIMÁTICAS E PRODUTIVIDADE DE GRÃOS

Este capítulo expõe o método de revisão sistemática com os critérios adotados para a escolha dos artigos e o mapa conceitual da produtividade de grãos com o fluxo de pesquisas. Também, são apresentadas as mudanças climáticas.

### 2.1 Mudanças climáticas

Mudanças climáticas vem sendo observadas desde o século XIX (RAYNER *et al.*, 2003). A partir de 1880 a 2012 as mudanças confirmadas foram de  $0,85 \pm 0,21^{\circ}\text{C}$  na superfície do planeta. Se considerado o cenário mais pessimista (CRP8.5) do Painel Internacional de Mudanças Climáticas (IPCC), as mudanças climáticas podem chegar a  $3,7 \pm 0,55^{\circ}\text{C}$  até o ano de 2100 (IPCC, 2014). Os efeitos do aquecimento sistemático podem ser observados através da alteração nos níveis dos mares, derretimento das calotas polares, aquecimento dos oceanos e da atmosfera.

O IPCC definiu mudanças climáticas como “uma variação estatisticamente significativa tanto no estado médio do clima como na sua variabilidade, persistindo por um período prolongado, tipicamente décadas ou mais” (IPCC, 2001, p. 788). Nos últimos 20 anos, a existência, a causa e a gravidade das mudanças climáticas tem sido amplamente debatida (EVANGELISTA *et al.*, 2013).

As mudanças climáticas não são iguais em todas as regiões do globo, por exemplo, estão ocorrendo mais alterações na região dos trópicos (MCAFEE, 2013) e em algumas regiões apresentam tendências de esfriar (BAJAT *et al.*, 2015). Na bacia hidrográfica de Murray Darling (Austrália) a temperatura média entre 1900-2007 oscilou predominando o aquecimento e na precipitação, nos últimos anos, teve uma forte redução (KAMRUZZAMAN *et al.*, 2011). Para Mekasha *et al.* (2014) não está ocorrendo tendência significativa na temperatura mínima, porém, na temperatura máxima sim, como tendências de esquentar e esfriar. No mesmo estudo, em 60% das Estações Meteorológicas (EMs), a precipitação está aumentando, e nas demais está reduzindo.

Na Alemanha a primeira tendência climática foi confirmada em 1951 com duração de 25 anos e elevação de  $1,5^{\circ}\text{C}$  na temperatura (TORGOVITSKI, 2015). Também, observa-se que as mudanças climáticas alteraram o comprimento das estações do inverno e verão

em -9 e 4 dias, nessa ordem (KUTTA; HUBBART, 2016). No Brasil estão ocorrendo tendências de esquentar e esfriar, essas tendências são diferentes por regiões e altitude. Para mais detalhes veja o Apêndice ??.

A temperatura e a precipitação influenciaram na produtividade do milho e do trigo, reduzindo-a em escala global, entretanto, mudanças na produtividade de arroz e soja não foram significativas (LOBELL *et al.*, 2011). Na Argentina, as mudanças climáticas têm reduzido a produtividade da soja, milho e trigo (VERÓN *et al.*, 2015). No Brasil, a produtividade de soja reduziu em 42% e 39% no Paraná e Rio Grande do Sul, nessa ordem, devido a alteração na precipitação nos anos de *El-niño* e *La-niño* (FERREIRA, 2010). Por outro lado, com frequência, os níveis de CO<sub>2</sub> tem aumentado a produtividade de grãos (HAO *et al.*, 2014; WANG *et al.*, 2015b).

## 2.2 Método sistemático

A temática abordada neste capítulo é a discutida na seção anterior (introdução previa), para explicá-la foi aplicado um método sistemático, explícito e reprodutivo. Esta revisão foi conduzida da seguinte forma: extrair os pontos impactantes na produtividade de grãos; expor as áreas onde ocorreram as pesquisas; e auxiliar no desenvolvimento de novas pesquisas.

De acordo com Johansson e Sundin (2014) uma revisão sistemática da literatura é um método que utiliza fases para garantir uma abordagem transparente e precisa. Para Ceulemans *et al.* (2015) a revisão sistemática possui cinco etapas: (1) formular a questão; (2) localizar os estudos; (3) selecionar e avaliar os estudos; (4) analisar e sintetizar; e (5) elaboração do relatório e uso dos resultados. Nesta revisão, a subseção mudanças climáticas corresponde à primeira etapa, as subseções método sistemático e mapa conceitual da produtividade de grãos, descrevem as etapas 2, 3 e 4, por último, na etapa 5 são apresentados os impactos observados.

Para manter a transparência é necessário informar a base de dados, termos chave e escolhas dos dados utilizados (SAUNDERS *et al.*, 2009). Os artigos foram baixados através do portal de Periódicos da CAPES, que possui dados eletrônicos e detêm o acesso a outras bases como Elsevier, Emerald, Taylor & Francis, IEEE, Springer, Wiley e outros.

A pesquisa bibliográfica se limitou às palavras-chave nas revistas de Engenharias

III, como: *Agricultural Systems; Atmospheric Environment; Atmospheric Research; Climatic Change; Food Control; Food Quality and Preference; Food Research International; Hydrological Process; Journal of Food Engineering; Journal of Food Science and Technology; Marine Geology; Marine Environmental Research; e Science of the Total Environment.*

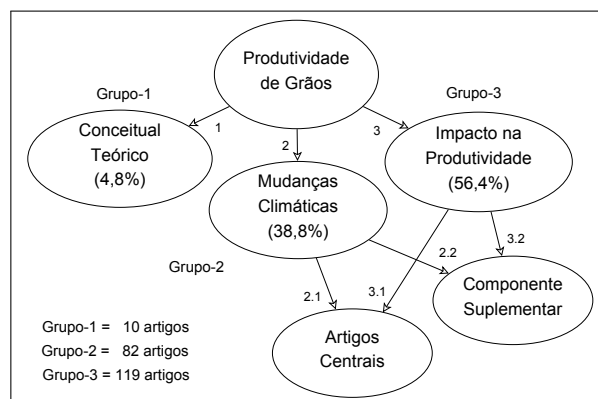


Figura 1 – Classificação temática dos artigos preliminares incluídos na revisão

Devido a extensa lista de artigos publicados (1.092), optou-se por adotar os com maior relevância para a estrutura do trabalho, publicados principalmente entre os anos de 2013 a 2017. Para selecionar os artigos, partiu-se da produtividade de grãos e então, criaram-se três grupos: (1) conceitual teórico; (2) mudanças climáticas; e (3) impacto na produtividade (ver a Figura 1). No grupo-2 foram incluídos artigos que analisaram um período superior a 30 anos e/ou realizaram previsões de produtividade de grãos. No grupo-3 foram incluídos artigos que continham a produtividade de soja, arroz, milho, trigo e sorgo. Essa técnica de subdivisão em grupos foi aplicada por [Ceulemans et al. \(2015\)](#) e [Garza-Reyes \(2015\)](#) de forma a classificar os artigos a serem incluídos ou excluídos da revisão.

As revistas *Food Quality and Preference*, *Hydrological Process*, *Journal of Food Engineering*, *Journal of Food Science and Technology*, *Marine Geology* e *Marine Environmental Research* foram excluídas por não terem artigos classificados. Foram analisados 211 artigos que incluíam a produtividade de grãos (ver Figura 2). No grupo 1, 2 e 3 foram incluídos 10, 82 e 119 artigos, nessa ordem. No total, foram excluídos 881 artigos.

Posteriormente, essa classificação levou a um mapa conceitual que resultou no fluxo de pesquisas e subseções. Também, durante a revisão foi observado: (1) definições e especificações; (2) as comparações dos resultados; (3) método de tratamento estatístico; e (4) os resultados estatísticos ([BERKHOUT et al., 2015](#)). Essas informações foram



utilizadas para construção das seções posteriores, metodologia e resultados.

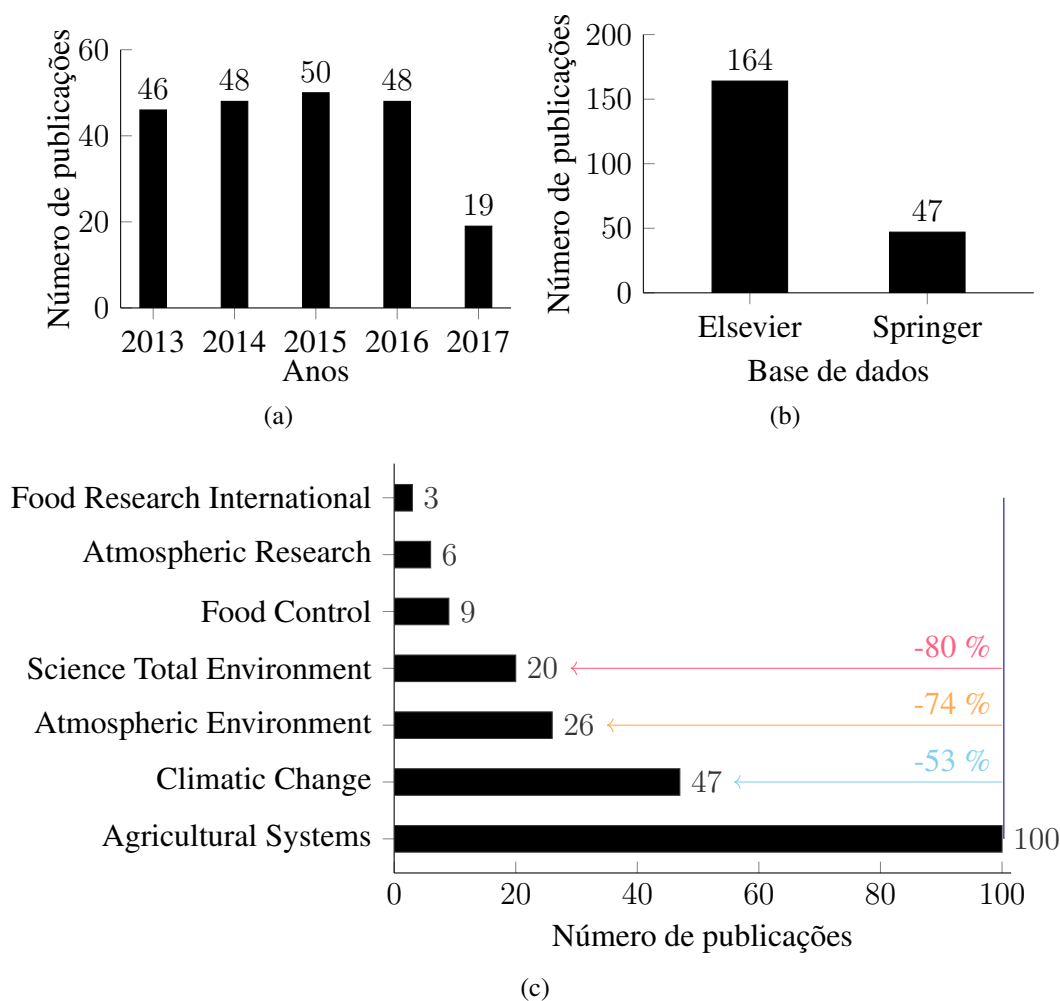


Figura 2 – Descrição das publicações sobre produtividade de grãos e mudanças climáticas

### 2.3 Mapa conceitual da produtividade de grãos

O mapa conceitual foi criado para classificar, organizar, visualizar e estruturar as discussões desta revisão sistemática da literatura (ver Figura 3). Os trabalhos que mais contribuíram foram [Cammarano et al. \(2016\)](#), [Lobell et al. \(2011\)](#), [Kool et al. \(2014\)](#), [Pereira et al. \(2015\)](#) e [Twine et al. \(2013\)](#), atuando de forma interdisciplinar no cenário.

No mapa conceitual foram encontrados diversos fluxos de pesquisas dentro das subdivisões cenário e modelo de agricultura. No cenário foram identificadas pesquisas como mudanças e influência do clima, alteração na produção de grãos por melhoramento de sementes ([TACK; HOLT, 2016](#)) e gestão ([BERKHOUT et al., 2015](#)). Mais detalhes sobre a relação entre o clima e a produtividade de grãos veja a seção 2.4.

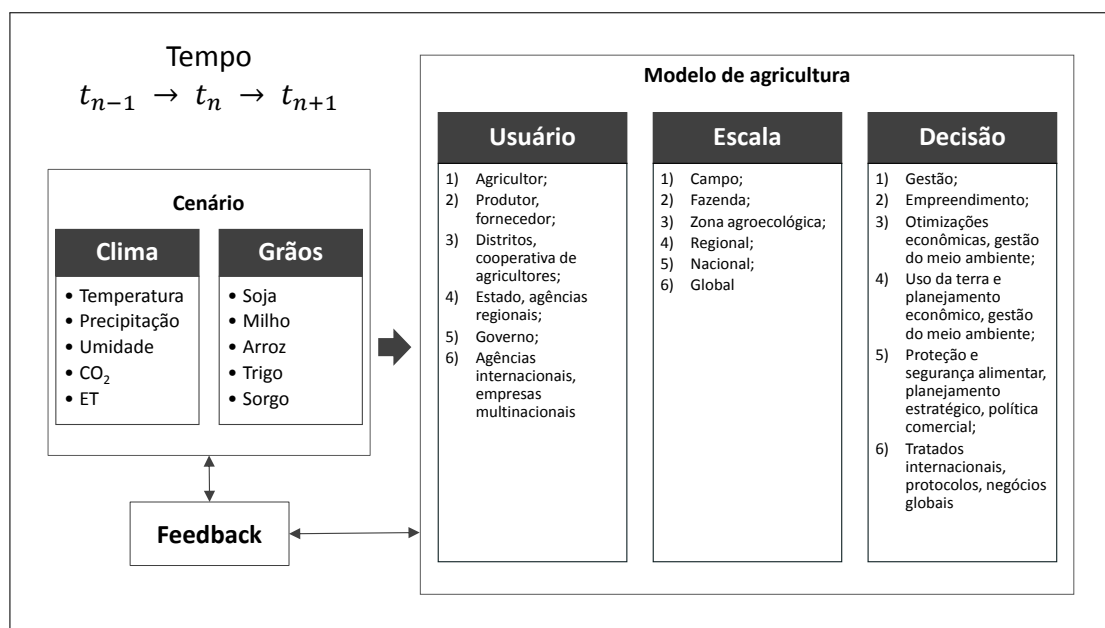


Figura 3 – Mapa conceitual da produtividade de grãos no tempo e escala global

No modelo de agricultura os usuários buscam maneiras de gerenciar condições que envolvam variáveis climáticas, físicas, biológicas e socioeconômicas, que variam em escala e tempo (JONES *et al.*, 2017a). O processo de tomada de decisão em um sistema agrícola pode ser estruturado (ROBERT *et al.*, 2016), como: (1) definir um problema para construção de teorias a serem testadas; (2) selecionar casos como pesquisas estatísticas; (3) levantar dados e realizar análise; e (4) realizar um modelo conceitual.

Assim, as previsões poderiam fornecer aos agricultores informações suficientes para adaptar suas ações às mudanças climáticas correspondentes e, conseqüentemente, ajustar seus comportamentos de oferta e mercado (FERNANDEZ *et al.*, 2016). Tais previsões potencialmente produzem informações para que o agricultor ou tomador de decisão (do nível de campo ao global, ver Figura 3) adote medidas pelas quais estaria disposto a pagar.

Modelada a influência das mudanças climáticas na produtividade de grãos podem ser realizadas análises, como o desenvolvimento do índice de área folhear pela disponibilidade de água. Por exemplo, onde os níveis de evaporação são altos, a disponibilidade de água é baixa (LIU *et al.*, 2016). Então, o plantio deve ser dimensionado de forma adequada (dentro da escala no modelo de agricultura, ver Figura 3). Assim, com a compreensão do cenário e escala podem ser aplicadas práticas de gestão agrícola, como por exemplo aumentar a transpiração (KOOL *et al.*, 2014). No entanto, métodos que contenham a evapotranspiração de referência fornece uma interpretação robusta baseada em dados

climatológicos e aplicação das curvas de cultura (isso ocorre porque a  $ET_0$  é baseada na grama), o que permite projetar a necessidade da superfície de irrigação embasado no balanço hídrico (PEREIRA *et al.*, 2015).

No caso do estresse de temperaturas extremas, alguns grãos são sensíveis a temperaturas alta e baixa, podendo reduzir sua produtividade pela metade (WANG *et al.*, 2015b). Então, as mudanças climáticas (ver cenário na Figura 3) devem ser observadas periodicamente por afetarem diretamente a produtividade de grãos (LOBELL *et al.*, 2011). No modelo de agricultura, é importante garantir o aumento na produtividade para manter a produção agrícola, podendo ser obtida através de práticas de gestão (GAO *et al.*, 2014), melhoria genética (TACK; HOLT, 2016) e interação com o clima (VERÓN *et al.*, 2015).

As alterações no preço de commodities também serão um desafio, isso porque o valor pago pelos grãos está aumentando (MORI *et al.*, 2010). O preço da produção (de forma genérica) está associado a sua produtividade, ao preço pago pelo produtor por matéria-prima e insumo, além da lei de oferta e procura (CALDARELLI; BACCHI, 2012). Sendo assim, de forma fácil, para formar o preço dos grãos (análise em maior escala, como regional e global, ver Figura 3) podem ser aplicados modelos que contenham a análise do cenário e do modelo de agricultura cruzando com valores econômicos. Deve-se salientar também, que a alteração da demanda impacta na área plantada, o que ocasiona o avanço da extensão do plantio para responder a demanda mundial. Além de que, quando tomada a decisão de plantar uma cultura é deixado de plantar outra.

## 2.4 Produtividade de grãos

As mudanças climáticas afetaram negativamente a produtividade de grãos (WANG *et al.*, 2015b) principalmente através de alterações na temperatura, precipitação, concentração de  $CO_2$ , variabilidade climática e escoamento superficial da água (CALZADILLA *et al.*, 2013). A produtividade de grãos também é afetada pela alteração na frequência de eventos climáticos extremos (JU *et al.*, 2013) como inundações, secas, ondas de calor, tornados, entre outros (FERREIRA, 2010), que estão associados a alteração na temperatura dos oceanos (FERNANDEZ *et al.*, 2016).

Se aumentar em  $1^\circ C$  a temperatura média, o milho e o trigo irão reduzir o tempo da estação de crescimento em 6 e 10 dias, nessa ordem (CAMMARANO *et al.*, 2016). O

aquecimento interfere na evaporação da umidade do solo, aumentando-a. Esse fator agravou a salinidade e alcalinidade do solo do rio Yellow, na China, reduzindo a produtividade de culturas (JU *et al.*, 2013), além de proporcionar uma tendência de desertificação.

Na região dos Pampas no oeste da Argentina, com as mudanças climáticas o padrão da chuva mudou, tornando assim, possível plantar trigo no inverno (ASSENG *et al.*, 2013). Porém, na Austrália Ocidental se as mudanças climáticas permanecerem, as condições serão desfavoráveis para plantio de trigo (ASSENG; PANNELL, 2013). O excesso de precipitação pode trazer danos a produção de grãos, como: favorecer o ataque de pragas; prejudicar a aeração do solo; alterar a incidência de radiação; e dificultar o processo de colheita. No entanto, mesmo em regiões chuvosas o déficit hídrico também pode prejudicar a cultura, caso haja um intervalo de 15 dias sem água disponível, sendo necessária a irrigação (GAVA, 2014).

Atualmente os níveis de CO<sub>2</sub> em média são de 375 ppm e se aumentado para 550 ppm o milho e a soja irão responder com o fechamento dos estômatos, o que irá diminuir o fluxo de água para a atmosfera e alterar o índice de área foliar, incrementando a produtividade da soja (TWINE *et al.*, 2013). Em outra análise, Hussain *et al.* (2013) ao trabalhar com a saturação de 541, 545 e 585 ppm chegaram a conclusão de que o milho e a soja irão perder sua produtividade. Então, Hao *et al.* (2014) conduziram dois experimentos na China com o uso de Mini-FACE e observaram que com o aumento de CO<sub>2</sub> o metabolismo da soja é acelerado, aumentando a produtividade (26 e 31%), a quantidade de óleo extraído (29 e 34%) e a concentração de proteína (24%). Assim, a alteração dos níveis de CO<sub>2</sub> pode compensar o aumento da temperatura e a perda de umidade do solo (JONES *et al.*, 2017a).

Com alterações na concentração de CO<sub>2</sub> confirmadas, a produtividade de grãos será afetada por períodos frios e quentes. Para o milho e o trigo a produtividade irá aumentar em 4,8% e 14% (de 310 a 450 ppm) em períodos quentes; 3,8% e 11% (de 320 a 450 ppm) para períodos frios; e 2,2% e 7% (de 354 a 450 ppm) em períodos neutros (CAMMARANO *et al.*, 2016).

As culturas de grãos são propensas a doenças causadas por fungos e bactérias, o que pode gerar impactos econômicos significativos devido a perdas na produção, e também ameaças a segurança alimentar (OLIVEIRA *et al.*, 2017; SAVI *et al.*, 2016). A contaminação dos grãos pode ocorrer durante o cultivo, a colheita, o armazenamento e o

transporte (TWARUŹEK *et al.*, 2013). Essas epidemias nas culturas (KIM; CHO, 2016) podem estar relacionadas às práticas de gestão, como rotação de culturas, plantio direto e cultivares (KOS *et al.*, 2017; MÜLLER; KORN, 2013).

As mudanças climáticas interferem no processo de crescimento e desenvolvimento das plantas, assim como na ocorrência de doenças (JUNK *et al.*, 2016). Durante o cultivo dos grãos, as condições climáticas podem levar a um aumento do crescimento de fungos e da produção de toxinas. Além do impacto na produtividade de grãos, as mudanças climáticas também podem impactar na qualidade desses grãos, como no caso dos grãos de café que em condições mais úmidas e quentes são contaminados por fungos (PATERSON *et al.*, 2014) ou no caso do milho que é contaminado por toxinas devido a secas prolongadas e ao aumento da temperatura, na Sérvia (KOS *et al.*, 2013; KOS *et al.*, 2014).

No caso do milho, na Servia, os principais fatores que interferem na contaminação dos grãos por determinados tipos de fungos são seca prolongada, alta temperatura, tempo e condições de armazenamento (KOS *et al.*, 2013). Na Alemanha, a contaminação do trigo por substâncias tóxicas produzidas por fungos (micotoxinas) predominou em anos úmidos (MÜLLER; KORN, 2013). No Brasil, as micotoxinas também tem impactado na qualidade dos grãos de milho (OLIVEIRA *et al.*, 2017).

A produtividade e a qualidade dos grãos também podem ser afetadas pela presença de metais pesados no solo. Para reduzir a absorção desses metais pesados pelas plantas e aumentar o rendimento são utilizadas práticas agronômicas e reguladores químicos. A variedade do grão também pode influenciar no nível de absorção (CAO *et al.*, 2014).

Na China, devido ao aquecimento global, os padrões de cultivo têm sido alterados, de sistemas de culturas simples para rotação de culturas (JU *et al.*, 2013). Isso impacta na produtividade de grãos. Em um estudo na China, com a adoção de sistemas de tripla cultura (trigo de inverno, milho de verão e milho de primavera), os resultados apontam altos rendimentos dos grãos, diminuição das emissões de gases de efeito estufa (óxido nitroso) e uso sustentável de águas subterrâneas (GAO *et al.*, 2014).

Na produção de arroz (com uma cultura) as alterações climáticas, sem considerar os níveis de CO<sub>2</sub>, causaram perda econômica líquida entre \$17,4-53,0 milhões, na China, entre os anos de 1996 a 2009 (CHEN *et al.*, 2016). E, nos Estados Unidos, para garantir a segurança alimentar são investidos entre \$0,8–3,48 bilhões de dólares por ano, caso

contrário a agricultura perderia \$3,2–8,0 bilhões de dólares (JU *et al.*, 2013).

### 3 METODOLOGIA

A produtividade de grãos vem sofrendo alterações devido as mudanças climáticas. Esses impactos podem ser avaliados por diferentes métodos. Neste trabalho, será utilizada a regressão PLS para modelar e verificar a influência das mudanças climáticas na produtividade de grãos.

No presente capítulo é apresentada a caracterização da pesquisa, a definição da área de estudo e base de dados, os métodos utilizados para completar as falhas nas séries históricas e a definição dos métodos para calcular a evapotranspiração. Por último, o ajuste da produtividade de grãos utilizando séries temporais e regressão PLS.

#### 3.1 Caracterização da pesquisa

Ambas as abordagens (qualitativa e quantitativa) têm seus pontos fortes e fracos durante o projeto e aplicação. A abordagem a ser usada no planejamento de uma pesquisa depende de vários fatores, como o tipo de pergunta(s), treinamento ou experiência do pesquisador e o público alvo (YILMAZ, 2013). Neste trabalho, a abordagem é quantitativa aplicada, ou seja, é apropriada para testar e gerar teorias, permitindo também o trabalho de forma exploratória. Assim, tem-se a finalidade de esclarecer os melhores métodos para se obter os dados que irão modelar a produtividade de grãos e verificar qual método se ajusta melhor, além de permitir verificar hipóteses.

A revisão de literatura aponta quais métodos são adequados para alcançar os objetivos, além de possibilitar a identificação de pesquisas científicas em aberto (CAUCHICK-MIGUEL, 2007). Para elaboração dos métodos a revisão se concentrou nas revistas *Journal Statistical Software* e *Chemometrics and Intelligent Laboratory Systems*.

O método aplicado é de modelagem estocástica, pois, em particular, permite a construção de modelos conceituais e experimentais (BERTRAND; FRANSOO, 2002) e a aplicação de ferramentas estatísticas como séries temporais e regressão em dados históricos (CAUCHICK-MIGUEL *et al.*, 2012). Os métodos foram divididos em duas etapas: (1) levantar os dados que influenciam na produtividade de grãos, como: temperatura; umidade; precipitação; evapotranspiração; e concentração de dióxido de carbono (CO<sub>2</sub>) no ar; e (2) avaliar o ajuste da produtividade de grãos por séries temporais e pelas variáveis climáticas.

Na primeira etapa, a decisão do melhor método para fazer a previsão da temperatura, precipitação e umidade foi o teste de Tukey e para a evapotranspiração foi o coeficiente de Eficiência Nash-Sutcliffe (NSE). O teste de Tukey foi aplicado para avaliar o melhor método para completar os dados. O coeficiente NSE foi aplicado porque, além de medir os erros, ele verifica a magnitude de variabilidade.

Na segunda etapa, no ajuste da produção de grãos por séries temporais foi adotado o Critério de Informação de Akaike (AIC) e Bayesian (BIC), por penalizar a quantidade de parâmetros no modelo. No ajuste pelos dados climáticos, o melhor modelo foi selecionado, durante a validação-cruzada, pelo critério de  $Q_{cum}^2$ , que mede a capacidade de predição do modelo.

### 3.2 Área de estudos e base de dados

Neste trabalho é estudada a produtividade de grãos no Brasil. O país possui mais de 16 mil quilômetros de fronteira e ocupa a área de 47,89% da América do Sul ([ALMANAQUE-BRASIL, 2012](#)). Em consequência do relevo e extensão territorial, o seu clima é variável, predominado o tropical, equatorial, árido e subtropical.

Os grãos com maior produção no Brasil são a soja (*Glycine max*) e o milho (*Zea mays*). No total, foram produzidas 238.000 MT em 2016 (ver Figura 4). As regiões que concentram a maior produção de grãos é o Centro-Oeste (44%) e o Sul (35%). As informações da produtividade (produção/área) de grãos estão disponíveis no site da Companhia Nacional de Abastecimento ([CONAB](#)). Os dados estão disponíveis a partir de 1977, entretanto, as análises foram iniciadas em 1980 devido aos dados climáticos (ver seção 4.1).

Os dados climáticos diários foram acessados no banco de dados do Instituto Nacional de Meteorologia ([INMET](#)) e do *National Oceanic and Atmospheric Administration* ([NOAA](#)). Para baixar os dados no INMET foi realizada uma adaptação no algoritmo de [Tatsch \(2016\)](#) para acessar os metadados e para os dados do NOAA foi elaborado um *script* em R com as informações de [Piccirilli \(2015\)](#). Estão disponíveis 265 e 922 Estações Meteorológicas (EMs) no INMET e NOAA, nessa ordem.

As variáveis disponibilizadas pelo INMET são: temperatura média, mínima e máxima em graus Celsius ( $^{\circ}\text{C}$ ); precipitação em milímetros (mm); umidade relativa do ar em porcentagem (%); velocidade dos ventos em metros por segundos ( $\text{m}\cdot\text{s}^{-1}$ ) e evapotrans-



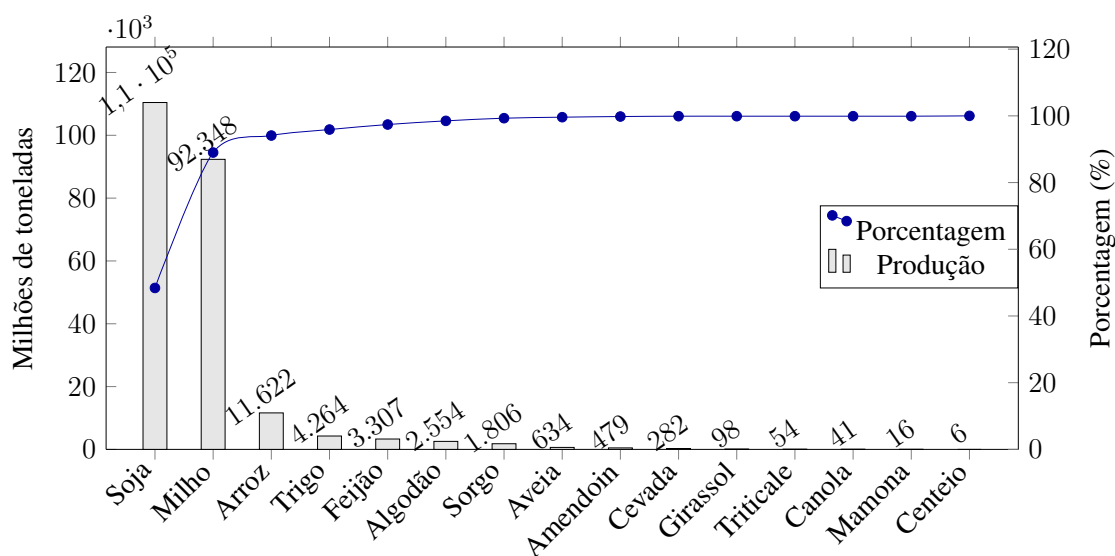


Figura 4 – Diagrama de Pareto da produtividade de grãos no Brasil em 2016

piração em milímetros por dia ( $\text{mm}\cdot\text{d}^{-1}$ ). As variáveis disponibilizadas pelo NOAA são: temperatura média ( $^{\circ}\text{C}$ ); e velocidade dos ventos ( $\text{m}\cdot\text{s}^{-1}$ ).

As séries históricas possuem dados faltantes, entretanto, foi aplicado o método de [Vicente-Serrano et al. \(2010\)](#) e [Ferrari e Ozaki \(2014\)](#) para completar as falhas. O preenchimento dos dados faltantes por essas técnicas faz uso de informações de EMs vizinhas. Devido aos dois bancos de dados não possuírem as mesmas variáveis, os dados do NOAA não foram completados, mas utilizados para completar dados do INMET. Mais detalhes sobre preenchimento dos dados faltantes e do cálculo da evapotranspiração para determinar a matriz com informações do clima, serão abordados na seção 3.3 e 3.4.

Os dados climáticos diários foram ajustados para o período das estações (primavera, verão, outono e inverno). Nesse intervalo: a precipitação e a evapotranspiração de referência ( $\text{ET}_0$ ) foram acumuladas; na temperatura e umidade (máxima e mínima) foram aplicados os extremos; e na temperatura e umidade (média) foram utilizadas as médias. O intervalo diurno da temperatura ( $\text{T}_{\text{dtr}}$ ) e a umidade do ar ( $\text{UR}_{\text{dtr}}$ ) também foram calculadas (ex.: máximo - mínimo). O nível de  $\text{CO}_2$  mensal, de 1980 a 2016, foi disponibilizado pelo NOAA. Os dados de  $\text{CO}_2$  gerenciados pelo NOAA são gerados pela técnica de simulação de Monte Carlo ([CONWAY et al., 1994](#)).

Durante a modelagem, existe a hipótese de algumas EMs na região  $r = \text{Centro-Oeste}$  e Sul não serem importantes e outras, fora da divisão política da região, sim. Por isso, as EMs que representam cada região devem ser selecionadas com base na representativi-

dade da região agrícola (CAMMARANO *et al.*, 2016). Para determinar quais EMs que representam cada região, foram calculadas as matrizes de ruídos aleatórios  $\mathbf{A}^k$  das séries

$$\mathbf{A}^k = \mathbf{X}^k - \mathbf{S}^k - \mathbf{TD}^k \quad (3.1)$$

em que:  $\mathbf{A}^k$  são matrizes de ruídos aleatórios com média zero e variância constante  $\sim N(0, \Sigma^2)$  para as  $k$  variáveis climáticas;  $\mathbf{X}^k$  são matrizes com as séries observadas;  $\mathbf{S}^k$  são matrizes com as sazonalidades; e  $\mathbf{TD}^k$  são matrizes com as tendências. Todas as matrizes são de ordem  $t \times v$ , onde  $v$  é o número de EMs e  $t$  é o tempo.

A matriz de correlação  $c_{ij}(v \times v)$  para as  $k$  variáveis climáticas com correlação acima de 50% em relação as capitais dos estados foram agrupadas em macrorregiões ( $r =$  Centro-Oeste e Sul). Para ilustração (ver Figura 5), foi realizada a interpolação multivariada para obter os limites de correlação (AKIMA, 1978). Para a temperatura na capital de Goiás, devem ser utilizadas EMs no estado do Pará, Tocantins, Maranhão e Minas Gerais. Também, existem algumas EMs em outros estados, como na Bahia e São Paulo.

Determinadas as EMs que representam cada região, foi utilizada a média das  $k$  variáveis no instante  $t$ , então para cada  $k$  obtemos um vetor coluna ( $\mathbf{z}_i^c = 1/v \sum_{j=1}^v x_{ij}^k$ , com  $i = 1, \dots, t$ ). Resultando nas matrizes  $\mathbf{Z}^r$  com informação do clima (ordem  $t \times k$ ). O *boxplot* com os dados da matriz  $\mathbf{Z}^r$  está na Figura 6-a e 6-b. Devido as diferenças na unidade de medida, os dados climáticos que representam cada região foram escalados para variância unitária, pela subtração da média e divisão pelo desvio padrão de sua respectiva variável ( $\mathbf{x}_i^c$  e  $\mathbf{y}_i^c$ ) (ver Figura 6-c e 6-d). Isso corresponde a  $\mathbf{Z}^*$  e  $\mathbf{Y}^*$  com a mesma importância e peso em uma análise a priori (ZHANG *et al.*, 2009)

$$z_{ij}^* = \frac{z_{ij} - \bar{z}_j}{s_j^z}, y_{ij}^* = \frac{y_{ij} - \bar{y}_j}{s_j^y} \quad (3.2)$$

em que:  $\bar{z}_j = \frac{1}{n} \sum_{i=1}^n z_{ij}$  e  $\bar{y}_j = \frac{1}{n} \sum_{i=1}^n y_{ij}$  são vetores com as médias;  $s_j^z = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (z_{ij} - \bar{z}_j)^2}$  e  $s_j^y = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_{ij} - \bar{y}_j)^2}$  são vetores com os desvios padrão.

Os dados de produtividade de grãos são fornecidos por safras (anual) e os dados climáticos estão ajustados por estações (trimestral). Por esse motivo, os dados das variáveis predictoras foram ajustados de onze para 44 (exemplo: precipitação na primavera,  $\dots$ ,

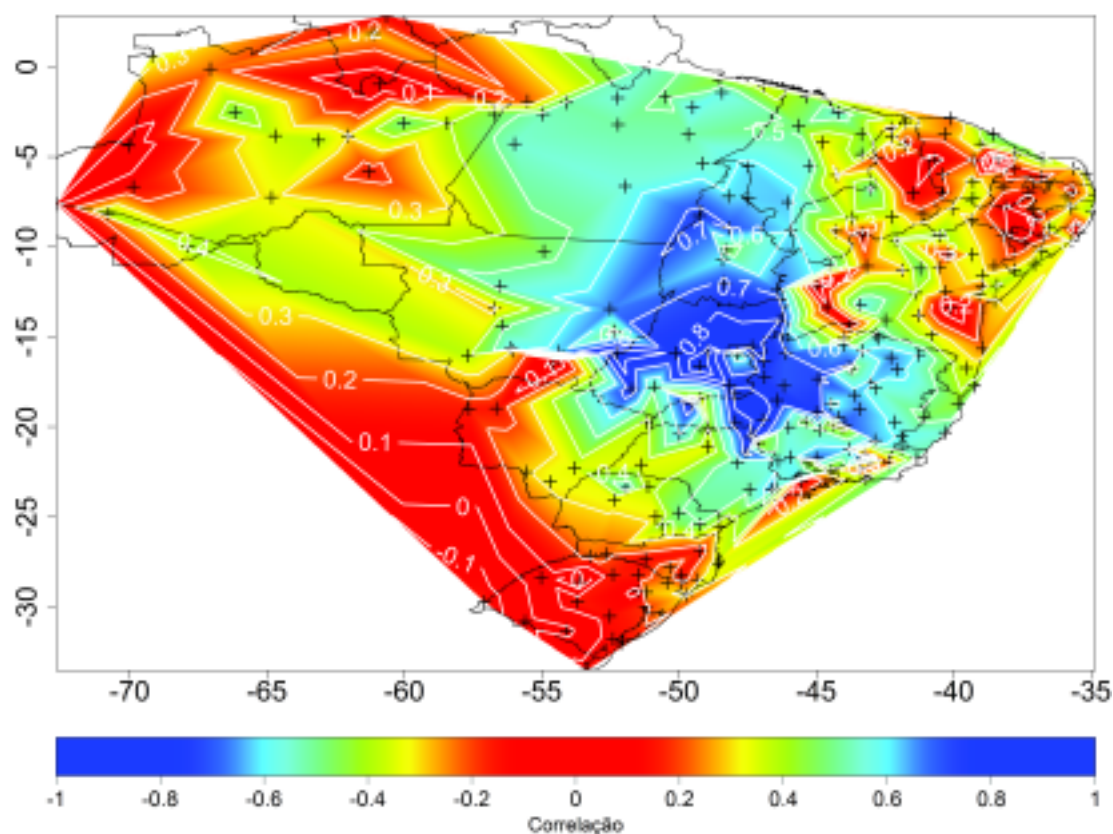


Figura 5 – Agrupamento das estações meteorológicas para a variável temperatura no estado de Goiás

precipitação no inverno) e as observações foram atualizadas para  $i = 1, \dots, n$ , onde  $n$  é igual a 37 anos.

### 3.3 Variáveis climáticas

Os impactos causados pelas mudanças climáticas fizeram com que a demanda por dados históricos sobre o clima aumentasse significativamente (FERNÁNDEZ-FERNÁNDEZ *et al.*, 2014; RODRIGO, 2012). Entretanto, esses estudos requerem um conjunto de dados controlados que geram algumas dificuldades (SERRANO-NOTIVOLI *et al.*, 2017), como encontrar séries de dados climáticos com qualidade e dados em regiões específicas onde não há EMs. A baixa qualidade pode ocorrer devido à falta de dados em dias de manutenção (INMET, 2017b), avaria de estações meteorológicas, erros de digitação (VINCENT *et al.*, 2005) e erros nos metadados (VIANNA *et al.*, 2017).

Para obter os dados históricos foram adotadas as seguintes etapas: (1) obter metadados (informações do equipamento que coleta os dados, como código de identificação, localização das EMs e a altura em relação ao nível do mar); (2) analisar as distâncias

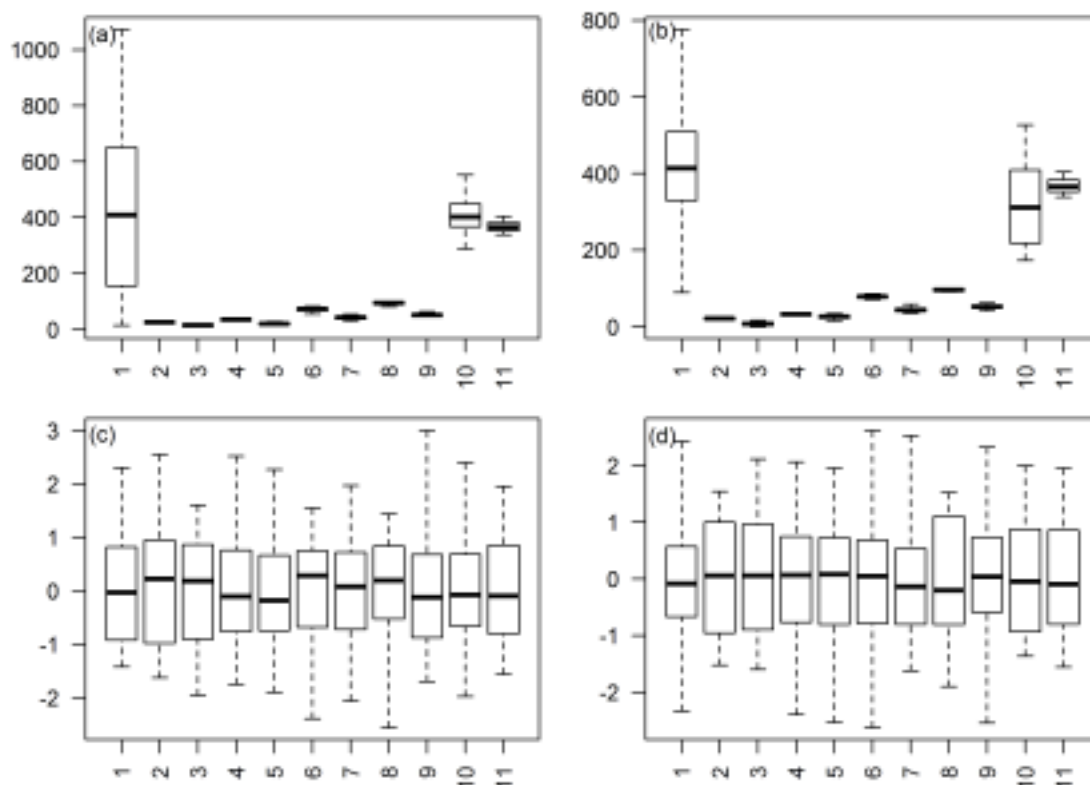


Figura 6 – *Boxplot* com informações do clima de 1980 a 2016. (a) variáveis originais da região Centro-Oeste. (b) variáveis originais da região Sul. (c) variáveis transformadas da região Centro-Oeste. (d) variáveis transformadas da região Sul

Nota. As variáveis no eixo x são: 1 é precipitação; 2 é temperatura média; 3 é a temperatura mínima; 4 é a temperatura máxima; 5 é o intervalo diurno da temperatura; 6 é a umidade média; 7 é a umidade mínima; 8 é a umidade máxima; 9 é o intervalo diurno da umidade; 10 é a evapotranspiração; 11 é a concentração de dióxido de carbono no ar.

entres as EMs; (3) baixar os dados do clima; (4) corrigir as datas faltantes; (5) determinar o melhor método para completar os dados; (6) completar os dados faltantes; e (7) verificar erros e *outliers*.

Devido a uma EM poder estar fora da região agrícola  $r$ , foram baixados os dados do clima do Brasil. A área possui 8.516.000 Km<sup>2</sup>, com a densidade espacial de dados considerada razoável (uma estação meteorológica por 7.174 Km<sup>2</sup>). No total foram analisadas 1.187 EMs no Brasil. As estações utilizadas estão ilustradas na Figura 7.

No INMET, 65% das EMs possuem mais de 50 anos. A EM mais velha possui 120 anos de funcionamento. Entretanto, os dados do INMET estão disponíveis a partir de 1961, por esse motivo, iniciou-se o preenchimento de 1961 a 2016 (56 anos ou 20.454 observações).

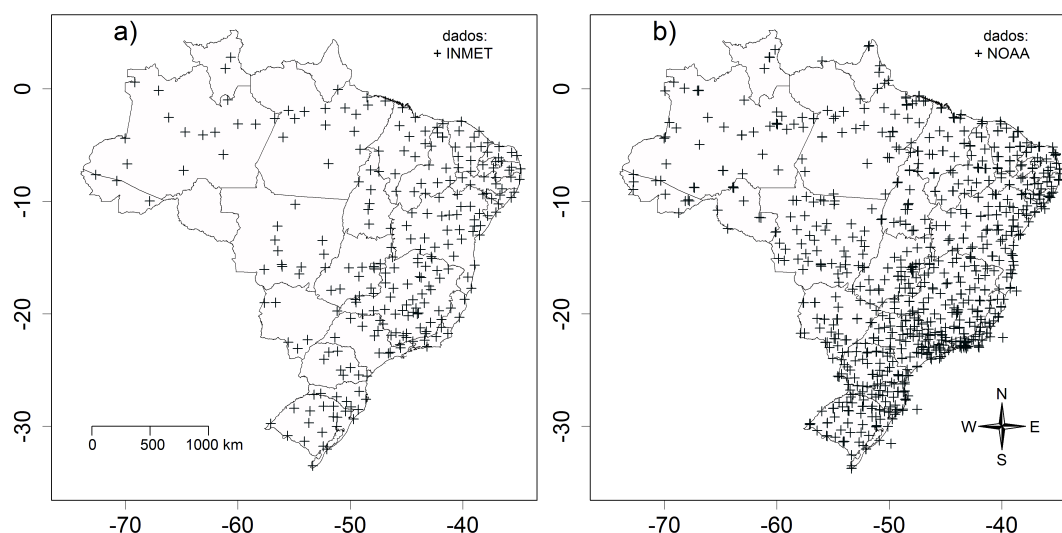


Figura 7 – Estações meteorológicas utilizadas no Brasil. (a) Localização dos dados do INMET. (b) Localização dos dados do NOAA

Antes de determinar o melhor método, as variáveis em cada EM foram classificadas em três grupos “x”, “y” e “z”. Isso faz necessário para determinar o melhor método com as melhores séries. As EMs que possuem menos de 25 anos de funcionamento ou séries (variável  $k$ ) com mais de 7.300 observações faltantes foram consideradas curtas para serem utilizadas no início da reconstrução. Esse grupo foi rotulado por “z” e aplicado na reconstrução de outras séries.

A série “x” inclui EMs com mais de 25 anos e séries com menos de 3.650 observações faltantes. A série “y” inclui EMs com mais de 25 anos e séries com 3.650 até 7.300 observações faltantes. Assim, as séries “x” e “y” foram consideradas melhores para iniciar a reconstrução por possuírem EMs com mais de 25 anos e séries com menos falhas. Para a precipitação acumulada ( $P_{ad}$ ) 79 EMs possuem entre 3.650 e 5.475 observações faltantes (ver Tabela 1), em um total de 20.545 observações (56 anos), o que corresponde a falhas entre 17% a 27% dos dados.

### 3.3.1 Método de preenchimento com informações de estações vizinhas

Para completar os dados foram considerados procedimentos baseados na informação de EMs vizinhas disponível a uma determinada distância. Foram testados três métodos diferentes: o vizinho mais próximo; a interpolação ponderada pela distância inversa; e o método de regressão linear (FERRARI; OZAKI, 2014). No método de vizinho mais próximo, os dados são preenchidos diretamente com os dados da EM disponível mais

Tabela 1 – Análise do número de estações meteorológicas com dados faltantes das variáveis climáticas

Observações*	$P_{ad}$	$T_{med}$	$T_{max}$	$T_{min}$	$UR_{med}$	$UR_{max}$	$U_z$
0-182	0	3	0	0	0	0	0
182-365	3	3	2	3	0	1	2
365-730	17	14	10	10	6	9	12
730-1.825	29	23	21	28	16	23	17
1.825-3.650	67	73	61	57	58	59	59
3.650-5.475	79	70	75	69	62	70	62
5.475-7.300	44	46	62	61	68	63	62
7.300-9.125	12	18	17	20	32	24	26
9.125-10.950	8	8	10	9	9	9	15
10.950-12.775	4	3	3	5	5	2	2
12.775-14.600	0	2	2	1	5	2	2

Nota:  $P_{ad}$  é a precipitação acumulada;  $T_{med}$  é a temperatura média;  $T_{max}$  é a temperatura máxima;  $T_{min}$  é a temperatura mínima;  $UR_{med}$  é a umidade média;  $UR_{max}$  é a umidade máxima; e  $U_z$  é a velocidade dos ventos. \* dados em dias

próxima.

Para interpolação [Equação (3.3)] todas as séries vizinhas são consideradas baseando na ponderação de distância inversa, de forma que  $\mathbf{z}_j^k$  é uma variável de uma EM e que o valor  $\mathbf{z}_j^k(t)$  é faltante no instante  $t$ , assim

$$\hat{\mathbf{z}}_j^k(t) = \frac{\sum_{i=1}^v \mathbf{z}_i^k(t) \cdot d_{ij}^{-2}}{\sum_{i=1}^v d_{ij}^{-2}} \quad (3.3)$$

em que:  $\mathbf{z}_j^k$  é um vetor de variável climática ( $k =$  temperatura, precipitação e umidade);  $\hat{\mathbf{z}}_j(t)$  é o valor estimado de acordo com a média ponderada dos vizinhos na posição ( $t$ )  $\{\mathbf{z}_i = \mathbf{z}_1(t), \mathbf{z}_2(t), \dots, \mathbf{z}_v(t)\}$ , onde  $v$  é o número de EMs vizinhas dentro de uma distância específica  $\{i = 1, \dots, v\}$ ; e  $d$  é a distância entre  $\mathbf{z}_i$  e  $\mathbf{z}_j$  (em Km).

No método de regressão linear [Equação (3.4)], os dados faltantes foram modelados pelos vizinhos mais próximos. O intercepto foi forçado a passar pela origem, com o objetivo de evitar valores negativos que possam ocorrer nas variáveis quando a precipitação é baixa (ALLEN; DEGAETANO, 2001)

$$\hat{\mathbf{z}}_j^k(t) = \sum_{i=1}^v \beta_i \mathbf{z}_i^k(t) + \epsilon = \left\langle \beta, \mathbf{Z}(t) \right\rangle + \epsilon \quad (3.4)$$

em que:  $\hat{z}_j^k(t)$  é o dado estimado na posição ( $t$ ) para variável  $k$ ;  $Z_i$  são os dados das estações dentro de uma distância;  $\beta$  é o parâmetro de inclinação; e  $\epsilon$  é o erro.

Para verificar quais métodos de preenchimento com informação das EMs vizinhas possui a melhor acurácia, foi utilizado a análise de variância (ANOVA). Quando a anova dá um resultado significativo, isso indica que pelo menos um grupo difere dos outros (WALPOLE *et al.*, 2016). Porém, não informa quanto é diferente. Para analisar essa diferença, com frequência são utilizadas análises em pares.

A comparação pareada pode ser realizada pela diferença significativa honesta (em inglês, *honestly significant difference* – HSD), conhecida por teste de Tukey. O teste de Tukey verifica a diferença entre duas médias usando a distribuição estatística de Student. Esse procedimento fornece a maior diferença exata entre um conjunto de médias provenientes da mesma população (ABDI; WILLIAMS, 2010).

### 3.3.2 Qualidade dos dados

Ao analisar o histórico climático podem ser observados alguns erros como o de digitação. Vincent *et al.* (2005) encontraram 178,1°C em Ponta Grossa no Paraná. Em uma série temporal, esse tipo de erro pode ser detectado separando os dados por meses para retirar a sazonalidade (janeiro, ..., dezembro) e então é observado se existem valores maiores que os limites de três desvios padrões (SLONOSKY *et al.*, 2001). Assim, se ocorrer um erro de digitação grosseiro ele será identificado.

Na precipitação essa análise não é adequada devido aos dados possuírem uma alta variabilidade espacial e temporal (VICENTE-SERRANO *et al.*, 2010). Por esse motivo, os dados foram convertidos em percentis. Os percentis acima de 99% e abaixo de 1% foram analisados com os vizinhos mais próximos. Caso os valores sejam diferentes, esse dado é tratado como faltante e substituído pelo vizinho mais próximo.

## 3.4 Cálculo da evapotranspiração de referência

Quanto se tem o objetivo de determinar, planejar ou manejar a irrigação um dos parâmetros é calcular a necessidade de água da cultura (ROMERO; DUKES, 2016). Essa definição pode ser feita através do balanço hídrico do solo, que tem como principais variáveis a precipitação e a evapotranspiração (ET). O balanço hídrico é a soma das

entradas e saídas de água no solo (KOOL *et al.*, 2014), então podemos equacionar a ET de uma área como

$$ET = IR + P - RO - DP \pm \Delta S \quad (3.5)$$

em que:  $IR$  é a irrigação;  $P$  é a precipitação;  $RO$  é a superfície de escoamento;  $DP$  é a drenagem interna; e  $\Delta S$  é a alteração no armazenamento de água no solo.

O termo ET foi definido no início na década de 40 quando Thornthwaite expressa o termo evaporação e transpiração de forma simultânea. Depois do encontro organizado pelo Grupo *Food and Agriculture Organization* (FAO), no Líbano em 1971 e Roma em 1972, foi definida a Evapotranspiração de Referência ( $ET_0$ ) como: “a taxa de evapotranspiração de uma superfície coberta com grama com altura uniforme entre 8 a 15 cm, verde, crescendo ativamente, protegendo completamente o chão e sem déficit de água” (DOORENBOS; PRUITT, 1977, p. 1).

A evapotranspiração pode ser medida por métodos diretos e indiretos por informações climáticas. Os métodos diretos incluem diversos tipos de lisímetros e balanço de água no solo (ROMERO; DUKES, 2016). Os métodos com informações de dados climáticos podem ser determinados por Thornthwaite (1948), Penman (1948), Monteith (1965), entre outros.

No método indireto, Penman (1948) combinou variáveis climáticas para estimar a evaporação de “água aberta”, determinando que a transferência de massa de vapor de água é influenciada pelo gradiente de vapor (isto é, a diferença entre a pressão de vapor saturada e a real, que está relacionada com a umidade relativa e temperatura) e a velocidade do vento. Ao lado da superfície evaporativa, uma fina camada de ar não turbulenta fornece resistência ao fluxo de evaporação, conhecida como resistência aerodinâmica, dada como

$$\lambda E = \left[ \Delta (R_n - G) \right] + \left[ \frac{\gamma \lambda E_a}{(\Delta + \gamma)} \right] \quad (3.6)$$

em que:  $\lambda E$  é o fluxo de calor latente de evaporação ( $\text{MJ} \cdot \text{m}^{-2} \cdot \text{d}^{-1}$ );  $\Delta$  é a inclinação da curva de pressão de vapor saturado (kPa);  $R_n$  é o fluxo de radiação ( $\text{MJ} \cdot \text{m}^{-2} \cdot \text{d}^{-1}$ );  $G$  é um fluxo de calor sensível no solo ( $\text{MJ} \cdot \text{m}^{-2} \cdot \text{d}^{-1}$ );  $\gamma$  é a constante psicrométrica ( $\text{kPa} \cdot ^\circ\text{C}^{-1}$ ); e  $E_a$  é o fluxo de transporte de vapor ( $\text{mm} \cdot \text{d}^{-1}$  ou  $\text{in} \cdot \text{d}^{-1}$  ou  $\text{m}^{-2} \cdot \text{d}^{-1}$ ).



A resistência superficial também é importante, pois a transpiração é regulada pelo grau de abertura estomática nas folhas (MONTEITH, 1991). Assim, Monteith (1965) equacionou essa resistência para as folhas de plantas, como

$$\lambda ET = \frac{\left[ \Delta (R_n - G) \right] + \left[ 86.400 \rho_a C_p (e_s^o - e_a) \right] / r_{av}}{\Delta + \gamma (1 + r_s / r_{av})} \quad (3.7)$$

em que:  $\rho_a$  é a densidade do ar ( $\text{kg}\cdot\text{m}^{-3}$ );  $C_p$  é o calor específico do ar ( $\text{kJ}\cdot\text{kg}^{-1}\cdot\text{K}^{-1}$ );  $r_{av}$  é a resistência aerodinâmica da superfície de massa de vapor de água ( $\text{s}\cdot\text{m}^{-1}$ );  $r_s$  é a resistência de superfície do dossel; e  $e^o$  é a pressão de vapor saturado ( $\text{kPa}$ ). Em casos onde a evaporação é denominada em unidade de massa por tempo, elas são expressas por  $E$ ,  $ET$  e  $T$  (usualmente,  $\text{kg}\cdot\text{m}^{-2}\cdot\text{d}^{-1} \approx \text{mm}\cdot\text{d}^{-1}$ ), enquanto que a evaporação em unidade de energia é referida como  $\lambda E$ ,  $\lambda ET_o$  e  $\lambda T$ , em que  $\lambda$  é o fluxo de calor latente ( $\text{J}\cdot\text{kg}^{-1}$ ) (KOOL *et al.*, 2014).

Tabela 2 – Métodos para estimar a evapotranspiração de referência

Modelos de ET	$T_{max}$ , $T_{min}$	$UR_{max}$ , $UR_{min}$	$R_s$	$U_z$
Penman-Monteith FAO-56 (Allen <i>et al.</i> , 1998)	✓	✓	✓	✓
Matt-Shuttleworth (Shuttleworth e Wallace, 2009)	✓	✓	✓	✓
Makkink (DeBruin, 1987)	✓		✓	
Blaney-Criddle (Allen e Pruitt, 1986)	✓	✓	✓	✓
Turc (Guo <i>et al.</i> , 2016 apud Turc, 1961)	✓		✓	
Hargreaves-Samani (Hargreaves e Samani, 1985)	✓			

Nota:  $T_{max}$  e  $T_{min}$  é a temperatura máxima e mínima ( $^{\circ}\text{C}$ );  $R_s$  é a radiação solar ( $\text{MJ}\cdot\text{m}^{-2}$ );  $UR_{max}$  e  $UR_{min}$  é a umidade relativa máxima e mínima do ar (%), e  $U_z$  é a velocidade dos ventos ( $\text{m}\cdot\text{s}^{-1}$ )

A junção desses métodos ficou conhecida como Penman-Monteith e atualmente é utilizado pelo FAO, além de ser bem aceito na área ambiental, agricultura, hidrologia, agronomia, engenharias e outras, também possui mais de 11.500 citações (PEREIRA *et al.*, 2015). Outros métodos para estimar a  $ET_o$  são demonstrados na Tabela 2. Combinando o balanço energético e os componentes de transferência de massa, as quatro variáveis climáticas chave relacionadas com  $ET_o$  são temperatura, umidade relativa, radiação solar e velocidades dos ventos (GUO *et al.*, 2016).

Entre os métodos para calcular a  $ET_o$ , foram testados os métodos de Penman-Monteith, Blaney-Criddle e Hargreaves-Samani. Esses métodos foram testados para verifi-

car qual é o melhor, quanto a umidade ou velocidade dos ventos está ausente. A  $ET_o$  pelo método Penman-Monteith (ALLEN *et al.*, 1998) é calculada pela Equação (3.8)

$$ET_o = \frac{\left[0,408\Delta(R_n - G)\right] + \left[\gamma \frac{C_n}{(T_{cal} + 273)} u_2 (e_s - e_a)\right]}{\Delta + \gamma(1 + C_d u_2)} \quad (3.8)$$

em que:  $ET_o$  é a evapotranspiração de referência ( $\text{mm}\cdot\text{d}^{-1}$  ou  $\text{mm}\cdot\text{h}^{-1}$ );  $\Delta$  é a declividade da curva de pressão de vapor de saturação ( $\text{kPa}\cdot\text{°C}^{-1}$ );  $R_n$  é o saldo de radiação à superfície da cultura ( $\text{MJ}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$ );  $G$  é a densidade do fluxo de calor do solo ( $\text{MJ}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$ );  $T_{cal}$  é a temperatura média calculada do ar a 2 m de altura ( $\text{°C}$ );  $u_2$  é a velocidade de vento a 2 m de altura ( $\text{m}\cdot\text{s}^{-1}$ );  $e_s$  é a pressão de vapor de saturação ( $\text{kPa}$ );  $e_a$  é a pressão parcial de vapor ( $\text{kPa}$ );  $\gamma$  é o coeficiente psicrométrico ( $\text{kPa}\cdot\text{°C}^{-1}$ );  $C_n$  e  $C_d$  são parâmetros constantes que mudam com intervalo de tempo (para dados diários  $C_n = 900$  e  $C_d = 0,34$ ; dados diurno  $C_n = 37$  e  $C_d = 0,24$ ; e dados noturno  $C_n = 37$  e  $C_d = 0,96$ ).

Para estimar o método de Blaney-Criddle [Equação (3.9)] também são necessários os dados de temperatura mínima e máxima, umidade relativa mínima e máxima, radiação solar e velocidade dos ventos. A  $ET_o$  pode ser estimada (ALLEN; PRUITT, 1986) por

$$ET_o = 0,0043UR_{min} - \frac{nh}{NH} - 1,41 + b \cdot P(0,46T_{med} + 8,13) \left[1 + 0,1\left(\frac{E_{elev}}{1000}\right)\right] \quad (3.9)$$

em que:  $ET_o$  é a evapotranspiração de referência ( $\text{mm}\cdot\text{d}^{-1}$ );  $b$  é um valor de ajuste;  $P$  é a percentagem de horas de Sol em um intervalo de tempo (%);  $T_{med}$  é a temperatura média diária ( $\text{°C}$ ); e  $E_{elev}$  é a altura em relação ao nível do mar (m).

A variável  $b$  pode ser ajustada pela regressão como na Equação (3.10) a seguir

$$b = 0,81917 - UR_{min} + 1,0705 \frac{nh}{NH} + 0,065649u_2 - 0,0059684UR_{min} \frac{n}{N} \quad (3.10)$$

em que:  $u_2$  é a velocidade dos ventos a 2 metros de altura ( $\text{m}\cdot\text{s}^{-1}$ );  $UR_{min}$  é a umidade relativa mínima (%);  $nh$  é a duração de Sol real; e  $NH$  é a duração máxima de Sol (horas).

Na ausência das variáveis umidade relativa, radiação solar e velocidade dos ventos, pode ser aplicado o método de [Hargreaves e Samani \(1985\)](#). O Método de Hargreaves-Samani ([SAMANI, 2000](#)) é estimado pela Equação (3.11)

$$ET_o = 0,0135K_r \frac{R_a}{\lambda} (T_{max} - T_{min})^{0,5} (T_{cal} + 17,8) \quad (3.11)$$

em que:  $K_r$  é igual a  $0,00185(T_{max} - T_{min})^2 - 0,0433(T_{max} - T_{min}) + 0,4023$ ;  $T_{cal}$  é a temperatura média calculada  $(T_{max} + T_{min}) \cdot 2^{-1}$ ; e  $\lambda$  é a constante do calor latente da vaporização (2,45 a 20°C). Esse método pode estimar a  $ET_o$  diária, mensal ou anual.

Na ausência da variável climática velocidade dos ventos foi aplicada uma velocidade constante de  $2 \text{ m}\cdot\text{s}^{-1}$ . Essa estimativa foi determinada através da média de 2.000 EMs no mundo ([ALLEN \*et al.\*, 1998](#)). A radiação solar [Equação (3.12)] pode ser determinada por [Hargreaves e Samani \(1985\)](#)

$$R_s = K_r \cdot (T_{max} - T_{min})^{0,5} \cdot R_a \quad (3.12)$$

em que:  $R_s$  é a radiação solar ( $\text{MJ}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$ );  $K_r$  é o coeficiente de ajuste; e  $R_a$  é a radiação extraterrestre ( $\text{MJ}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$ ). A pressão de vapor  $e^o(T)$  em (KPa) pode ser determinada por

$$e^o(T_{med}) = 0,6108 \exp\left(\frac{17,27T_{med}}{T_{med} + 237,3}\right) \quad (3.13)$$

em que:  $T$  é a temperatura (°C). A pressão de vapor saturada pode ser determinada por  $e_s = (e^o(T_{max}) - e^o(T_{min})) \cdot 2^{-1}$ .

### 3.5 Ajuste da produtividade de grãos por séries temporais

As séries temporais fazem a modelagem de uma variável  $\mathbf{y}_i^c$  ao longo do tempo, em que  $i = 1, \dots, n$ , onde  $n$  é o número de observações. Supondo o modelo  $\mathbf{y}_i^c = f(\mathbf{y}_{i-1}, \mathbf{y}_{i-2}, \dots, \epsilon)$  em função do tempo, então, deve ser especificada a forma funcional do modelo  $f(\cdot)$ .

Na aplicação dos modelos de Box e Jenkins, onde são integrados os modelos de média móvel e autoregressivo, faz necessário verificar se a série temporal estudada é

não-estacionária, podendo apresentar tendência ou sazonalidade (MORETTIN; TOLOI, 2011). A série temporal pode ser decomposta como apresentado a seguir

$$\mathbf{y}_i^c = \mathbf{td}_i^c + \mathbf{s}_i^c + \mathbf{a}_i^c \quad (3.14)$$

em que:  $\mathbf{td}_i^c$  são as tendências ao longo do período;  $\mathbf{s}_i^c$  são as sazonalidades ocorrida no período; e  $\mathbf{a}_i^c$  são ruídos aleatórios ou ruídos brancos, que possuem oscilação aleatória com média zero e variância constante  $\sim N(0, \sigma^2)$ .

É possível aplicar os modelos de Box e Jenkins quando um processo apresenta característica estacionária (estocástico com propriedade probabilística), ou seja, permanece com média  $\mu$  e variância  $\sigma^2$  constante ao longo do tempo (BOX *et al.*, 2008), sendo necessário outros modelos. Entretanto, a produtividade da soja e do milho são não-estacionárias.

Quando a série temporal não apresenta sazonalidade e o seu comportamento é não-estacionário, pode ser aplicado o modelo ARIMA (p,d,q) (MONTGOMERY *et al.*, 2015). Essas especificações representam os parâmetros, onde:  $p$  é o número de termos autoregressivo;  $d$  é a quantidade de diferenciações; e  $q$  é o número de média-móvel. Assim, o modelo pode ser representado por

$$\phi(L)(1-L)^d \mathbf{y}_i^c = \delta + \theta(L) \mathbf{a}_i^c \quad (3.15)$$

em que:  $\phi(L)$  é igual ao polinômio autoregressivo de ordem ( $p$ ) ( $1 - \phi_1 L^1 - \dots - \phi_p L^p$ );  $\theta(L)$  é igual ao polinômio de média-móvel de ordem ( $q$ ) ( $1 - \theta_1 L^1 - \dots - \theta_q L^q$ );  $d$  é a quantidade de diferenciações necessárias para tornar a série estacionaria;  $\delta$  é a média  $\mu(1 - \phi_1 - \dots - \phi_p)$ ;  $L$  é o operador de retardo ( $L^d \mathbf{y}_i^c = \mathbf{y}_{i-d}^c$ ); e  $(1-L) \mathbf{y}_i^c = \mathbf{y}_i^c - \mathbf{y}_{i-1}^c$ , onde há defasagens o vetor  $\mathbf{y}_i^c$  é ajustado para  $i = 1, \dots, n - d$ .

Quando a série temporal apresenta um comportamento sazonal em determinado período, faz-se necessário acrescentar esse componente no modelo. A distância entre os intervalos sazonais pode ser medida por  $(1 - L^d) \mathbf{y}_i^c = \mathbf{y}_i^c - \mathbf{y}_{i-d}^c$ . Determinado o intervalo sazonal, o modelo SARIMA (p,d,q)(P,D,Q)s pode ser equacionado por

$$\phi_p(L) \Phi_P(L^s) (1-L)^d (1-L^s)^D \mathbf{y}_i^c = \delta + \theta_q(L) \Theta_Q(L^s) \mathbf{a}_i^c \quad (3.16)$$

em que:  $\Phi_P(L^s)$  e  $\Theta_Q(L^s)$  são polinômios sazonais autoregressivo e média móvel de ordem (P) e (Q), nessa ordem, e D é a quantidade necessária de *lags* para retirar a sazonalidade.

Para ajuste do modelo são necessárias algumas etapas, como: verificar a necessidade de fazer a transformação nos dados originais para reduzir sua variância (MORETTIN; TOLOI, 2011); determinar a quantidade de diferenciações; estimar os parâmetros desconhecidos; testar o ajuste dos resíduos; e avaliar a previsão do modelo (TSENG; TZENG, 2002).

Para identificar o parâmetro  $d$ , são feitas as diferenciações  $(1 - L)^d$ , onde  $d = 0, 1, \dots$ . As diferenciações são realizadas até que a série seja estacionária. A característica estacionária é confirmada quando a série não possui raiz unitária, para isso, pode ser aplicado o teste de Dickey e Fuller (1979). Se uma série temporal  $\mathbf{y}_i^c$  converge para  $i = \infty$ , então ela será estacionária para  $|\phi| < 1$ . Logo, o teste consiste em:

$$\begin{aligned} \mathbf{y}_i^c - \mathbf{y}_{i-1}^c &= \phi \mathbf{y}_{i-1}^c - \mathbf{y}_{i-1}^c + \mathbf{a}_i^c \\ (1 - L)\mathbf{y}_i^c &= (\phi - 1)\mathbf{y}_{i-1}^c + \mathbf{a}_i^c \\ (1 - L)\mathbf{y}_i^c &= \pi \mathbf{y}_{i-1}^c + \mathbf{a}_i^c \end{aligned} \quad (3.17)$$

em que:  $\pi$  é o valor crítico, onde  $\pi = (\phi - 1)$ . Se  $|\phi| = 1$ , então  $\pi$  é igual a zero, confirmando a presença de raiz unitária. O teste verifica a veracidade da hipótese  $H_o : \pi = 0$ , *versus*  $H_a : \pi < 0$ .

Entretanto, esse teste não verifica se os resíduos são autocorrelacionado, podendo ser ampliado (DICKEY; FULLER, 1981), incluindo defasagem  $(1 - L)\mathbf{y}_{i-l}^c$  na Equação (3.17), onde  $l$  varia de zero até que o Critério de Informação de Akaike (AIC) seja minimizado. Também, pode ser estendido a séries com *drift* (nesse caso é o intercepto) e tendência. As diferenciações são realizadas até que a série não possua mais raiz unitária (ver Pseudocódigo 1).

---

### Pseudocódigo 1 Série estacionária. Teste sequencial de Dickey-Fuller

---

Passo 1: Estimado a regressão inicial,  $\pi = 0?$ ;

Passo 2: Com a presença de tendência incluída, o coeficiente da regressão é zero dado  $\pi = 0?$ ;

Passo 3: Com a presença de *drift* incluída, o coeficiente da regressão é zero dado  $\pi = 0?$ ;

Passo 4: Sendo os Passo 1 até o 3 diferente de zero, então a série não possui raiz unitária.

---

Para identificar quantas diferenciações  $d$ , também podem ser utilizados os coeficientes da função de autocorrelação amostral (FAC) e a função de autocorrelação amostral parcial (FACP). Os valores precisam cair exponencialmente e estabilizar entre os limites do intervalo de confiança (IC). O IC é definido por  $t_\gamma/\sqrt{n}$ , onde  $t_\gamma$  é o valor da estatística  $t$  de Student com  $n - 1$  graus de liberdade (MONTGOMERY *et al.*, 2015; MORETTIN; TOLOI, 2011).

Os parâmetros de  $p$  e  $q$  podem ser determinados pelo comportamento da FACP e FAC, onde se espera o decaimento exponencial da FAC e trancamento da FACP, o que as vezes dificulta a análise. Porém, como alternativa, para identificar os parâmetros podem ser adotados o Critério de Informação de Akaike (AIC) e o Critério de Informação de Bayesian (BIC) como seleção. Os dois critérios penalizam a soma dos resíduos ao quadrado quando incluído um parâmetro no modelo (AKAIKE, 1974; SCHWARZ, 1978). Definidos por:

$$AIC = \log \left( 1/n \sum_{i=1}^n e_i^l e_i^c \right) + 2\xi \cdot n^{-1} \quad (3.18)$$

$$BIC = \log \left( 1/n \sum_{i=1}^n e_i^l e_i^c \right) + \xi \log(n) \cdot n^{-1} \quad (3.19)$$

$$e_i^c = \mathbf{y}_i^c - \left( \hat{\delta} + \sum_{j=1}^p \hat{\phi} \mathbf{y}_{i-j}^c - \sum_{j=1}^q \hat{\theta} e_{i-j}^c \right)$$

em que:  $\xi$  o total de parâmetros no modelo, onde  $\xi = p + q$ .

Dentre os vários modelos gerados, escolhe-se aquele que apresenta os menores valores de AIC e BIC. Vale salientar também, que a escolha deverá levar em consideração os modelos que apresentam o menor número de parâmetros. Por fim, é recomendável que os critérios AIC e BIC sejam avaliados conjuntamente, pois são complementares e não excludentes. Isso ocorre porque o BIC penaliza um pouco mais pelo tamanho da amostra ( $\xi \log(n)$ ) e o AIC pelo número de parâmetros ( $2\xi$ ).

### 3.6 Ajuste da produtividade de grãos pela regressão PLS

O método de Mínimos Quadrados Parciais (do inglês *Partial Least Square Regression* – PLSR) consiste em um modelo preditivo que pode ser aplicado quando as variáveis são colineares, na existência de mais de uma variável de resposta e quando o número de variáveis é maior que as observações. Consiste na redução das dimensões em uma situação

que queremos relacionar um conjunto de variáveis resposta a um conjunto de variáveis preditoras (PÉREZ-ENCISO; TENENHAUS, 2003).

O PLS foi criado por Herman Ole Andreas Wold (1908-1992). Durante sua vida, ele trabalhou com muitos modelos, mas sua preferência era o método dos mínimos quadrados. Em 1964, após uma conferência sobre o método *Fixed-Point* (de sua autoria) na Universidade da Carolina do Norte (WOLD, 2001), Wold decidiu estendê-lo a técnicas multivariadas. Então, Herman Ole A. Wold dá esta técnica como encerrada no final da década de 80 (SVANTE WOLD, 2015), quando apresenta o algoritmo *Non-linear Iterative Partial Least Squares* (NIPALS) que calcula componentes principais e correlação canônica com o método dos mínimos quadrados ordinários (MATEOS-APARICIO, 2011).

Para Krishnan *et al.* (2011) e Wold (2015) o método PLS originou várias técnicas, entre elas: uma análise que verifica a associação de correlação entre dois conjuntos de dados (*Partial Least Squares Correlation – PLSC*); modelagem de equações estruturais (*Partial Least Squares Path Modeling – PLS-PM*); e regressão que utiliza variáveis latentes para estimar um conjunto de dados (*Partial Least Squares Regression – PLSR*).

O PLSR é um método que relaciona duas matrizes  $\mathbf{X}$  e  $\mathbf{Y}$  por meio de um modelo multivariado (YUN *et al.*, 2014), onde a relação pode ser expressa por  $\mathbf{Y} = f(\mathbf{X})$ . Sendo  $\mathbf{X}$  e  $\mathbf{Y}$  duas matrizes de ordem  $n \times k$  e  $n \times m$  nessa ordem, onde:  $n$  é o número de observações ( $i = 1, \dots, n$ );  $k$  é número de variáveis em  $\mathbf{X}$  ( $j = 1, \dots, k$ ); e  $m$  é o número de variáveis em  $\mathbf{Y}$  ( $j = 1, \dots, m$ ).

Independente da resposta, foi testada a amostra para verificar a distribuição dos dados usando o teste de Kolmogorov-Smirnov para as distribuições: normal; log-normal; exponencial; weibull; gamma; beta; uniforme; e logística. Os resultados indicam que os dados podem ser normais, mais detalhes veja Merelles *et al.* (2017) e para complementar os resultados desse artigo com os dados das regiões estudadas veja o Apêndice B.

Antes de aplicar a regressão PLS, com frequência, as variáveis de  $\mathbf{X}$  e  $\mathbf{Y}$  são transformadas (centradas e/ou normalizadas). Essa transformação ocorre porque o PLS absorve a diferença na escala dos dados. Quando os dados são medidos na mesma unidade não são necessárias transformações. Entretanto, quando aplicado em dados com unidades de medidas diferentes (exemplo: chuva em milímetros, temperatura em °C) é necessário centrar na média e normalizar os dados (GELADI; KOWALSKI, 1986). Também, nessa

transformação, quando é conhecida a variável  $\mathbf{x}_i^c$  mais importante para o modelo, ela pode ser multiplicada por um fator para aumentar a sua importância.

Com objetivo de modelar as variáveis em  $\mathbf{Y}$ , faz-se necessário encontrar algumas variáveis latentes ortogonais ( $H =$  número de componentes no modelo) que são os *scores* de  $\mathbf{X}$  e  $\mathbf{Y}$  (PÉREZ-ENCISO; TENENHAUS, 2003; WOLD *et al.*, 2001), representadas por  $\mathbf{T}$  e  $\mathbf{U}$  (ordem  $n \times h$ ), respectivamente. De acordo com Rosipal e Krämer (2006), no NIPALS, são determinados os pesos  $\mathbf{w}_h^l$  dado que a covariância é maximizada entre as variáveis latentes  $\left(\text{cov}\left(\mathbf{t}_h^c, \mathbf{u}_h^c\right)^2 \equiv \left(1/n \sum \mathbf{t}_h^l \mathbf{u}_h^c\right)^2\right)$ .

---

### Pseudocódigo 2 Passos para modelar a relação entre $\mathbf{X}$ e $\mathbf{Y}$ com PLS

---

- Passo 1: Pré-tratamento dos dados;
- Passo 2: Calcular os vetores de peso (*weight*):  $\mathbf{w}_h^l$ ;
- Passo 3: Calcular os vetores de pontuação (*scores*):  $\mathbf{t}_h^c, \mathbf{u}_h^c$ ;
- Passo 4: Calcular os valores das cargas (*loading*):  $\mathbf{p}_h^l, \mathbf{c}_h^l$ ;
- Passo 5: Determinar as matrizes de deflação:  $\mathbf{E}_h, \mathbf{F}_h$ ;
- Passo 6: Determinar um ponto de parada.

Nota: mais detalhes em Geladi e Kowalski (1986) e Wold *et al.* (2001)

---

No PLS, durante a extração de cada componente são realizadas algumas interações para certificar a convergência das variáveis latentes. Durante essa etapa, os *scores* ( $\mathbf{t}_h^c$  e  $\mathbf{u}_h^c$ ) são atualizados até que seja confirmado um pequeno erro  $\epsilon$  entre a interação atual e a anterior ( $\epsilon = 10^{-6}$  ou  $10^{-8}$ ). Os passos para extrair as variáveis latentes estão no Pseudocódigo 2. Nele é extraído somente a primeira componente  $h$ , para as demais componentes, repita o passo dois até que o passo seis seja satisfatório. Em cada componente é realizada a deflação (GODOY *et al.*, 2014) nas matrizes  $\mathbf{E}_h$  e  $\mathbf{F}_h$ , onde  $\mathbf{E}_0$  é igual a  $\mathbf{X}$  e  $\mathbf{F}_0$  é igual a  $\mathbf{Y}$ . Assim, as matrizes de deflações podem ser atualizadas por

$$\mathbf{E}_h = \mathbf{E}_{h-1} - \mathbf{t}_h^c \mathbf{p}_h^l \quad (3.20)$$

$$\mathbf{F}_h = \mathbf{F}_{h-1} - \mathbf{t}_h^c \mathbf{c}_h^l \quad (3.21)$$

Devido as deflações sucessivas durante as extrações das componentes, pode ocorrer uma complicação nos pesos  $\mathbf{W}'$ , sendo necessária sua atualização (WEHRENS; MEVIK, 2007). Essa atualização, também, aumenta a relação entre as matrizes  $\mathbf{X}$  e  $\mathbf{Y}$ , e pode ser escrita como  $\mathbf{R} = \mathbf{W}(\mathbf{P}'\mathbf{W})^{-1}$ , onde  $\mathbf{R}'$ ,  $\mathbf{W}'$  e  $\mathbf{P}'$  são de ordem  $h \times k$ .

O PLS converge na primeira interação se o número de variáveis  $m = 1$ . Caso contrário ( $m > 1$ ), também é possível fazer a convergência na primeira interação (WOLD



*et al.*, 2001). Isso ocorre porque o primeiro vetor de pesos  $\mathbf{w}_1^l$  corresponde ao primeiro autovetor da combinação da matriz de variância-covariância  $\mathbf{S}_1 = \mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}$ . Para as demais  $h$  componentes, os pesos  $\mathbf{w}_h^l$  podem ser determinados fazendo as deflações na matriz de variância-covariância

$$\begin{cases} \mathbf{S}_2 = \mathbf{S}_1 - \mathbf{t}_1^c \mathbf{p}_1^l \\ \mathbf{S}_h = \mathbf{S}_{h-1} - \mathbf{t}_{h-1}^c \mathbf{p}_{h-1}^l \end{cases} \quad (3.22)$$

Para o primeiro vetor de *scores*  $\mathbf{t}_1^c$ , de forma similar, corresponde ao primeiro autovetor de  $\mathbf{X}\mathbf{X}'\mathbf{Y}\mathbf{Y}'$ . Para as demais  $h$  componentes de  $\mathbf{t}_h^c$ , é aplicado o primeiro autovetor de  $\mathbf{S}_h \mathbf{S}_h' \mathbf{Y}\mathbf{Y}'$ .

Na predição, no PLSR é suposto um ponto de parada para  $h$ -ésima componente, onde se obtém as componentes principais ( $\mathbf{T} = \mathbf{t}_1^c, \dots, \mathbf{t}_H^c$ ), assim, o modelo pode ser expresso por

$$\hat{\mathbf{Y}} = \mathbf{T}\mathbf{C}' = \mathbf{X}\mathbf{R}\mathbf{C}' = \mathbf{X}\mathbf{B} \quad (3.23)$$

em que:  $\mathbf{B}$  é uma matriz (ordem  $k \times m$ ) de coeficiente da regressão PLS, onde  $\mathbf{B} = \mathbf{R}\mathbf{C}'$  com  $h$  componentes. Então, finalmente, para modelar a produção de grãos  $\mathbf{Y}$  sob as mudanças climáticas  $\mathbf{Z}$ , é retirada a transformação dos dados [Equação (3.2)] e atualizada a modelagem

$$\hat{\mathbf{Y}}^r = \begin{bmatrix} \alpha_1^l \\ \vdots \\ \alpha_n^l \end{bmatrix}_{ij}^r + \sum_{l=1}^k z_{il}^r \psi_{lj}^r + \epsilon_{ij}^r = \begin{bmatrix} \alpha_1^l \\ \vdots \\ \alpha_n^l \end{bmatrix}^r + \mathbf{Z}^r \mathbf{\Psi}^r + \mathbf{E}^r \quad (3.24)$$

em que:  $\alpha_m^l$  é o vetor de intercepto, onde  $\alpha_m^l = \bar{y}_j - \sum_{j=1}^k \bar{z}_{1 \times j} \Psi_{j \times m}^r$ ;  $\mathbf{\Psi}^r$  é a matriz de coeficientes de  $\mathbf{Z}$  sem transformação, onde  $\mathbf{\Psi}^r = \mathbf{A}_{s_z}^d \mathbf{B} \mathbf{A}_{s_y}^d$  (ordem  $k \times m$ );  $\mathbf{A}_{s_z}^d = 1/s_j^z$  se  $i = j$  e 0 se  $i \neq j$  e  $\mathbf{A}_{s_y}^d = s_j^y$  se  $i = j$  e 0 se  $i \neq j$  são as matrizes diagonais com os vetores de desvios  $s_j^z$  e  $s_j^y$  na diagonal principal.

Os dados foram divididos em dois conjuntos, treinamento (1980-2011) e teste (2012-2016). A validação-cruzada (CV) foi aplicada para determinar o melhor modelo, onde foram excluídos os seguimentos do conjunto de treinamento. Nos seguimentos foram

incluídas as combinações de 2, 3 e 4 datas por vez (foram testados 41.416 modelos), o melhor modelo foi selecionado pelo critério  $Q_{cum}^2$  [Equação 3.26]. Também, foi aplicada a CV para corrigir o critério de parada das componentes. Assim, o conjunto de treinamento foi dividido em 10 grupos aleatórios, então são deletados cada grupo, um a um, e modelado novamente.

O número de componentes ótimas é um critério que deve ser determinado e para isso existem várias técnicas (HÖSKULDSSON, 1996). Neste trabalho foi utilizado o índice  $Q_h^2$  que mede a contribuição de cada componente. No dimensionamento, são deletadas observações e modelado novamente, durante isso, a soma dos quadrados residuais previsto (PRESS) e a soma dos quadrados residuais (RSS) são calculadas. Então, a  $h$ -ésima componente é selecionada se satisfazer o critério

$$Q_h^2 = 1 - \left( \frac{\sum_{j=1}^m PRESS_{jh}}{\sum_{j=1}^m RSS_{j(h-1)}} \right) \quad (3.25)$$

em que:  $h$  está variando de um até  $Q_h^2 \geq (1 - 0,95^2)$ , para todas variáveis de resposta  $m$ , ou o número de componente  $h$  ser igual ao mínimo entre  $n$  e  $k$ . Também, a contribuição global das componentes selecionadas pode ser medida por

$$(Q_{cum}^2)_h = 1 - \prod_{h=1}^H \left( \frac{\sum_{j=1}^m PRESS_{jh}}{\sum_{j=1}^m RSS_{j(h-1)}} \right) \quad (3.26)$$

As variáveis resposta em  $\mathbf{Y}$  podem ser consideradas bem modeladas pelas  $h$  componentes se  $Q_{cum}^2 \geq 1/2$ , que são limites arbitrários (TENENHAUS, 1998).

No PLS, o *score* com a influência da variável na projeção (**VIP**) faz a seleção de variáveis preditoras em  $\mathbf{X}$ , isso ocorre porque o **VIP** faz a relação da contribuição de cada variável em  $\mathbf{X}$  para construção dos pesos  $\mathbf{W}'$  e os *scores*  $\mathbf{T}$ .

No método **VIP**, a contribuição de uma variável preditora para construção do *score*  $\mathbf{t}_h^c$  é medida pelos pesos  $\mathbf{w}_h^l$ , onde para cada  $h$  componente, a soma dos pesos  $\sum (\mathbf{w}_h^l)^2$  de todas variáveis em  $\mathbf{X}$  é igual a um. A contribuição de uma variável preditora para construção das variáveis resposta é medida pela componente  $\mathbf{t}_h^c$  (PÉREZ-ENCISO; TENENHAUS, 2003) e o seu poder explicativo é medido por  $\mathbf{cm}_j^l = 1/m \sum_{i=1}^m c_{ij}$ , onde  $c_{ij} = \text{cor}(\mathbf{Y}, \mathbf{T})^2$ . A matriz de *score* **VIP** (ordem  $k \times h$ ), com as contribuições das variáveis preditoras na

construção de  $\mathbf{Y}$ , pode ser calculada por

$$\mathbf{VIP}_{ij} = \left\{ \sum_{j=1}^h w_{ij}^2 \left( \sum_{i=1}^h a_{ji}^{ts} a_{ij}^d \right) \right\}^{1/2} = \left( \mathbf{W}^2 \mathbf{A}_{vip}^{ts} \mathbf{A}_{vip}^d \right)^{1/2} \quad (3.27)$$

em que:  $\mathbf{A}_{vip}^d$  é uma matriz quadrada diagonal (ordem  $h \times h$ ), onde  $\mathbf{ry}_h^l$  se  $i = j$  e 0 se  $i \neq j$ ;  $\mathbf{ry}_h^l$  é um vetor de proporção, onde  $\mathbf{ry}_h^l = k / \sum_{j=1}^h \mathbf{cm}_j^l$ ; e  $\mathbf{A}_{vip}^{ts}$  é uma matriz triangular superior (ordem  $h \times h$ ) com o coeficiente de determinação ao quadrado, onde  $\mathbf{A}_{vip}^{ts} = [\mathbf{cm}_1^c, \dots, \mathbf{cm}_h^c]_{ij}$  se  $i \leq j$  e 0 se  $i > j$  (com esse procedimento todos os elementos abaixo da diagonal principal são iguais a zero). O **VIP**, também, tem por propriedade a soma dos quadrados igual ao número de variáveis preditoras em  $\mathbf{X}$  (TENENHAUS, 1998), para todas as componentes ( $\sum_{j=1}^n vip_{ij}^2 = k, \forall i = 1, \dots, h$ ).

Quantificada a influência sobre a resposta para cada variável preditora e todas as componentes em relação a soma dos quadrados totais do **VIP**, faz com que ele seja uma medida intuitiva com o efeito global das variáveis climáticas  $\mathbf{Z}$  na produção de grãos. Durante a seleção, na primeira análise foi realizada a modelagem com todas as variáveis preditoras, e os *scores* **VIP** fez a sua classificação. Posteriormente, somente as variáveis com **VIP** > 1 foram selecionadas e incluídas na segunda análise.

### 3.7 Métricas de avaliação dos ajustes

Existem alguns critérios que podem ser aplicados para avaliar a qualidade de ajuste do modelo (MORIASI *et al.*, 2007; KRAUSE *et al.*, 2005). Neste trabalho, o desempenho do modelo de previsão proposto foi avaliado aplicando as seguintes métricas: coeficiente de determinação ( $R^2$ ); média absoluta percentual dos erros (MAPE); coeficiente de eficiência Nash-Sutcliffe (NSE); percentual de viés (PBIAS); erro absoluto médio (MAE); e raiz do erro quadrático médio (RMSE).

Com frequência, o coeficiente de determinação ( $R^2$ ) é o mais aplicado (WALPOLE *et al.*, 2016). Essa métrica indica quanto o modelo é capaz de explicar a amostra. O  $R^2$  [Equação 3.28] assume valores de zero a um, sendo um, o melhor ajuste. O NSE [Equação 3.29] avalia a magnitude de variância dos resíduos em relação a variância da amostra (NASH; SUTCLIFFE, 1970). O NSE assume valores entre  $-\infty$  a 1. Geralmente, são aceitos valores de zero a um, caso contrário (valores < 0), o valor médio da observação

prediz melhor que os dados estimados.

$$R^2 = \left[ \frac{\sum \mathbf{y}_i \hat{\mathbf{y}}_i - \sum \mathbf{y}_i \sum \hat{\mathbf{y}}_i}{\sqrt{[n \sum \mathbf{y}_i^2 - (\sum \mathbf{y}_i)^2][n \sum \hat{\mathbf{y}}_i^2 - (\sum \hat{\mathbf{y}}_i)^2]}} \right]^2 \quad (3.28)$$

$$NSE = 1 - \left[ \frac{\sum_{i=1}^n (\mathbf{y}_i - \hat{\mathbf{y}}_i)^2}{\sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}}_i)^2} \right] \quad (3.29)$$

em que:  $\hat{\mathbf{y}}_i$  é um vetor com os valores previstos;  $\mathbf{y}_i$  é um vetor com os valores observados; e  $n$  é o número de amostras.

O PBIAS [Equação 3.30] mensura quanto os dados estimados estão acima ou abaixo da amostra em média (em porcentagem). PBIAS com valor positivo indica que o modelo possui viés de subestimar os dados e quando o valor é negativo, indica que o modelo possui viés de superestimar. Também, com unidade em porcentagem, o MAPE [Equação 3.31] mede a porcentagem de erro em média, podendo assumir valores maiores que zero. PBIAS e MAPE com o valor de zero é a situação ótima. Apesar do MAPE ser aplicado com frequência, em circunstâncias onde exista discrepância nos erros, pode ocasionar um resultado acima do valor real (GOODWIN; LAWTON, 1999).

$$PBIAS = 100 \left[ \frac{\sum_{i=1}^n (\mathbf{y}_i - \hat{\mathbf{y}}_i)}{\sum_{i=1}^n \mathbf{y}_i} \right] \quad (3.30)$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{\mathbf{y}_i - \hat{\mathbf{y}}_i}{\mathbf{y}_i} \right| \quad (3.31)$$

O MAE [Equação 3.32] mede a magnitude média dos erros sem considerar sua direção, onde todos os erros têm o mesmo peso. De forma similar, o RMSE [Equação 3.33] também mede a magnitude média dos erros, porém os erros são elevados ao quadrado. Isso faz com que modelos que possuam um elemento com erro alto seja penalizado. Ambos, possuem valores que podem variar de zero a  $\infty$ , e casos onde o erro é zero, o ajuste é ideal.

Entretanto, o MAE é a medida mais simples aplicada, além de ser adicionada a outras métricas (WILLMOTT; MATSUURA, 2005). De acordo com Chai e Draxler (2014), a presença de *outliers* é melhor descrita pelo RMSE, porém o MAE é mais apropriado para dados com distribuição normal.

$$MAE = \frac{1}{n} \sum_{i=1}^n \left| \mathbf{y}_i - \hat{\mathbf{y}}_i^* \right| \quad (3.32)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left| \mathbf{y}_i - \hat{\mathbf{y}}_i \right|^2} \quad (3.33)$$

## 4 RESULTADOS E DISCUSSÕES

Neste capítulo, são apresentados os resultados e discussões dos melhores métodos para completar dados faltantes e estimar a evapotranspiração. Também, é apresentada a performance dos modelos e a influência das mudanças climáticas na produtividade de grãos.

### 4.1 Dados climáticos

Antes de preencher os dados com informações dos vizinhos, primeiro é necessário determinar a distância das EMs que serão aceitas (VICENTE-SERRANO *et al.*, 2010). Para estabelecer essa distância foi determinado o número médio de vizinhos em determinadas distâncias. O *boxplot* com o número de vizinhos por distância é apresentado na Figura 8.

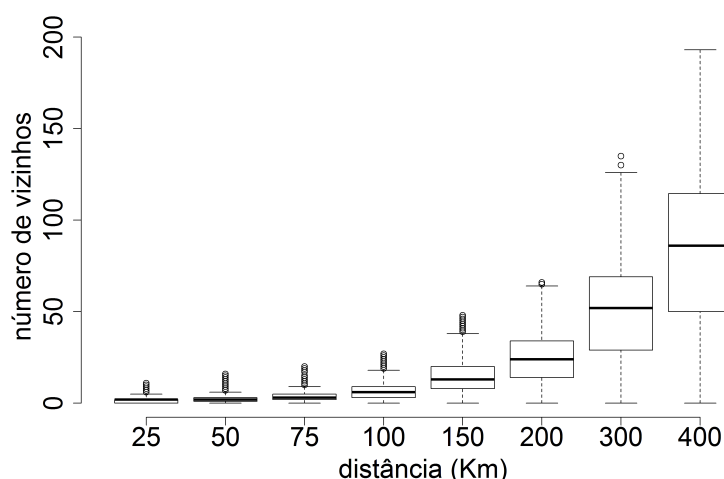


Figura 8 – Número de vizinhos com as distâncias equivalentes

Na distância de 400 Km, 50% das estações possuem mais de 86 EMs vizinhas e no primeiro quartil são 50 EMs. No percentil 1,5% existem duas estações com zero vizinhos, elas foram excluídas porque uma fica na Ilha da Trindade no oceano Atlântico e a outra é isolada na divisa do Pará com o Suriname. Assim, essa distância passou a ser considerada aceitável, por todas as estações possuírem vizinhos.

Para cada EM foi determinada a matriz de correlação com os vizinhos disponíveis até 400 Km. Correlações superiores a 0,5 foram consideradas satisfatórias, caso contrário, reduzia-se a distância. Os valores de correlação para cada variável climática são apresentados na Tabela 3. Como a distância mínima para velocidade dos ventos é de 5 Km, essa variável foi excluída por não ter vizinhos.

Tabela 3 – Distância máxima aceitável pelos vizinhos

Variável	R-Pearson	Km
Precipitação	0,509	115
Temperatura média	0,759	400
Temperatura máxima	0,728	400
Temperatura mínima	0,699	400
Umidade relativa	0,568	400
Velocidade dos ventos	0,545	5

Para determinar o melhor método de preenchimento, excluiu-se de forma aleatória 1, 5, 10, 15 e 20% dos dados rotulados por “x” e “y” com períodos em comum de no mínimo 3 anos. De acordo com a quantidade de dados faltantes, essa análise permite determinar qual método será aplicado. Após serem completados os dados excluídos, foi aplicada a raiz do erro quadrático médio (RMSE) para avaliar o melhor método.

Dos três métodos, na análise de 10% (ver Tabela 4), o método de interpolação proporcionou os melhores resultados para precipitação e umidade relativa do ar, com um RMSE médio de 8,49 mm e 7,5%, respectivamente. O método de regressão linear demonstrou excelente desempenho para a temperatura média, mínima e máxima, e a abordagem dos vizinhos mais próximos apresentou resultados aceitáveis, mas não foi o melhor.

Tabela 4 – Estatística RMSE da análise dos métodos para completar dados climáticos com 10% de dados

Variáveis	Interpolação	Regressão	Vizinho
Precipitação (mm)	<b>8,493</b>	8,785	10,421
Temperatura média (°C)	1,617	<b>1,369</b>	2,015
Temperatura máxima (°C)	1,794	<b>1,558</b>	2,199
Temperatura mínima (°C)	1,901	<b>1,766</b>	2,303
Umidade relativa (%)	<b>7,502</b>	8,168	9,344

No Sul, de acordo com [Bier e Ferraz \(2017\)](#), o erro médio é acima de 20mm. Provavelmente, os erros altos ocorreram devido aos dados não serem classificados em grupos, também pode ser devido a distância considerada entre as EMs ou alguns cuidados nos métodos, como: testar o método de substituição pelo vizinho mais próximo; e o intercepto da regressão.

Como o RMSE penaliza somente os altos erros, a decisão de qual método utilizar não

pode ser aplicada somente com sua análise, porque a modelagem pode alterar os valores dos extremos (temperatura) e a frequência de dias secos e úmidos (VICENTE-SERRANO *et al.*, 2010). Assim, aplicou-se ANOVA e o teste de Tukey para medir a diferença entre o que ocorreu e o estimado pelos métodos de preenchimento com informação dos vizinhos mais próximos (ver Apêndice A).

A determinação do mesmo método de preenchimento de falhas para diferentes variáveis pode levar a redução da explicação de uma série estimada. Exemplo: para a temperatura máxima, a escolha dos vizinhos poderia alterar a explicação dos valores extremos; para a precipitação, se aplicada a interpolação, a frequência de dias úmidos e secos seria alterada.

Os testes com 1 e 5% apresentaram mais de um método para preenchimento (interpolação/regressão) nas variáveis temperatura média e máxima, mas as análises com 10, 15 e 20% foram determinantes para escolha do método. Entretanto, ao contrário da análise RMSE, para completar os dados com falhas foram aplicados os três métodos. Na precipitação foi aplicado o método dos vizinhos mais próximos; na temperatura máxima foi utilizada a interpolação por distância inversa; e na umidade relativa do ar, temperatura média e mínima foi aplicado o método de regressão linear. A síntese dos melhores métodos para cada variável está na Tabela 5.

Tabela 5 – Melhores métodos para completar os dados faltantes

Descrição (percentual faltante)	1%	5%	10%	15%	20%
Precipitação	vizinho	vizinho	vizinho	vizinho	vizinho
Temperatura média	regressão	regressão/ interpolação	regressão	regressão	regressão
Temperatura máxima	interpolação/ regressão	interpolação	interpolação	interpolação	interpolação
Temperatura mínima	regressão	regressão	regressão	regressão	regressão
Umidade relativa	regressão	regressão	regressão	regressão	regressão

A variação da percentagem de exclusão foi importante na escolha do melhor método. Se fossem consideradas pequenas exclusões, um método inadequado poderia ser selecionado. Por exemplo, Vicente-Serrano *et al.* (2010) analisaram a precipitação excluindo apenas 1% das observações, já determinando que o melhor método seria o método dos vizinhos. Se essa mesma análise fosse estendida para temperatura média e máxima, a determinação do melhor método poderia ser equivocada.



Validados os melhores métodos, os dados rotulados por “x” foram completados utilizando os dados dos três grupos. Completou-se a série “y” com os dados de “y”, “z” e “x” atualizados. Então, foram completadas as séries rotuladas por “z”, com a mesma lógica. Após completar os dados, foi observado que algumas EMs ainda possuíam falhas. A quantidade de dados faltantes após aplicar o método de preenchimento está na Tabela 6. Nota-se que na temperatura média ( $T_{med}$ ) 233 EMs possuem falhas entre 0 e 91 observações, não consecutivas.

Tabela 6 – Análise do número de estações meteorológicas com dados faltantes após as séries serem completadas com informações dos vizinhos mais próximos

Observações*	$P_{ad}$	$T_{max}$	$T_{min}$	$T_{med}$	$UR_{med}$	$UR_{max}$
0-91	2	52	49	233	23	25
91-182	17	133	131	2	127	129
182-274	7	12	9	3	9	6
274-365	4	6	5	13	2	7
365-547	24	19	19	3	32	26
547-730	16	6	7	1	7	7
730-1.825	54	23	30	1	31	31
1.825-3.650	67	10	11	7	16	16
3.650-5.475	57	2	2	0	11	12
5.475-7.300	7	0	0	0	3	2
7.300-9.125	5	0	0	0	2	2
9.125-10.950	3	0	0	0	0	0

Nota:  $P_{ad}$  é a precipitação acumulada;  $T_{med}$  é a temperatura média;  $T_{max}$  é a temperatura máxima;  $T_{min}$  é a temperatura mínima;  $UR_{med}$  é a umidade média;  $UR_{max}$  é a umidade máxima; e  $U_z$  é a velocidade dos ventos.\* dados em dias.

Para reduzir o número de falhas, o período de estudo foi ajustado de 1980 a 2016 (37 anos), período com falhas inferiores a 14 dias (EISCHEID *et al.*, 1995). Essas falhas foram completadas pelo método de média móvel. Para a temperatura média, todas as falhas estavam completas.

Nos dados preenchidos, durante a correção dos dados, foram retirados 10.128 *outliers* acima de três desvio padrão (99,7%) dos 24.881.115 registros observados. Também, foram retidos 95 erros de digitação. Os erros de digitação foram da seguinte forma: temperatura máxima de 2°C (05 de fevereiro de 2010 em Esperantina-PI); temperatura mínima de 36,5°C (23 de fevereiro de 1986 em Manicoré-AM); temperatura média de 0°C (02 de novembro de 1996 em Irecê-BA); umidade mínima de 212% (07 de março de 1998 em

Uruguaiana-RS); umidade média de 0% (31 de maio de 1993 em Paranaguá-PR); e umidade máxima de 11,2% (21 de outubro de 2016 em Iauaretê-AM). Os *outliers* e os erros de digitação foram substituídos pela informação dos vizinhos mais próximos, com método apropriado para cada variável, desde que os dados dos vizinhos fossem adequados. Na precipitação os valores abaixo de 1% percentil como 0,001 mm foram considerados como zero e os dados acima de 99% percentis foram substituídos pelo vizinho.

## 4.2 Evapotranspiração

Uma das dificuldades no cálculo da evapotranspiração é os dados faltantes. Assim, o cálculo de  $ET_0$  foi realizado para determinar o melhor método na ausência de algumas variáveis climáticas (velocidade dos ventos e umidade relativa). Para determinação do melhor método foi aplicado o apresentado no boletim do FAO 56 (Penman-Monteith) com todas as variáveis e, posteriormente, a velocidade dos ventos e umidade relativa foram estimadas, entre os anos de 1980 a 2016. Depois foi calculada a  $ET_0$  utilizando as variáveis estimadas com os métodos de Penman-Monteith, Hargreaves-Samani e Blaney-Criddle (ALENCAR *et al.*, 2015).

Foi utilizado o NSE como base de avaliação do melhor modelo, pois o coeficiente é sensível nos períodos em que a evapotranspiração está baixa e reage bem nos períodos de evapotranspiração máxima (KRAUSE *et al.*, 2005). Isso ocorre por ele determinar a magnitude relativa entre a variância residual e a variância dos dados observados (MORIASI *et al.*, 2007).

Considerando o NSE, o cálculo da  $ET_0$  pelo método de Penman-Monteith sem a umidade relativa é a melhor estimativa. Na ausência dos ventos, o cálculo da  $ET_0$  pelo método de Penman-Monteith não é o mais adequado, sendo melhor o método de Hargreaves-Samani. Os dados das métricas estão na Tabela 7.

Para Alencar *et al.* (2015) o método de Penman-Monteith é melhor com a ausência dos ventos, porém não levaram em consideração o coeficiente  $K_r$  [Equação (3.11)] no método de Hargreaves-Samani. O coeficiente  $K_r$  é ajustado pela amplitude da temperatura, nesse caso corresponde ao ajuste da posição geográfica e a altitude da região (SAMANI, 2000).

Tabela 7 – Métricas para estimar o melhor método de evapotranspiração de referência

Métrica	Penman-Monteith sem umidade	Penman-Monteith sem ventos	Hargreaves Samani	Blaney Criddle
RMSE (mm)	1,029±0,18	1,215±0,17	1,364±0,45	5,381±1,04
MAE (mm)	0,863±0,16	1,103±0,20	1,045±0,27	5,074±1,02
MAPE (%)	20,384±3,81	22,463±3,70	27,356±8,50	119,583±24,07
RSQ (%)	54,085±22,16	62,252±24,28	30,768±20,37	70,764±13,05
PBIAS (%)	10,686±8,48	25,014± 9,07	-1,566±13,45	-53,145±4,74
NSE (índice)	0,298±0,31	-0,616±1,36	-0,320±1,06	-18,269±8,92

### 4.3 Ajuste de produtividade de grãos

A produção mundial de grãos (soja e milho) vem aumentando, em 2017 é esperada a produção de 4.910 milhões de toneladas (MT) com a demanda de 4.844 MT de grãos (MAFF, 2017). Entretanto, a produtividade ( $\text{kg}\cdot\text{ha}^{-1}$ ) de grãos vem sofrendo alteração no mundo desde a década de 80 (LOBELL *et al.*, 2011). Na China, Sul da Ásia e no Norte da Europa foi observada estagnação na produtividade do arroz e trigo a partir do ano 2000 (BAI *et al.*, 2016). No Brasil, é esperada a produção de 226 MT de grãos, no ano de 2017 (CONAB), também foi observado um comportamento ascendente da produtividade a partir de 1980 (ver Figura 9).

Tabela 8 – Medidas descritivas da produtividade de Soja e Milho na região Centro-Oeste e Sul do Brasil

Medidas descritivas	Centro-Oeste		Sul	
	Soja	Milho	Soja	Milho
Valor mínimo (tonelada)	1.657,00	1.813,00	1.091,00	1.567,00
Primeiro quartil (tonelada)	2.027,00	2.726,00	1.771,00	2.446,00
Média (tonelada)	2.469,00	4.347,00	2.143,00	3.774,00
Mediana (tonelada)	2.548,00	4.204,00	2.086,00	3.311,00
Terceiro quartil (tonelada)	2.940,00	5.419,00	2.519,00	4.758,00
Valor máximo (tonelada)	3.137,00	7.697,00	3.124,00	7.412,00
Desvio padrão (tonelada)	481,90	1.885,90	545,62	1678,24
Coefficiente de variação (%)	19,50	43,40	25,46	44,46
Assimetria	-0,26	0,38	0,18	0,74
Curtose	-1,45	-1,05	-0,83	-0,64

Para melhor compreender a produtividade de grãos, inicialmente, foram determinadas suas medidas de dispersão. A análise corresponde a produção anual na região Centro-Oeste e Sul do Brasil durante o período de 1980 a 2016 (ver Tabela 8). A produtividade de soja

e milho em média é de 15% maior na região Centro-Oeste. O autovalor para o desvio padrão e o coeficiente de variação nas séries indicam que a média é pouco representativa. A produtividade máxima de soja ocorreu em 2011 e a mínima foi em 1980, para o milho a produtividade máxima ocorreu em 2015 e a mínima em 1986.

A produtividade de soja e milho são correlacionadas, no Centro-Oeste a relação é de 0,922 ( $p$ -valor =  $2,1 \cdot 10^{-15}$ ) e no Sul de 0,915 ( $p$ -valor =  $4,7 \cdot 10^{-15}$ ). As correlações entre as variáveis preditoras e as variáveis resposta estão na Tabela 9. Observa-se que os níveis de  $\text{CO}_2$  possuem correlação positiva alta, isso já era esperado devido ao metabolismo das plantas ser acelerado por essa variável (HAO *et al.*, 2014; TWINE *et al.*, 2013). No geral, a umidade relativa do ar possui correlação negativa com a produtividade de grãos, ou seja, cenários onde a umidade é alta a produtividade de grãos é reduzida. A temperatura mínima pode reduzir a produtividade de grãos (JONES *et al.*, 2017a), entretanto, a temperatura máxima pode incrementar a produtividade de grãos. A perda de produtividade quando a umidade e a temperatura mínima do ar aumentam pode estar associada ao aumento de doenças nas plantas causadas principalmente por fungos (JUNK *et al.*, 2016; PATERSON *et al.*, 2014). De modo geral, a variabilidade climática pode alterar a produtividade de grãos (JU *et al.*, 2013).

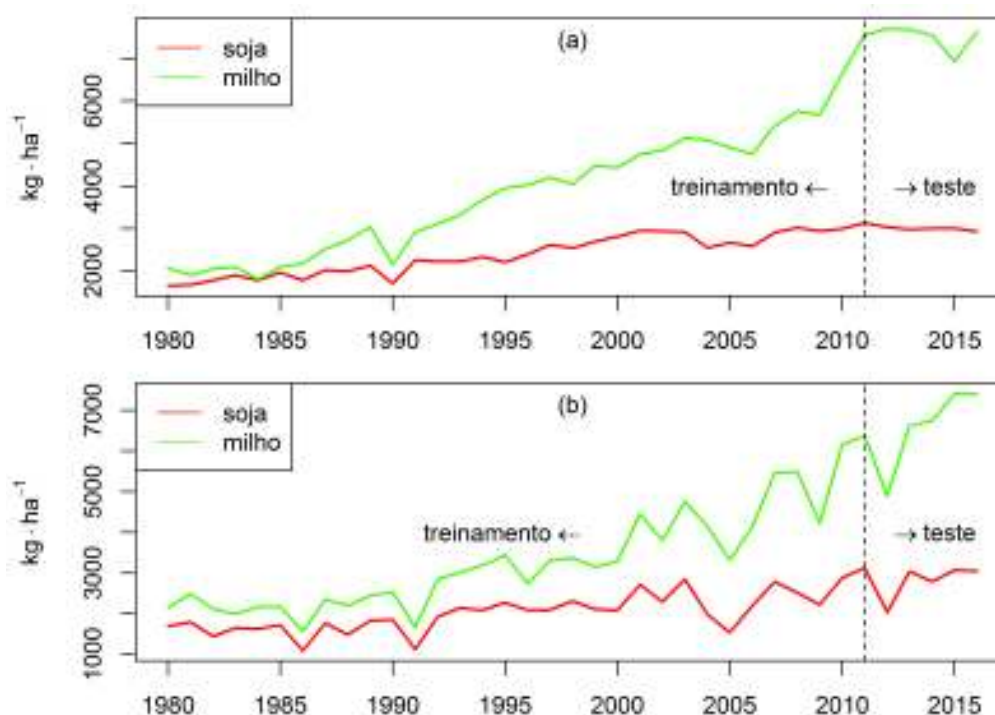


Figura 9 – Série histórica da produtividade de grãos de 1980 a 2016. (a) produtividade da região Centro-Oeste. (b) produtividade da região Sul

Tabela 9 – Correlação entre as variáveis preditoras e resposta

Variáveis	Centro-Oeste		Sul		Variáveis	Centro-Oeste		Sul	
	Soja	Milho	Soja	Milho		Soja	Milho	Soja	Milho
$P_{ad}:1$	<b>0,479</b>	<b>0,478</b>	0,271	0,090	$UR_{med}:3$	<b>-0,674</b>	<b>-0,686</b>	0,078	0,050
$P_{ad}:2$	<b>-0,214</b>	<b>-0,046</b>	<b>-0,090</b>	<b>-0,025</b>	$UR_{med}:4$	<b>-0,494</b>	<b>-0,483</b>	0,100	0,119
$P_{ad}:3$	<b>-0,009</b>	<b>-0,025</b>	<b>0,056</b>	<b>0,005</b>	$UR_{min}:1$	-0,839	-0,856	-0,323	-0,426
$P_{ad}:4$	-0,852	-0,777	0,415	0,470	$UR_{min}:2$	-0,867	-0,807	-0,441	-0,473
$T_{med}:1$	<b>0,482</b>	0,518	0,371	0,420	$UR_{min}:3$	-0,786	-0,748	-0,376	-0,351
$T_{med}:2$	0,486	<b>0,500</b>	-0,093	-0,001	$UR_{min}:4$	-0,813	-0,790	-0,439	-0,413
$T_{med}:3$	0,730	0,789	0,476	0,617	$UR_{max}:1$	-0,867	-0,787	-0,694	-0,757
$T_{med}:4$	<b>0,571</b>	0,560	0,267	0,364	$UR_{max}:2$	-0,880	-0,781	-0,703	-0,741
$T_{min}:1$	<b>-0,738</b>	<b>-0,695</b>	-0,151	-0,223	$UR_{max}:3$	-0,728	-0,676	-0,571	-0,614
$T_{min}:2$	-0,798	-0,694	-0,359	-0,404	$UR_{max}:4$	-0,752	-0,678	-0,571	-0,601
$T_{min}:3$	-0,730	<b>-0,630</b>	-0,289	-0,236	$UR_{dtr}:1$	<b>0,658</b>	<b>0,757</b>	-0,037	0,058
$T_{min}:4$	<b>-0,725</b>	<b>-0,677</b>	-0,266	-0,216	$UR_{dtr}:2$	<b>0,036</b>	<b>0,118</b>	<b>0,182</b>	<b>0,202</b>
$T_{max}:1$	0,737	0,762	0,563	0,653	$UR_{dtr}:3$	<b>-0,326</b>	<b>-0,277</b>	0,137	<b>0,075</b>
$T_{max}:2$	0,698	0,704	0,456	0,547	$UR_{dtr}:4$	<b>0,474</b>	<b>0,520</b>	0,084	0,022
$T_{max}:3$	0,814	0,780	0,558	0,575	$ET_o:1$	-0,787	-0,717	-0,460	-0,362
$T_{max}:4$	0,790	0,798	0,576	0,614	$ET_o:2$	-0,759	-0,708	-0,505	-0,543
$T_{dtr}:1$	0,826	0,819	0,350	0,438	$ET_o:3$	<b>-0,478</b>	<b>-0,320</b>	-0,148	-0,077
$T_{dtr}:2$	0,844	0,767	0,419	0,483	$ET_o:4$	<b>-0,641</b>	<b>-0,536</b>	-0,229	-0,228
$T_{dtr}:3$	0,793	0,716	0,450	0,428	$CO_2:1$	0,921	0,975	0,795	0,930
$T_{dtr}:4$	0,820	0,801	0,487	0,468	$CO_2:2$	0,920	0,976	0,795	0,930
$UR_{med}:1$	-0,510	<b>-0,604</b>	0,366	0,086	$CO_2:3$	0,920	0,976	0,797	0,931
$UR_{med}:2$	<b>-0,601</b>	<b>-0,498</b>	0,027	0,075	$CO_2:4$	0,921	0,976	0,798	0,932

Nota: “1” corresponde ao verão; “2” corresponde ao outono; “3” corresponde ao inverno; e “4” corresponde a primavera. Os valores numéricos em negrito indicam as correlações que não foram significâncias ao nível de  $p < 0,05$ .

Em média, no mundo, em 1980 os níveis de  $CO_2$  eram de 339ppm e aumentou para 402,76ppm em 2016 (tendência de  $4,34\text{ppm}\cdot\text{década}^{-1}$ ). Nas regiões Centro-Oeste e Sul do Brasil foram encontradas tendências na temperatura, precipitação, umidade relativa do ar e evapotranspiração, essas tendências são apresentadas nas Figuras 10 e 11.

A precipitação no Centro-Oeste e Sul do Brasil apresenta tendências diferentes, no Centro-Oeste a tendência é negativa ( $-6,9\text{ mm}\cdot\text{década}^{-1}$ ) e no Sul é positiva ( $3,7\text{ mm}\cdot\text{década}^{-1}$ ). Entretanto, na região Centro-Oeste existe uma alteração no regime de chuvas, havendo tendência de reduzir a precipitação na primavera, e aumentar no verão. Em outros países, como na Argentina, Verón *et al.* (2015) e Asseng *et al.* (2013) encontraram que a precipitação está aumentando na região dos Pampas. Já na Austrália, a precipitação está diminuindo na região do cinturão do trigo (ASSENG; PANNELL, 2013).

As tendências na evapotranspiração são negativas para ambas as regiões:  $-4,22\text{ mm}\cdot\text{década}^{-1}$ , no Centro-Oeste; e  $-2,15\text{ mm}\cdot\text{década}^{-1}$ , no Sul. Esse efeito negativo pode

estar associado ao aumento dos níveis de CO<sub>2</sub> além da influência de outras variáveis como temperatura e umidade relativa do ar (WANG *et al.*, 2015a).

A temperatura média apresenta uma pequena tendência no Centro-Oeste (0,06 °C·década<sup>-1</sup>) e Sul (0,04 °C·década<sup>-1</sup>). A temperatura máxima também apresenta tendência de 0,19 °C·década<sup>-1</sup>, no Centro-Oeste, e 0,21 °C·década<sup>-1</sup>, no Sul. Porém, a temperatura mínima está reduzindo 0,21 °C·década<sup>-1</sup> no Centro-Oeste e 0,15 °C·década<sup>-1</sup> no Sul. Isso fez com que o intervalo diurno de temperatura apresentasse tendência ascendente no Centro-Oeste (0,41 °C·década<sup>-1</sup>) e no Sul (0,35 °C·década<sup>-1</sup>). Existem comportamentos interatuais que são discutidos no Apêndice ??.

No Centro-Oeste a umidade relativa média está reduzindo (-0,37%·década<sup>-1</sup>), enquanto no Sul quase não houve tendência (0,006%·década<sup>-1</sup>). A umidade relativa máxima tem tendência de redução de -0,60%·década<sup>-1</sup> no Centro-Oeste e -0,47%·década<sup>-1</sup> no Sul. Da mesma forma a umidade relativa mínima também apresenta tendência de redução no Centro-Oeste (-0,76%·década<sup>-1</sup>) e no Sul (-0,54%·década<sup>-1</sup>). Ambas, umidade mínima e máxima, possuem tendências negativas, porém a umidade mínima possui o coeficiente maior, fazendo com que a tendência na amplitude seja ascendente (Centro-Oeste, 0,16%·década<sup>-1</sup>; e Sul, 0,07%·década<sup>-1</sup>).

Os dados de 1980 a 2011 foram usados para ajustar o modelo e os dados de 2012 a 2016 foram reservados para teste. As variáveis de resposta para a modelagem é a produtividade do milho e da soja (ver Figura 9).

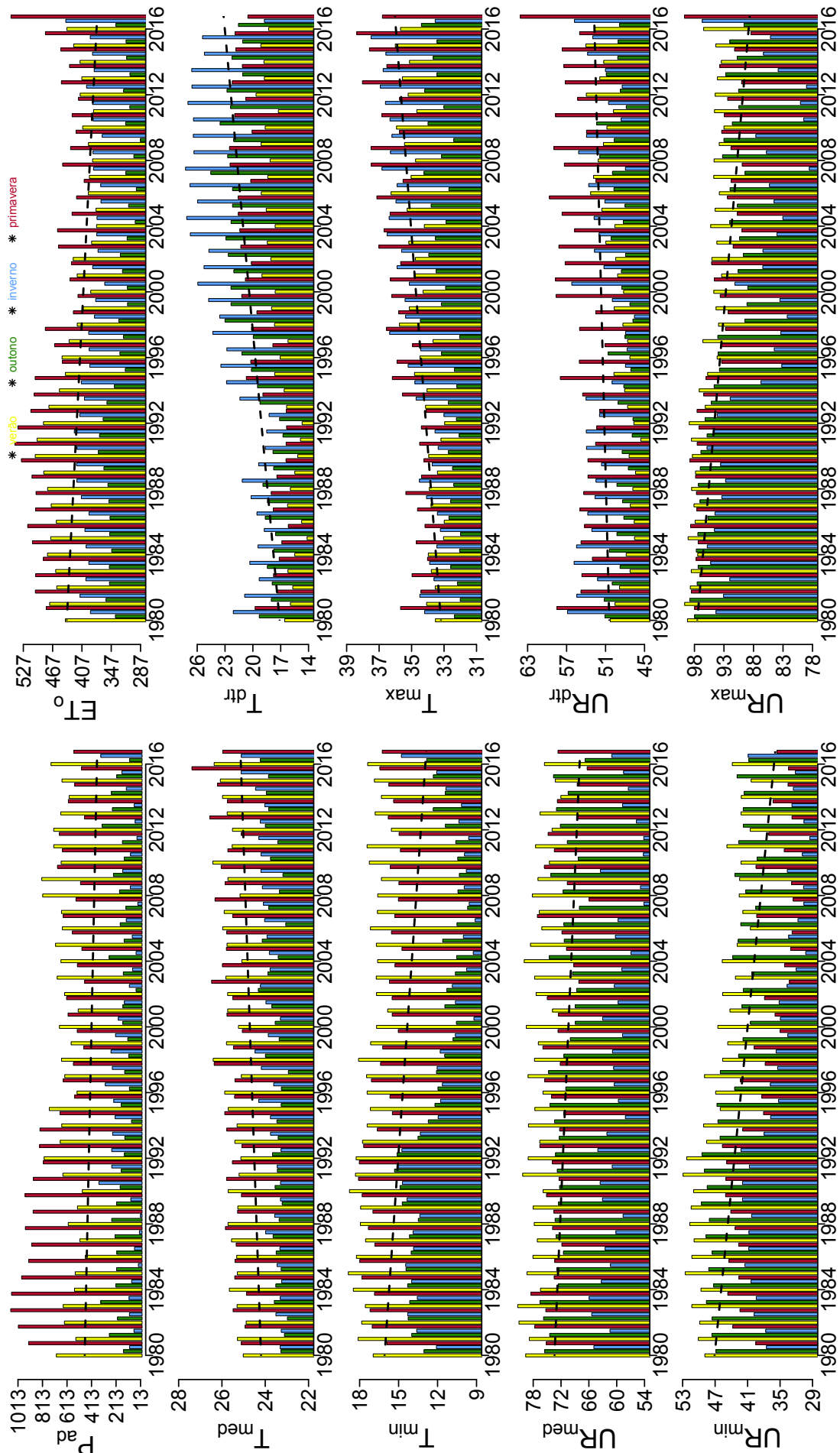


Figura 10 – Tendências climáticas na região Centro-Oeste do Brasil de 1980 a 2016

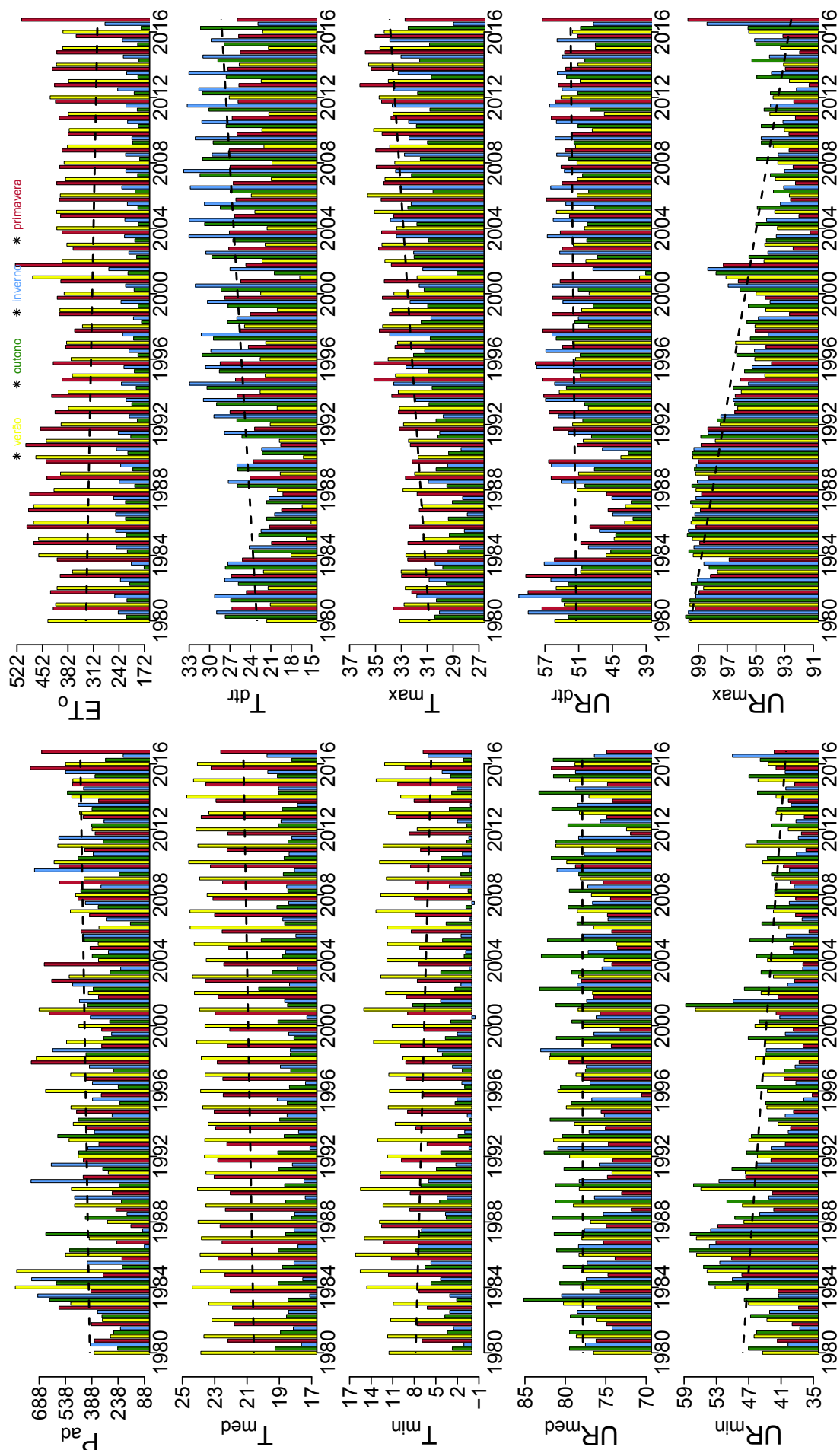


Figura 11 – Tendências climáticas na região Sul do Brasil de 1980 a 2016



### 4.3.1 Modelagem por séries temporais

A modelagem de Box-Jenkins foi realizada em quatro etapas. Nesta seção foi determinada a quantidade de diferenciações, estimação dos parâmetros desconhecidos, e teste de ajuste dos resíduos. A avaliação da previsão do modelo foi realizada juntamente com os resultados da regressão PLSR (ver seção 4.3.3).

O teste sequencial de [Dickey e Fuller \(1981\)](#) foi aplicado para determinar a quantidade de diferenciações para tornar as séries estacionárias (ver Tabela 10). Ambas as séries se tornaram estacionárias com uma diferenciação e com exceção da produtividade da soja na região Sul, as séries são estacionárias com *drift*.

Tabela 10 – Teste sequência de Dickey e Fuller para produtividade de grãos na região Centro-Oeste e Sul do Brasil

Região	Cultura	Parâmetro	Valor crítico	Nível de significância		
				1%	5%	10%
Centro	soja	$\tau$	-8,953	-3,58	-2,93	-2,60
		$\phi$	40,084	7,06	4,86	3,94
Oeste	milho	$\tau$	-6,096	-3,58	-2,93	-2,60
		$\phi$	18,578	7,06	4,86	3,94
Sul	soja	$\tau$	-7,732	-2,62	-1,95	-1,61
		$\tau$	-7,734	-3,58	-2,93	-2,60
	milho	$\phi$	29,908	7,06	4,86	3,94

Nota:  $\tau$  é a estatística do teste de  $\pi$ ; e  $\phi$  é o nível de significância do intercepto (*drift*). Os valores críticos de  $\tau$  tem que ser menores que os níveis de significância; e  $\phi$  tem que ser maior que os níveis de significância.

Para ajustar os parâmetros do modelo, dentre os existentes, foi aplicado o algoritmo de [Hyndman e Khandakar \(2008\)](#) nas séries temporais. O algoritmo percorreu um espaço de busca com os modelos possíveis e selecionou o que apresentou menor critério de AIC. Os modelos selecionados foram da família ARIMA (p, 1, q). Os parâmetros estimados, erro padrão e intervalo de confiança estão apresentados na Tabela 11, também os valores dos critérios de informações AIC e BIC são apresentados.

Na análise de resíduos, com nível de significância de 5%, o teste Shapiro-Wilk e Jarque-Bera não refugam a hipótese de normalidade dos resíduos. O teste Ljung-Box e Box-Pierce confirmam que os resíduos não são correlacionados (ver Tabela 12). Por tanto,

Tabela 11 – Parâmetros ajustados pelo modelo ARIMA para produtividade de grãos

Região	Centro-Oeste				Sul			
	soja		milho		soja		milho	
	Arima(0,1,1)		Arima(0,1,0)		Arima(2,1,0)		Arima(0,1,1)	
Modelo	$\theta_1$	Drift	Drift	$\phi_1$	$\phi_2$	$\theta_1$	Drift	
Parâmetro estimado	-0,6484	45,7868	186,9206	-0,6236	-0,4677	-0,7053	122,8291	
Erro padrão	0,2416	10,7727	63,2496	0,1576	0,1667	0,1939	34,3572	
Intervalo de confiança 2,5%	-1,1219	24,6726	62,9537	-0,9325	-0,3148	-1,0069	55,4951	
Intervalo de confiança 97,5%	-0,1748	66,9008	310,8875	-0,7943	-0,1410	-0,4036	190,1680	
Critério de Informação de Akaike	409,17		455,55	495,15		490,80		
Critério de Informação de Bayesian	413,47		458,42	499,64		495,10		

com os pressupostos exigidos pelo modelo Box-Jenkins atendidos, os modelos podem ser utilizados.

Tabela 12 – P-valor dos testes de normalidade e autocorrelação para os resíduos da região Centro-Oeste e Sul

Modelo	Teste de normalidade		Teste de autocorrelação	
	Shapiro-Wilk	Jarque-Bera	Box-Pierce	Ljung-Box
Centro-Oeste				
Soja	0,081	0,068	0,974	0,973
Milho	0,222	0,345	0,527	0,511
Sul				
Soja	0,078	0,427	0,417	0,398
Milho	0,364	0,688	0,847	0,814

#### 4.3.2 Modelagem pela regressão PLSR

Em resumo, o PLSR foi construído separadamente para a região Centro-Oeste e Sul do Brasil para as culturas de soja e milho (ver Tabela 13). Na validação-cruzada, o melhor modelo da região Centro-Oeste foi modelado sem o segmento dos anos de 1983, 1995 e 2011 e na região Sul foi 1992, 1995 e 1999. Após a modelagem sem os segmentos, a seleção das principais variáveis preditoras foi realizada pelo *scores VIP* > 1. O modelo construído para a região Sul foi forte ( $R^2 > 0,5$  e  $Q_h^2 > 0,0975$ ), como apresentado na Tabela 13. Porém, a região Centro-Oeste foi considerada com baixo poder de preditividade devido ao valor de  $R^2 < 0,5$ .

A explicação da variabilidade na produtividade de grãos na região Centro-Oeste é de 83,89% na primeira componente. A adição de mais três componentes (número de

Tabela 13 – Resumo da modelagem da produtividade de grãos na região Centro-Oeste e Sul pela regressão PLS

Região	$R^2_{soja}$	$R^2_{milho}$	Comp	$Q^2_{cum}$	Explicação de variabilidade em Y (%)	Explicação de variabilidade em Y (%) acumulada	$Q^2_{soja}$	$Q^2_{milho}$	$Q^2_h$
Centro Oeste	94,42 <sup>a</sup>	96,16 <sup>a</sup>	1	0,8238	83,89	83,89	0,8421	0,8054	0,8237
	32,73 <sup>b</sup>	18,39 <sup>b</sup>	2	0,8887	7,33	91,22	0,0107	0,6612	0,3689
			3	0,9163	3,51	94,73	0,3593	-0,0343	0,2475
			4 <sup>c</sup>	0,9169	0,73	95,46	-0,0379	0,0594	0,0076
			5	0,8895	0,34	95,80	-0,1313	-0,3718	-0,2301
Sul	89,29 <sup>a</sup>	94,66 <sup>a</sup>	1	0,6542	68,72	68,72	0,5677	0,7408	0,6543
	97,72 <sup>b</sup>	95,56 <sup>b</sup>	2	0,7943	16,20	84,92	0,4154	0,3875	0,4051
			3	0,7925	3,21	88,13	0,0579	-0,1216	-0,0088
			4	0,8175	2,46	90,59	-0,0608	0,3235	0,1205
			5 <sup>c</sup>	0,8212	1,28	91,88	0,0655	-0,0635	0,0202
			6	0,7506	0,97	92,84	-0,1475	-0,3654	-0,2281

Nota: **a** é o coeficiente de determinação de calibração; **b** é o coeficiente de determinação da validação; e **c** indica o número de componentes apropriadas. O  $R^2$  bondade do ajuste,  $Q^2_{cum}$  bondade da predição na validação-cruzada, e  $Q^2_h$  bondade da componente na predição na validação-cruzada foram calculados para o modelo PLSR.

componentes total igual a 4) levou a explicação acumulada de variância total do modelo para 95,46% das variáveis preditoras (ver Tabela 13). Na primeira componente dos pesos **W** não teve variáveis preditoras maiores que 0,3 em módulo (ver Tabela 14), porém os pesos ficaram bem divididos. Por outro lado, na componente dois, a produtividade de grãos foi dominada pelos níveis de CO<sub>2</sub> com direção positiva. A terceira componente foi dominada pela precipitação, temperatura máxima e amplitude da umidade do ar, ambas com efeito negativo na produtividade grãos.

Para a região Sul, a explicação da variabilidade da produtividade de grãos é de 68,72% na primeira componente. Com cinco componentes, a explicação acumulada é atualizada para 91,88% (ver Tabela 13). Nos pesos **W**, a primeira componente foi dominada pelos níveis de CO<sub>2</sub> com direção positiva. Na segunda componente, a umidade do ar média teve maior contribuição positiva e a amplitude da umidade do ar teve dominância negativa.

No Centro-Oeste, as variáveis preditoras (ver Tabela 15) mais significantes para explicar a produtividade da soja foram a amplitude da umidade do ar e a precipitação, atuando negativamente; e os níveis de CO<sub>2</sub>, atuando de forma positiva. Na produtividade de milho, as variáveis preditoras mais significantes foram: a umidade mínima, reduzindo a produtividade; e ET<sub>0</sub> e níveis de CO<sub>2</sub> incrementando a produtividade.

Para explicar a produtividade do milho na região Sul as variáveis preditoras (ver Tabela 15) mais significantes foram a umidade máxima, amplitude de temperatura e

Tabela 14 – Valores da influência da variável na projeção (VIP) e regressão por mínimos quadrados parciais (PLSR) para região Centro-Oeste e Sul do Brasil

Região	Variáveis em X	VIP	$w_1^l$	$w_2^l$	$w_3^l$	$w_4^l$	$w_5^l$	$w_6^l$
Centro-Oeste								
	$P_{ad}:4$	1,0244	-0,2111	-0,0276	<b>-0,3225</b>	0,2539	<b>-0,3616</b>	
	$T_{med}:3$	0,9910	0,1945	0,2951	-0,0984	-0,2410	<b>0,3830</b>	
	$T_{min}:1$	0,8926	-0,1839	0,1027	-0,1983	<b>-0,3344</b>	-0,1009	
	$T_{min}:2$	0,9298	-0,1833	0,2669	-0,0610	-0,0035	0,1831	
	$T_{max}:1$	0,9758	0,1984	0,0527	<b>-0,3563</b>	-0,1770	-0,0570	
	$T_{max}:2$	0,8841	0,1809	0,0484	-0,1923	<b>-0,5394</b>	-0,1332	
	$T_{max}:3$	0,9158	0,1939	-0,1142	0,0540	0,1938	0,1937	
	$T_{max}:4$	0,9364	0,2023	-0,0214	-0,1124	-0,0881	0,2012	
	$T_{dtr}:1$	0,9918	0,2147	-0,0217	-0,1109	0,0674	0,0182	
	$T_{dtr}:2$	0,9619	0,2003	-0,1837	-0,0208	-0,1842	-0,1836	
	$T_{dtr}:3$	0,9237	0,1831	-0,2523	0,0671	0,1842	-0,0528	
	$T_{dtr}:4$	0,9524	0,2017	-0,1439	-0,0896	0,0135	-0,1350	
	$UR_{min}:1$	1,0135	-0,2154	0,0162	0,2350	0,0170	-0,0994	
	$UR_{min}:2$	0,9885	-0,2107	0,0419	-0,0737	-0,1206	<b>0,6324</b>	
	$UR_{min}:3$	0,9197	-0,1857	0,2340	0,0623	-0,1530	0,1283	
	$UR_{min}:4$	0,9385	-0,1982	0,1479	0,0718	0,1669	-0,1950	
	$UR_{max}:1$	0,9792	-0,2078	0,0082	-0,2358	0,0289	-0,0888	
	$UR_{max}:2$	0,9950	-0,2051	0,0086	-0,2996	<b>0,3081</b>	-0,1830	
	$UR_{dtr}:1$	1,0057	0,1812	-0,0200	<b>-0,5989</b>	-0,0034	0,0897	
	$ET_0:1$	0,9266	-0,1810	0,2855	0,0431	0,1498	-0,2778	
	$CO_2:1$	1,1882	0,2365	<b>0,3675</b>	0,1628	0,2546	-0,1464	
	$CO_2:2$	1,1880	0,2364	<b>0,3682</b>	0,1602	0,2497	-0,1411	
	$CO_2:3$	1,1864	0,2364	<b>0,3658</b>	0,1535	0,2399	-0,1471	
	$CO_2:4$	1,1868	0,2366	<b>0,3654</b>	0,1521	0,2409	-0,1546	
Sul								
	$T_{med}:3$	0,8533	0,2300	0,1978	-0,1904	<b>0,3406</b>	<b>-0,3929</b>	-0,110
	$T_{max}:1$	0,8608	0,2446	-0,1552	0,0289	-0,0668	-0,2401	<b>0,4190</b>
	$T_{max}:4$	0,8761	0,2449	-0,1307	0,2030	0,0754	<b>0,4835</b>	-0,0757
	$T_{dtr}:4$	0,8630	0,2125	-0,0336	<b>0,6504</b>	<b>0,4772</b>	<b>0,4774</b>	0,0594
	$UR_{med}:1$	1,1548	0,0929	<b>0,6667</b>	<b>0,7079</b>	-0,1164	-0,1315	0,0432
	$UR_{max}:1$	0,9966	-0,2963	0,0677	0,1023	0,2791	<b>0,3388</b>	<b>0,6130</b>
	$UR_{max}:2$	0,9962	-0,2926	0,0682	0,0713	<b>0,3616</b>	<b>0,3996</b>	<b>0,6255</b>
	$UR_{max}:3$	0,9262	-0,2424	0,2348	0,1211	<b>0,4197</b>	0,1183	-0,0919
	$UR_{max}:4$	0,8959	-0,2366	0,2094	0,0697	<b>0,3932</b>	-0,1264	-0,4273
	$UR_{dtr}:1$	0,8024	0,0090	<b>-0,4525</b>	0,1958	0,2712	<b>-0,3668</b>	-0,0473
	$CO_2:1$	1,1592	<b>0,3480</b>	0,2192	-0,0555	0,1839	0,1232	0,0985
	$CO_2:2$	1,1599	<b>0,3480</b>	0,2201	-0,0558	0,1861	0,1256	0,1165
	$CO_2:3$	1,1627	<b>0,3486</b>	0,2235	-0,0510	0,1891	0,1364	0,1434
	$CO_2:4$	1,1636	<b>0,3487</b>	0,2246	-0,0505	0,1887	0,1411	0,1513

Nota: **VIP** é a importância da variável independente em cada projeção do modelo; + ou - é a direção dos parâmetros de regressão. Valores > 0,3 em magnitude, que são apresentadas em negrito, indicam que na componente do PLS as variáveis correspondes são as que possuem maior peso.

Tabela 15 – Lista dos coeficientes da regressão para produtividade de grãos<sup>a</sup> com a regressão PLS para região Centro-Oeste e Sul

Variáveis predictoras	Variáveis resposta			
	Centro-Oeste		Sul	
	Soja	Milho	Soja	Milho
$P_{ad}:4$	<b>-0,1475</b>	0,0337		
$T_{med}:3$	0,0262	0,0270	-0,0300	0,1007
$T_{min}:1$	-0,1202	-0,0440		
$T_{min}:2$	-0,0292	0,0079		
$T_{max}:1$	-0,0903	0,0304	0,0601	0,1446
$T_{max}:2$	-0,0608	-0,0074		
$T_{max}:3$	0,0611	0,0030		
$T_{max}:4$	-0,0041	-0,0137	0,1803	0,0945
$T_{dtr}:1$	0,0082	0,0411		
$T_{dtr}:2$	0,0008	-0,0085		
$T_{dtr}:3$	0,0457	0,0001		
$T_{dtr}:4$	-0,0054	0,0209	<b>0,3897</b>	<b>0,2677</b>
$UR_{min}:1$	0,0401	-0,0137		
$UR_{min}:2$	-0,0703	<b>-0,1560</b>		
$UR_{min}:3$	0,0001	-0,0102		
$UR_{min}:4$	0,0103	0,0606		
$UR_{med}:1$			<b>0,4362</b>	0,1985
$UR_{max}:1$	-0,1253	-0,0293	<b>0,2544</b>	<b>0,2970</b>
$UR_{max}:2$	-0,1293	0,0242	<b>0,2721</b>	<b>0,3313</b>
$UR_{max}:3$			0,0914	0,0786
$UR_{max}:4$			-0,0903	-0,1047
$UR_{dtr}:1$	<b>-0,1760</b>	-0,0024	-0,2108	-0,1333
$ET_o:1$	0,0179	<b>0,1154</b>		
$CO_2:1$	<b>0,1657</b>	<b>0,2228</b>	<b>0,2300</b>	<b>0,2651</b>
$CO_2:2$	<b>0,1646</b>	<b>0,2213</b>	<b>0,2361</b>	<b>0,2735</b>
$CO_2:3$	<b>0,1613</b>	<b>0,2201</b>	<b>0,2493</b>	<b>0,2878</b>
$CO_2:4$	<b>0,1608</b>	<b>0,2213</b>	<b>0,2533</b>	<b>0,2919</b>

Nota: a são dados escalados para variância unitária pela Equação 3.2. Os coeficientes com maior valor em magnitude, apresentados em negrito, indicam que a variável predictoras possui maior explicação para construção da variável de resposta (soja e/ou milho). Ambos modelos com número de componentes apropriados.

níveis de CO<sub>2</sub>, incrementando a produtividade. Para a produtividade da soja, as variáveis preditoras mais importantes foram: amplitude de temperatura, umidade média e máxima; e os níveis de CO<sub>2</sub>, atuando positivamente na produtividade de grãos.

Embora os valores dos coeficientes da regressão (ver Tabela 15) indicarem a mudança individual nas variáveis preditoras na produtividade de grãos, uma expressão mais conveniente e abrangente da importância relativa das mudanças climáticas poderia ser obtida explorando os valores dos *scores* **VIP** (ver Tabela 14). Para a produtividade de grãos no Centro-Oeste, os valores dos *scores* **VIP** das variáveis preditoras mais importantes foram: precipitação ( $vip = 1,0244$ ); umidade mínima ( $vip = 1,0135$ ); amplitude da umidade do ar ( $vip = 1,0057$ ); e os níveis de CO<sub>2</sub>. Para a produtividade de grãos na região Sul, os valores dos *scores* **VIP** foram: umidade do ar média ( $vip = 1,1548$ ); e os níveis de CO<sub>2</sub>. Variáveis preditoras com **VIP** < 0,5 são consideradas insignificantes para predição (WOL-DESENBET *et al.*, 2017). Verón *et al.* (2015) relataram que, nos Pampas, a temperatura foi mais impactante na produtividade dos grãos de soja e milho do que a precipitação. A diferença das variáveis mais importantes pode estar associada às características regionais.

#### 4.3.3 Análise comparativa e previsão

Para analisar a produtividade de grãos é necessário que a modelagem possua um bom ajuste. Os valores de NSE, R<sup>2</sup>, PBIAS e MAPE para calibração e validação da produtividade de grãos anual estão na Tabela 16. Todos os valores de NSE e R<sup>2</sup>, para a região Sul, com o modelo de regressão PLS são altos (R<sup>2</sup> e NSE > 0,75). Os valores de PBIAS, na mesma região, são baixos pela sub e/ou sobre predição menor que 2%. Para região Centro-Oeste, os valores de NSE e R<sup>2</sup> foram altos para calibração no modelo de regressão PLS, mas na validação caíram. Porém, os valores de PBIAS e MAPE são baixos (PBIAS < 2%). No geral, a regressão PLS para predição da produtividade de grãos no Brasil apresentou melhores resultados que os modelos de Box-Jenkins. Os resultados da predição pela regressão PLS e séries temporais na região Centro-Oeste e Sul estão Figura 12.

Em média, entre os anos de 1980 a 2016, os níveis de CO<sub>2</sub> foram de aproximadamente 360 ppm. Se considerar todas as variáveis preditoras média e aumentar os níveis de CO<sub>2</sub> em 10%, a produtividade de soja será incrementada em 14% na região Sul e 29% no Centro-Oeste, a produtividade do milho irá aumentar em 52 e 55%, na mesma ordem.

Tabela 16 – Critério para examinar acurácia dos modelos na região Centro-Oeste e Sul

Índice	Região Centro-Oeste				Região Sul			
	Séries temporais		Regressão PLS		Séries temporais		Regressão PLS	
	Soja	Milho	Soja	Milho	Soja	Milho	Soja	Milho
<b>Calibração</b>								
NSE	-0,178	0,633	0,944	0,960	0,076	0,494	0,892	0,946
R <sup>2</sup> (%)	65,631	91,186	94,419	96,165	56,598	86,115	89,297	94,659
PBIAS (%)	21,320	18,600	0,031	0,739	-13,850	-17,062	-0,702	-1,062
MAPE (%)	17,274	18,822	3,723	6,649	20,012	24,504	7,438	10,197
<b>Validação</b>								
NSE	-74,213	-12,470	-3,150	-0,165	-0,639	-1,728	0,974	0,939
R <sup>2</sup> (%)	35,720	25,246	32,729	18,399	34,429	26,124	97,723	95,559
PBIAS (%)	8,986	20,324	-1,151	1,502	-9,788	-13,717	-0,277	1,689
MAPE (%)	9,030	16,824	1,245	4,066	13,662	20,638	1,883	1,784

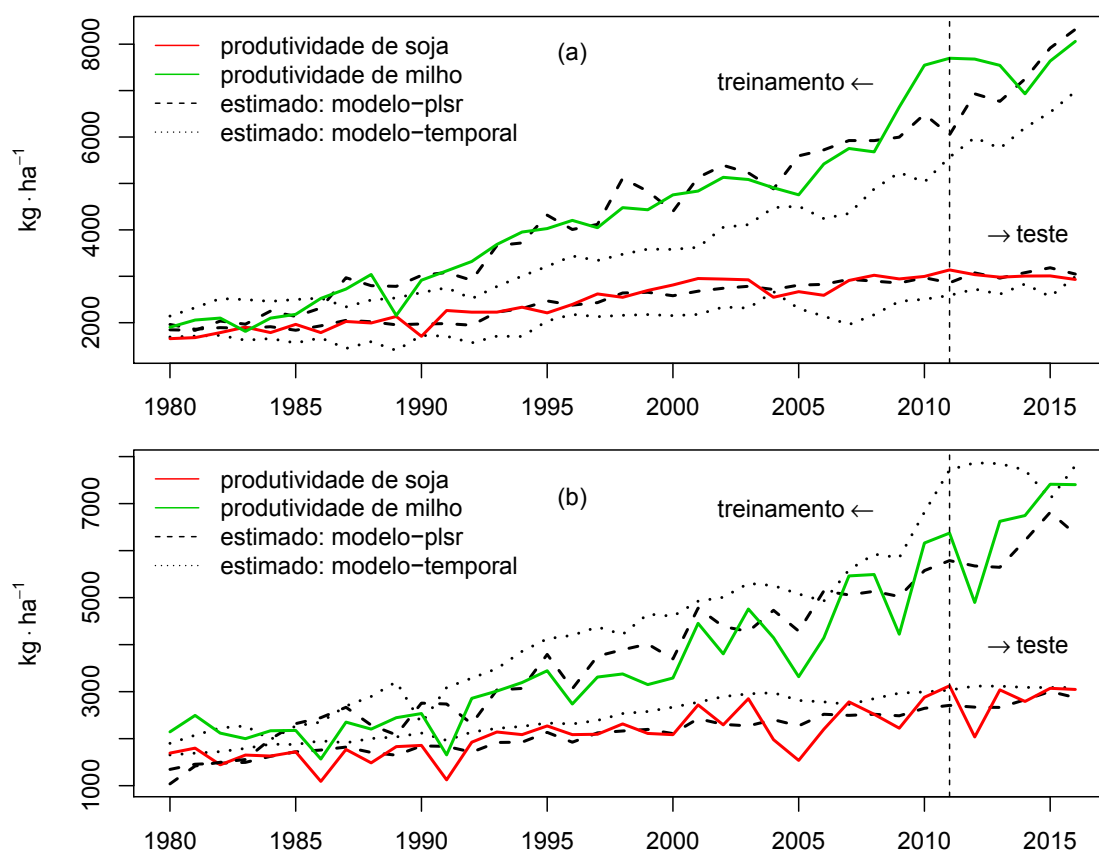


Figura 12 – Série histórica real da produtividade de grãos de 1980 a 2016 e a produtividade prevista por PLSR e séries temporais. (a) produtividade da região Centro-Oeste. (b) produtividade da região Sul

## **5 CONSIDERAÇÕES FINAIS**

Neste capítulo será apresentada a conclusão do método para obter as variáveis climáticas e da modelagem da produtividade de grãos.

### **5.1 Métodos para completar os dados climáticos**

Para determinar a influência das mudanças climáticas sobre a produtividade de grãos foi necessário identificar os métodos mais adequados para completar os dados climáticos e para calcular a evapotranspiração.

No Brasil, em regiões que possuem alta densidade de dados climáticos disponíveis, como Nordeste e Sudeste, é possível completar os dados históricos com maior qualidade. Os resultados apontam que quanto mais próximas as estações meteorológicas, mais as variáveis são correlacionadas.

É possível completar todos os dados climáticos do INMET analisados com as informações dos vizinhos mais próximos, e em algumas observações é necessário o uso de média móvel a partir de 1980. Antes desta data é necessário excluir algumas estações meteorológicas para se obter séries históricas completas.

Para avaliar quais os melhores métodos para completar os dados climáticos foram excluídas algumas observações (variando de 1 a 20%). Esta análise permitiu determinar o método de preenchimento de falhas mais adequado para cada variável climática, como: precipitação, vizinhos; temperatura média, regressão; temperatura máxima, interpolação; temperatura mínima, regressão; e umidade relativa, regressão.

Para calcular a evapotranspiração de referência onde somente dados de umidade relativa do ar estão ausentes, o melhor método é de Penman-Monteith. Na ausência de dados de velocidade dos ventos, o método de Hargreaves-Samani apresentou melhor desempenho.

### **5.2 Predição da produtividade de grãos**

Buscando compreender como as mudanças climáticas influenciam na produtividade dos grãos de milho e soja, foi utilizada a regressão PLS. As variáveis consideradas foram precipitação, temperatura, umidade, evapotranspiração e níveis de dióxido de carbono.



Foram incluídos dados das regiões Centro-Oeste e Sul do Brasil, correspondente aos anos de 1980 a 2016. A fim de validar os resultados do PLSR foi verificada a acurácia utilizando modelagem por séries temporais.

Para fazer a previsão da produtividade de grãos, a regressão PLS apresentou melhor ajuste. Na região Centro-Oeste, o modelo explicou 95,45% da variabilidade da produtividade de grãos, com vinte e quatro variáveis. Na região Sul, quatorze variáveis explicaram 91,88% da variabilidade da produtividade de grãos.

As variáveis que mais contribuíram na construção dos modelos foram o CO<sub>2</sub> e a umidade do ar; o CO<sub>2</sub>, influenciando positivamente na produtividade da soja e do milho; e a umidade do ar, influenciando negativamente.

Para a região Centro-Oeste, a validação do modelo não apresentou um ótimo desempenho. Isso pode estar associado ao número de Estações Meteorológicas, que é menor nessa região (ver Figura 7).

### 5.3 Trabalho futuro

Como pesquisa futura é sugerido utilizar dados de satélite no lugar de dados de estações meteorológicas. Na predição, pode ser avaliada a influência das mudanças climáticas pelo método de co-integração e por dados em painel (VERÓN *et al.*, 2015). Também, pode ser aplicada meta-heurística para melhorar a regressão PLS. Outras variáveis podem ser incluídas, como variáveis econômicas (CHEN *et al.*, 2016).

## REFERÊNCIAS

- ABDI, H.; WILLIAMS, L. J. Tukey's honestly significant difference (HSD) test. *Encyclopedia of Research Design. Thousand Oaks, CA: Sage*, p. 1–5, 2010. Citado na página 37.
- AHMADI, M.; HADDAD, O. B.; LOÁICIGA, H. A. Adaptive reservoir operation rules under climatic change. *Water Resour Manag*, v. 29, n. 4, p. 1247–1266, 2015. doi:10.1007/s11269-014-0871-0. Citado na página 17.
- AKAIKE, H. A new look at the statistical model identification. *IEEE transactions on automatic control*, v. 19, n. 6, p. 716–723, 1974. doi: 10.1109/TAC.1974.1100705. Citado na página 44.
- AKIMA, H. A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points. *ACM Trans Math Softw*, v. 4, n. 2, p. 148–159, 1978. doi: 10.1145/355780.355786. Citado 2 vezes nas páginas 18 e 32.
- ALENCAR, L.; SEDIYAMA, G.; MANTOVANI, E. Estimativa da evapotranspiração de referência (ET<sub>o</sub> padrão FAO), para minas gerais, na ausência de alguns dados climáticos. *Eng Agríc*, v. 35, n. 1, p. 39–50, 2015. doi: 10.1590/1809-4430. Citado na página 56.
- ALLEN, R. G. *et al.* Crop evapotranspiration-guidelines for computing crop water requirements-FAO irrigation and drainage paper 56. *FAO, Rome*, v. 300, n. 9, p. D05109, 1998. Citado 3 vezes nas páginas 39, 40 e 41.
- ALLEN, R. G.; PRUITT, W. O. Rational use of the FAO Blaney-Criddle formula. *J Irrig Drain Eng*, v. 112, n. 2, p. 139–155, 1986. doi: 10.1061/(ASCE)0733-9437(1986)112:2(139). Citado 2 vezes nas páginas 39 e 40.
- ALLEN, R. J.; DEGAETANO, A. T. Estimating missing daily temperature extremes using an optimized regression approach. *Int J Climatol*, v. 21, n. 11, p. 1305–1319, 2001. doi: 10.1002/joc.679. Citado 2 vezes nas páginas 18 e 36.
- ALMANAQUE-BRASIL. Atlas: O Brasil em mapas. Divisao politica, aspectos fisicos, sociais e economicos.(Brasil: Geografia). *Almanaque abril*, v. 38, p. 642, 2012. Citado na página 30.
- ASSENG, S.; PANNELL, D. J. Adapting dryland agriculture to climate change: Farming implications and research and development needs in western australia. *Clim Chang*, v. 118, n. 2, p. 167–181, 2013. doi: 10.1007/s10584-012-0623-1. Citado 3 vezes nas páginas 17, 26 e 59.
- ASSENG, S. *et al.* Has climate change opened new opportunities for wheat cropping in argentina? *Clim Chang*, v. 117, n. 1-2, p. 181–196, 2013. doi: 10.1007/s10584-012-0553-y. Citado 3 vezes nas páginas 17, 26 e 59.
- BAI, H. *et al.* Attribution of yield change for rice-wheat rotation system in china to climate change, cultivars and agronomic management in the past three decades. *Clim Chang*, v. 135, n. 3-4, p. 539–553, 2016. doi: 10.1007/s10584-015-1579-8. Citado na página 57.

- BAJAT, B. *et al.* Spatial analysis of the temperature trends in Serbia during the period 1961-2010. *Theor Appl Climatol*, v. 121, n. 1, p. 289–301, 2015. doi: 10.1007/s00704-014-1243-7. Citado 2 vezes nas páginas 17 e 20.
- BERKHOUT, E.; GLOVER, D.; KUYVENHOVEN, A. On-farm impact of the system of rice intensification (SRI): Evidence and knowledge gaps. *Agric Syst*, v. 132, p. 157–166, 2015. doi: 10.1016/j.agry.2014.10.001. Citado 3 vezes nas páginas 17, 22 e 23.
- BERTRAND, J. W. M.; FRANSOO, J. C. Operations management research methodologies using quantitative modeling. *Int J Oper Prod Man*, v. 22, n. 2, p. 241–264, 2002. doi: 10.1108/01443570210414338. Citado na página 29.
- BIER, A. A.; FERRAZ, S. E. T. Comparison of gap filling methodologies for meteorological data in southern Brazil stations. *Rev bras meteorol*, v. 32, n. 2, p. 215–226, 2017. doi: 10.1590/0102-77863220008. Citado 2 vezes nas páginas 18 e 53.
- BOX, G. E. *et al.* *Time series analysis: forecasting and control*. [S.l.]: John Wiley & Sons, 2008. Citado na página 42.
- CALDARELLI, C. E.; BACCHI, M. R. P. Fatores de influência no preço do milho no Brasil. *Nova Economia*, v. 22, p. 141–164, 2012. doi: 10.1590/S0103-63512012000100005. Citado na página 25.
- CALZADILLA, A. *et al.* Climate change impacts on global agriculture. *Clim Chang*, v. 120, n. 1-2, p. 357–374, 2013. doi: 10.1007/s10584-013-0822-4. Citado 2 vezes nas páginas 17 e 25.
- CAMMARANO, D. *et al.* Using historical climate observations to understand future climate change crop yield impacts in the southeastern US. *Clim Chang*, v. 134, n. 1-2, p. 311–326, 2016. doi: 10.1007/s10584-015-1497-9. Citado 4 vezes nas páginas 23, 25, 26 e 32.
- CAO, F. *et al.* Genotypic and environmental variation in cadmium, chromium, lead and copper in rice and approaches for reducing the accumulation. *Sci Total Environ*, v. 496, p. 275–281, 2014. doi: 10.1016/j.scitotenv.2014.07.064. Citado na página 27.
- CARKNER, M. K.; ENTZ, M. H. Growing environment contributes more to soybean yield than cultivar under organic management. *Field Crops Res*, v. 207, p. 42–51, 2017. doi: 10.1016/j.fcr.2017.02.019. Citado na página 18.
- CAUCHICK-MIGUEL, P. A. Estudo de caso na engenharia de produção: estruturação e recomendações para sua condução. *Production*, v. 17, n. 1, p. 216–229, 2007. doi: 396742029015. Citado na página 29.
- CAUCHICK-MIGUEL, P. A. *et al.* *Metodologia de pesquisa em engenharia de produção e gestão de operações*. 2. ed. Rio de Janeiro: Elsevier: ABREPO, 2012. Citado na página 29.
- CEULEMANS, K.; MOLDEREZ, I.; LIEDEKERKE, L. V. Sustainability reporting in higher education: a comprehensive review of the recent literature and paths for further research. *J Clean Prod*, v. 106, p. 127–143, 2015. doi: 10.1016/j.jclepro.2014.09.052. Citado 2 vezes nas páginas 21 e 22.

- CHAI, T.; DRAXLER, R. R. Root mean square error (RMSE) or mean absolute error (MAE)?—arguments against avoiding rmse in the literature. *Geoscientific Model Development*, v. 7, n. 3, p. 1247–1250, 2014. doi: 10.5194/gmd-7-1247-2014. Citado na página 50.
- CHEN, S.; CHEN, X.; XU, J. Assessing the impacts of temperature variations on rice yield in China. *Clim Chang*, v. 138, n. 1-2, p. 191–205, 2016. doi: 10.1007/s10584-016-1707-0. Citado 2 vezes nas páginas 27 e 71.
- CONAB. *Séries históricas de área plantada, produtividade e produção, relativas às safras*. 2017. Disponível em: <<http://www.conab.gov.br/conteudos.php?a=1252&t=2>>. Citado 2 vezes nas páginas 30 e 57.
- CONWAY, T. J. *et al.* Evidence for interannual variability of the carbon cycle from the National Oceanic and Atmospheric Administration/Climate Monitoring and Diagnostics Laboratory global air sampling network. *J Geophys Res Atmos*, Wiley Online Library, v. 99, n. D11, p. 22831–22855, 1994. Citado na página 31.
- DEBRUIN, H. From penman to makkink. *Evaporation and weather*, v. 39, p. 5–29, 1987. Citado na página 39.
- DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, v. 74, n. 366a, p. 427–431, 1979. doi: 10.1080/01621459.1979.10482531. Citado na página 43.
- DICKEY, D. A.; FULLER, W. A. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: Journal of the Econometric Society*, p. 1057–1072, 1981. doi: 10.2307/1912517. Citado 2 vezes nas páginas 43 e 63.
- DING, D. *et al.* Impact assessment of climate change and later-maturing cultivars on winter wheat growth and soil water deficit on the loess Plateau of China. *Clim Chang*, v. 138, n. 1-2, p. 157–171, 2016. doi: 10.1007/s10584-016-1714-1. Citado na página 17.
- DOORENBOS, J.; PRUITT, W. Guidelines for predicting crop water requirements. *FAO, Roma (Italia)*, v. 24, p. 144, 1977. Citado na página 38.
- EISCHEID, J. K. *et al.* The quality control of long-term climatological data using objective data analysis. *J Appl Meteorol*, v. 34, n. 12, p. 2787–2795, 1995. doi: 10.1175/1520-0450(1995). Citado 2 vezes nas páginas 18 e 55.
- EVANGELISTA, P.; YOUNG, N.; BURNETT, J. How will climate change spatially affect agriculture production in Ethiopia? case studies of important cereal crops. *Clim Chang*, v. 119, n. 3-4, p. 855–873, 2013. doi: 10.1007/s10584-013-0776-6. Citado na página 20.
- FERNÁNDEZ-FERNÁNDEZ, M. *et al.* The climate in zafra from 1750 to 1840: history and description of weather observations. *Clim Chang*, v. 126, n. 1-2, p. 107–118, 2014. doi: 10.1007/s10584-014-1201-5. Citado na página 33.
- FERNANDEZ, M. A. *et al.* Value of decadal climate variability information for agriculture in the Missouri river basin. *Clim Chang*, v. 139, n. 3-4, p. 517–533, 2016. doi: 10.1007/s10584-016-1807-x. Citado 2 vezes nas páginas 24 e 25.

FERRARI, G. T.; OZAKI, V. Missing data imputation of climate datasets: Implications to modeling extreme drought events. *Rev bras meteorol*, v. 29, n. 1, p. 21–28, 2014. doi: 10.1590/S0102. Citado 3 vezes nas páginas 18, 31 e 35.

FERREIRA, D. *Análise da variabilidade climática e suas consequências para a produtividade da soja na região Sul do Brasil*. Tese (Doutorado) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, SP, 2010. Citado 3 vezes nas páginas 17, 21 e 25.

GAO, B. *et al.* Nitrous oxide and methane emissions from optimized and alternative cereal cropping systems on the north China plain: a two-year field study. *Sci Total Environ*, v. 472, p. 112–124, 2014. doi: 10.1016/j.scitotenv.2013.11.003. Citado 2 vezes nas páginas 25 e 27.

GARZA-REYES, J. A. Green lean and the need for six sigma. *Int J Lean Six Sigma*, v. 6, n. 3, p. 226–248, 2015. doi: 10.1108/IJLSS-04-2014-0010. Citado na página 22.

GAVA, R. *Os efeitos do estresse hídrico na cultura da soja (Glycine max, (L.) Merrill)*. Tese (Doutorado) — Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, SP, 2014. Citado na página 26.

GELADI, P.; KOWALSKI, B. R. Partial least-squares regression: a tutorial. *Anal Chim Acta*, v. 185, p. 1–17, 1986. doi: 10.1016/0003-2670(86)80028-9. Citado 2 vezes nas páginas 45 e 46.

GODOY, J. L.; VEGA, J. R.; MARCHETTI, J. L. Relationships between pca and pls-regression. *Chemom Intell Lab Syst*, v. 130, p. 182–191, 2014. doi: 10.1016/j.chemolab.2013.11.008. Citado na página 46.

GOODWIN, P.; LAWTON, R. On the asymmetry of the symmetric mape. *International journal of forecasting*, v. 15, n. 4, p. 405–408, 1999. doi: 10.1016/S0169-2070(99)00007-2. Citado na página 50.

GUO, D.; WESTRA, S.; MAIER, H. R. An r package for modelling actual, potential and reference evapotranspiration. *Environ Model Softw*, v. 78, p. 216–224, 2016. doi: 10.1016/j.envsoft.2015.12.019. Citado na página 39.

HAO, X. *et al.* Effects of open-air elevated atmospheric CO<sub>2</sub> concentration on yield quality of soybean (*Glycine max (L.) Merr*). *Agric Ecosyst Environ*, v. 192, p. 80–84, 2014. doi: 10.1016/j.agee.2014.04.002. Citado 4 vezes nas páginas 17, 21, 26 e 58.

HARGREAVES, G. H.; SAMANI, Z. A. Reference crop evapotranspiration from temperature. *Appl Eng Agric*, v. 1, n. 2, p. 96–99, 1985. doi:10.13031/2013.26773. Citado 2 vezes nas páginas 39 e 41.

HÖSKULDSSON, A. Dimension of linear models. *Chemometr Intell Lab*, v. 32, n. 1, p. 37–55, 1996. doi: 10.1016/0169-7439(95)00058-5. Citado na página 48.

HUSSAIN, M. Z. *et al.* Future carbon dioxide concentration decreases canopy evapotranspiration and soil water depletion by field-grown maize. *Glob Change Biol*, v. 19, n. 5, p. 1572–1584, 2013. doi: 10.1111/gcb.12155. Citado na página 26.

HYNDMAN, R. J.; KHANDAKAR, Y. Automatic time series forecasting: The forecast package for R. *J Stat Softw*, v. 27, n. 3, p. 1–22, 2008. doi: 10.18637/jss.v027.i03. Citado na página 63.

INMET. *Banco de dados meteorológicos para ensino e pesquisa*. 2017. Disponível em: <<http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep>>. Citado na página 30.

INMET. *Rede de Estações Meteorológicas Automáticas do INMET. NOTA TÉCNICA No. 001/2011/SEGER/LAIME/CSC/INMET*. 2017. Disponível em: <[http://www.inmet.gov.br/portal/css/content/topo\\_iframe/pdf/Nota\\_TecnicaRede\\_estacoes\\_INMET](http://www.inmet.gov.br/portal/css/content/topo_iframe/pdf/Nota_TecnicaRede_estacoes_INMET)>. Citado na página 33.

IPCC. Climate change 2001: The scientific basis. In: *Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. New York: [s.n.], 2001. p. 893. Disponível em: <[https://www.ipcc.ch/ipccreports/tar/wg1/pdf/WGI\\_TAR\\_full\\_report.pdf](https://www.ipcc.ch/ipccreports/tar/wg1/pdf/WGI_TAR_full_report.pdf)>. Citado 2 vezes nas páginas 17 e 20.

IPCC. Climate change 2014: Synthesis report. In: *Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Geneva: [s.n.], 2014. Disponível em: <[https://www.ipcc.ch/pdf/assessment-report/ar5/syr/SYR\\_AR5\\_FINAL\\_full\\_wcover.pdf](https://www.ipcc.ch/pdf/assessment-report/ar5/syr/SYR_AR5_FINAL_full_wcover.pdf)>. Citado 2 vezes nas páginas 17 e 20.

JOHANSSON, G.; SUNDIN, E. Lean and green product development: two sides of the same coin? *J Clean Prod*, v. 85, p. 104–121, 2014. doi: 10.1016/j.jclepro.2014.04.005. Citado na página 21.

JONES, J. W. *et al.* Brief history of agricultural systems modeling. *Agric Syst*, v. 155, p. 240–254, 2017. doi: 10.1016/j.agsy.2016.05.014. Citado 4 vezes nas páginas 17, 24, 26 e 58.

JONES, J. W. *et al.* Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. *Agric Syst*, v. 155, p. 269–288, 2017. doi: 10.1016/j.agsy.2016.09.021. Citado na página 17.

JU, H. *et al.* The impacts of climate change on agricultural production systems in China. *Clim Chang*, v. 120, n. 1-2, p. 313–324, 2013. doi: 10.1007/s10584-013-0803-7. Citado 5 vezes nas páginas 25, 26, 27, 28 e 58.

JUNK, J. *et al.* Effects of regional climate change on brown rust disease in winter wheat. *Clim chang*, v. 135, n. 3-4, p. 439–451, 2016. doi: 10.1007/s10584-015-1587-8. Citado 3 vezes nas páginas 17, 27 e 58.

KAMRUZZAMAN, M.; BEECHAM, S.; METCALFE, A. Non-stationarity in rainfall and temperature in the Murray Darling basin. *Hydrol Process*, v. 25, n. 10, p. 1659–1675, 2011. doi: 10.1002/hyp.7928. Citado na página 20.

KIM, K.-H.; CHO, J. Predicting potential epidemics of rice diseases in Korea using multi-model ensembles for assessment of climate change impacts with uncertainty information. *Clim chang*, v. 134, n. 1-2, p. 327–339, 2016. doi: 10.1007/s10584-015-1503-2. Citado na página 27.

- KOOL, D. *et al.* A review of approaches for evapotranspiration partitioning. *Agric For Meteorol*, v. 184, p. 56–70, 2014. doi: 10.1016/j.agrformet.2013.09.003. Citado 4 vezes nas páginas 23, 24, 38 e 39.
- KOS, J. *et al.* The influence of climate conditions on the occurrence of deoxynivalenol in maize harvested in serbia during 2013–2015. *Food Control*, v. 73, p. 734–740, 2017. doi: 10.1016/j.foodcont.2016.09.022. Citado 2 vezes nas páginas 17 e 27.
- KOS, J. *et al.* Presence of fusarium toxins in maize from autonomous province of vojvodina, serbia. *Food control*, v. 46, p. 98–101, 2014. doi: 10.1016/j.foodcont.2014.05.010. Citado na página 27.
- KOS, J. *et al.* Natural occurrence of aflatoxins in maize harvested in serbia during 2009–2012. *Food Control*, v. 34, n. 1, p. 31–34, 2013. doi: 10.1016/j.foodcont.2013.04.004. Citado na página 27.
- KRAUSE, P.; BOYLE, D.; BÄSE, F. Comparison of different efficiency criteria for hydrological model assessment. *Adv Geosci*, v. 5, p. 89–97, 2005. doi: 1680-7359/adgeo/2005-5-89. Citado 2 vezes nas páginas 49 e 56.
- KRISHNAN, A. *et al.* Partial least squares (pls) methods for neuroimaging: a tutorial and review. *Neuroimage*, v. 56, n. 2, p. 455–475, 2011. doi: 10.1016/j.neuroimage.2010.07.034. Citado na página 45.
- KUTTA, E.; HUBBART, J. A. Reconsidering meteorological seasons in a changing climate. *Clim Chang*, v. 137, n. 3-4, p. 511–524, 2016. doi: 10.1007/s10584-016-1704-3. Citado na página 21.
- LIU, L. *et al.* Effect of warming and nitrogen addition on evapotranspiration and water use efficiency in a wheat-soybean/fallow rotation from 2010 to 2014. *Clim Chang*, v. 139, n. 3-4, p. 565–578, 2016. doi: 10.1007/s10584-016-1825-8. Citado 2 vezes nas páginas 17 e 24.
- LOBELL, D. B.; SCHLENKER, W.; COSTA-ROBERTS, J. Climate trends and global crop production since 1980. *Science*, v. 333, n. 6042, p. 616–620, 2011. doi: 10.1126/science.1204531. Citado 4 vezes nas páginas 21, 23, 25 e 57.
- LÓPEZ-MORENO, J. *et al.* Impact of climate and land use change on water availability and reservoir management: Scenarios in the upper aragón river, spanish pyrenees. *Sci Total Environ*, v. 493, p. 1222–1231, 2014. doi: 10.1016/j.scitotenv.2013.09.031. Citado na página 17.
- MAFF. *Departamento de Agricultura dos grãos, tais como a oferta e a demanda dos EUA relatório*. 2017. Disponível em: <[http://www.maff.go.jp/j/zyukyu/jki/j\\_usda/index.html](http://www.maff.go.jp/j/zyukyu/jki/j_usda/index.html)>. Citado na página 57.
- MATEOS-APARICIO, G. Partial least squares (PLS) methods: Origins, evolution, and application to social sciences. *Commun Stat Theory Methods*, v. 40, n. 13, p. 2305–2317, 2011. doi: 10.1080/03610921003778225. Citado na página 45.
- MCAFEE, S. A. Methodological differences in projected potential evapotranspiration. *Climatic change*, v. 120, n. 4, p. 915–930, 2013. doi: 10.1007/s10584-013-0864-7. Citado na página 20.

- MEKASHA, A.; TESFAYE, K.; DUNCAN, A. J. Trends in daily observed temperature and precipitation extremes over three ethiopian eco-environments. *Int J Climatol*, v. 34, n. 6, p. 1990–1999, 2014. doi: 10.1002/joc.3816. Citado na página 20.
- MERELLES, L. R. de O. *et al.* Teste de aderência e poder do teste pela simulação de monte carlo: Teste de distribuição nos dados climáticos do brasil. In: *XLIX SBPO*. Campus da FURB: [s.n.], 2017. Citado na página 45.
- MONTEITH, J. Weather and water in the sudano-sahelian zone. In: *Soil Water Balance in the Sudano-Sahelian Zone Proceedings of the Niamey Workshop*. [S.l.: s.n.], 1991. Citado na página 39.
- MONTEITH, J. L. Evaporation and environment. In: *Symp Soc Exp Biol*. [S.l.: s.n.], 1965. v. 19, n. 205-23, p. 205–234. Citado 2 vezes nas páginas 38 e 39.
- MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. *Introduction to time series analysis and forecasting*. 2. ed. [S.l.]: John Wiley & Sons, 2015. 671p p. Citado 2 vezes nas páginas 42 e 44.
- MORETTIN, P. A.; TOLOI, C. *Análise de séries temporais*. 2. ed. [S.l.]: Blucher, 2011. 538p p. Citado 3 vezes nas páginas 42, 43 e 44.
- MORI, S.; KATO, M.; IDO, T. Gisela–gis-based evaluation of land use and agriculture market analysis under global warming. *Appl Energy*, v. 87, n. 1, p. 236–242, 2010. doi: 10.1016/j.apenergy.2009.06.013. Citado na página 25.
- MORIASI, D. N. *et al.* Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *T ASABE*, v. 50, n. 3, p. 885–900, 2007. doi: 10.13031/2013.23153. Citado 2 vezes nas páginas 49 e 56.
- MÜLLER, M. E.; KORN, U. Alternaria mycotoxins in wheat—a 10 years survey in the northeast of germany. *Food Control*, v. 34, n. 1, p. 191–197, 2013. doi: 10.1016/j.foodcont.2013.04.018. Citado 2 vezes nas páginas 17 e 27.
- NASH, J. E.; SUTCLIFFE, J. V. River flow forecasting through conceptual models part i – a discussion of principles. *Journal of hydrology*, v. 10, n. 3, p. 282–290, 1970. doi: 10.1016/0022-1694(70)90255-6. Citado na página 49.
- NOAA. *Data access*. 2017. Disponível em: <<https://www.ncdc.noaa.gov/>>. Citado na página 30.
- OLIVEIRA, M. S. *et al.* Natural mycotoxin contamination of maize (zea mays l.) in the south region of brazil. *Food Control*, v. 73, p. 127–132, 2017. doi: 10.1016/j.foodcont.2016.07.033. Citado 3 vezes nas páginas 17, 26 e 27.
- PATERSON, R. R. M.; LIMA, N.; TANIWAKI, M. H. Coffee, mycotoxins and climate change. *Food Res Int*, v. 61, p. 1–15, 2014. doi: 10.1016/j.foodres.2014.03.037. Citado 2 vezes nas páginas 27 e 58.
- PENMAN, H. L. Natural evaporation from open water, bare soil and grass. *Proc R Soc A*, v. 193, n. 1032, p. 120–145, 1948. doi: 10.1098/rspa.1948.0037. Citado na página 38.



- PEREIRA, L. S. *et al.* Crop evapotranspiration estimation with FAO56: Past and future. *Agric Water Manage*, v. 147, p. 4–20, 2015. doi: 10.1016/j.agwat.2014.07.031. Citado 3 vezes nas páginas 23, 25 e 39.
- PÉREZ-ENCISO, M.; TENENHAUS, M. Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (pls-da) approach. *Human genetics*, v. 112, n. 5-6, p. 581–592, 2003. doi: 10.1007/s00439-003-0921-9. Citado 3 vezes nas páginas 45, 46 e 48.
- PICCIRILLI, M. *R package to download, transform, analyze, and plot NOAA ISD weather data*. 2015. Disponível em: <<https://github.com/mpiccirilli/weatheR>>. Citado na página 30.
- RAYNER, N. *et al.* Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J Geophys Res Atmos*, v. 108, n. D14, 2003. doi: 10.1029/2002JD002670. Citado 2 vezes nas páginas 17 e 20.
- ROBERT, M. *et al.* Cmfdm: A methodology to guide the design of a conceptual model of farmers' decision-making processes. *Agric Syst*, v. 148, p. 86–94, 2016. doi: 10.1016/j.agry.2016.07.010. Citado na página 24.
- RODRIGO, F. Completing the early instrumental weather record from cádis (southern spain): new data from 1799 to 1803. *Clim Chang*, v. 111, n. 3-4, p. 697–704, 2012. doi: 10.1007/s10584-011-0174-x. Citado na página 33.
- ROMERO, C. C.; DUKES, M. D. Review of turfgrass evapotranspiration and crop coefficients. *T ASABE*, v. 59, n. 1, p. 207–223, 2016. doi: 10.13031/trans.59.11180. Citado 2 vezes nas páginas 37 e 38.
- ROSIPAL, R.; KRÄMER, N. Overview and recent advances in partial least squares. *Lecture notes in computer science*, v. 3940, p. 34–51, 2006. doi: 10.1007/11752790\_2. Citado na página 46.
- SAMANI, Z. Estimating solar radiation and evapotranspiration using minimum climatological data. *J Irrig Drain Eng*, v. 126, n. 4, p. 265–267, 2000. doi: 10.1061/(ASCE)0733-9437(2000)126:4(265). Citado 2 vezes nas páginas 41 e 56.
- SAUNDERS, M.; LEWIS, P.; THORNHILL, A. *Research Methods for Business Students*. 5. ed. [S.l.]: Pearson Education, 2009. Citado na página 21.
- SAVI, G. D. *et al.* Deoxynivalenol in the wheat milling process and wheat-based products and daily intake estimates for the southern brazilian population. *Food Control*, v. 62, p. 231–236, 2016. doi: 10.1016/j.foodcont.2015.10.029. Citado na página 26.
- SCHWARZ, G. Estimating the dimension of a model. *Ann Stat*, v. 6, n. 2, p. 461–464, 1978. doi: 10.1214/aos/1176344136. Citado na página 44.
- SERRANO-NOTIVOLI, R.; LUIS, M. de; BEGUERÍA, S. An r package for daily precipitation climate series reconstruction. *Environ Model Softw*, v. 89, p. 190–195, 2017. doi: 10.1016/j.envsoft.2016.11.005. Citado na página 33.
- SHUTTLEWORTH, W. J.; WALLACE, J. S. Calculating the water requirements of irrigated crops in australia using the matt-shuttleworth approach. *T ASABE*, v. 52, n. 6, p. 1895–1906, 2009. doi: 10.13031/2013.29217. Citado na página 39.

- SLONOSKY, V.; JONES, P.; DAVIES, T. Instrumental pressure observations and atmospheric circulation from the 17th and 18th centuries: London and Paris. *Int J Climatol*, v. 21, n. 3, p. 285–298, 2001. doi: 10.1002/joc.611. Citado na página 37.
- TACK, J. B.; HOLT, M. T. The influence of weather extremes on the spatial correlation of corn yields. *Clim Chang*, v. 134, n. 1-2, p. 299–309, 2016. doi: 10.1007/s10584-015-1538-4. Citado 2 vezes nas páginas 23 e 25.
- TATSCH, J. *Inmetr: A Package to Import Historical Data from Brazilian Meteorological Stations*. 2016. Disponível em: <<https://github.com/jdtatsch/inmetr>>. Citado na página 30.
- TENENHAUS, M. *La régression PLS: théorie et pratique*. Paris: Editions technip, 1998. Citado 2 vezes nas páginas 48 e 49.
- THORNTHWAITE, C. W. An approach toward a rational classification of climate. *Geogr Rev*, v. 38, n. 1, p. 55–94, 1948. doi: 10.2307/210739. Citado na página 38.
- TORGOVITSKI, L. A Darling–Erdős-type CUSUM-procedure for functional data. *Metrika*, v. 78, n. 1, p. 1–27, 2015. doi: 10.1007/s00184-014-0487-7. Citado 2 vezes nas páginas 17 e 20.
- TSENG, F.-M.; TZENG, G.-H. A fuzzy seasonal arima model for forecasting. *Fuzzy Sets and Systems*, v. 126, n. 3, p. 367–376, 2002. doi: 10.1016/S0165-0114(01)00047-1. Citado na página 43.
- TURC, L. Estimation of irrigation water requirements, potential evapotranspiration: a simple climatic formula evolved up to date. *Ann Agron*, v. 12, n. 1, p. 13–49, 1961. Citado na página 39.
- TWARUŽEK, M. *et al.* Statistical comparison of fusarium mycotoxins content in oat grain and related products from two agricultural systems. *Food control*, v. 34, n. 2, p. 291–295, 2013. doi: 10.1016/j.foodcont.2013.05.010. Citado na página 27.
- TWINE, T. E. *et al.* Impacts of elevated co2 concentration on the productivity and surface energy budget of the soybean and maize agroecosystem in the midwest USA. *Glob Change Biol*, v. 19, n. 9, p. 2838–2852, 2013. doi: 10.1111/gcb.12270. Citado 3 vezes nas páginas 23, 26 e 58.
- VERÓN, S. R.; ABELLEYRA, D. de; LOBELL, D. B. Impacts of precipitation and temperature on crop yields in the pampas. *Clim Chang*, v. 130, n. 2, p. 235–245, 2015. doi: 10.1007/s10584-015-1350-1. Citado 6 vezes nas páginas 17, 21, 25, 59, 68 e 71.
- VIANNA, L. F. d. N. *et al.* Meteorological databases: Metadata analysis of the weather station network in the santa catarina state, brazil. *Rev bras meteorol*, v. 32, n. 1, p. 53–64, 2017. doi: 10.1590/0102-778632120150119. Citado na página 33.
- VICENTE-SERRANO, S. M. *et al.* A complete daily precipitation database for northeast Spain: reconstruction, quality control, and homogeneity. *Int J Climatol*, v. 30, n. 8, p. 1146–1163, 2010. doi: 10.1002/joc.1850. Citado 5 vezes nas páginas 18, 31, 37, 52 e 54.
- VINCENT, L. A. *et al.* Observed trends in indices of daily temperature extremes in South America 1960–2000. *J Climate*, v. 18, n. 23, p. 5011–5023, 2005. doi: 10.1175/JCLI3589.1. Citado 2 vezes nas páginas 33 e 37.

- WALPOLE, R. E. *et al.* *Probability & Statistics for Engineers & Scientists, MyStatLab*. 2. ed. [S.l.]: Pearson Higher Ed, 2016. Citado 2 vezes nas páginas 37 e 49.
- WANG, H.; CHEN, Y.; LI, W. Characteristics in streamflow and extremes in the tarim river, china: trends, distribution and climate linkage. *Int J Climatol*, v. 35, n. 5, p. 761–776, 2015. doi: 10.1002/joc.4020. Citado 2 vezes nas páginas 17 e 60.
- WANG, Z. *et al.* Modeling the impacts of climate change on nitrogen losses and crop yield in a subsurface drained field. *Clim chang*, v. 129, n. 1-2, p. 323–335, 2015. doi: 10.1007/s10584-015-1342-1. Citado 2 vezes nas páginas 21 e 25.
- WEHRENS, R.; MEVIK, B.-H. The pls package: principal component and Partial Least Squares Regression in r. *J Stat Softw*, v. 18, n. 2, p. 1–23, 2007. doi: 10.18637/jss.v018.i02. Citado 2 vezes nas páginas 17 e 46.
- WILLMOTT, C. J.; MATSUURA, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, v. 30, n. 1, p. 79–82, 2005. doi: 24869236. Citado na página 50.
- WOLD, S. Personal memories of the early PLS development. *Chemometr Intell Lab*, v. 58, n. 2, p. 83 – 84, 2001. doi: 10.1016/S0169-7439(01)00152-6. Citado na página 45.
- WOLD, S. Chemometrics and Bruce: Some fond memories. In: *40 Years of Chemometrics—From Bruce Kowalski to the Future*. [S.l.]: ACS Publications, 2015. p. 1–13. Citado na página 45.
- WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. *Chemometr Intell Lab*, v. 58, n. 2, p. 109–130, 2001. doi: 10.1016/S0169-7439(01)00155-1. Citado 3 vezes nas páginas 17, 46 e 47.
- WOLDESENBET, T. A. *et al.* Hydrological responses to land use/cover changes in the source region of the upper blue nile basin, ethiopia. *Sci Total Environ*, v. 575, p. 724–741, 2017. doi: 10.1016/j.scitotenv.2016.09.124. Citado 2 vezes nas páginas 18 e 68.
- XING, W.; WANG, B.; YIM, S.-Y. Long-lead seasonal prediction of china summer rainfall using an EOF–PLS regression-based methodology. *J. Climate*, v. 29, n. 5, p. 1783–1796, 2016. doi: 10.1175/JCLI-D-15-0016.1. Citado na página 18.
- YANG, Y.; XU, X. Post-disaster grain supply chain resilience with government aid. *Transport Res E Log*, v. 76, p. 139–159, 2015. doi: 10.1016/j.tre.2015.02.007. Citado na página 17.
- YILMAZ, K. Comparison of quantitative and qualitative research traditions: Epistemological, theoretical, and methodological differences. *European Journal of Education*, v. 48, n. 2, p. 311–325, 2013. doi: 10.1111/ejed.12014. Citado na página 29.
- YUN, Y.-H. *et al.* A simple idea on applying large regression coefficient to improve the genetic algorithm-pls for variable selection in multivariate calibration. *Chemom Intell Lab Syst*, v. 130, p. 76–83, 2014. doi: 10.1016/j.chemolab.2013.09.007. Citado na página 45.
- ZHANG, M. *et al.* Forecasting the transport energy demand based on pls method in china. *Energy*, v. 34, n. 9, p. 1396–1400, 2009. doi: 10.1016/j.energy.2009.06.032. Citado 2 vezes nas páginas 18 e 32.

## **Apêndices**

## APÊNDICE A – TESTE DO MELHOR MÉTODO PARA COMPLETAR OS DADOS

Tabela 17 – Valores de p-valor do teste de Tukey para a diferença entre o real e o previsto nos pontos extremos e a frequência dos dados climáticos

Em comum	Variáveis climáticas	Métodos		
		Interpolação	Regressão	Vizinho
1%	Freq. Precipitação (dia úmido)	0,000	0,000	<b>0,959</b>
	Freq. Precipitação (dia seco)	0,000	<b>0,999</b>	<b>0,999</b>
	Temperatura média - alta	0,949	<b>0,999</b>	0,967
	Temperatura média - baixa	0,781	<b>0,993</b>	0,046
	Temperatura máxima - alta	0,973	0,931	<b>0,974</b>
	Temperatura máxima - baixa	<b>0,999</b>	0,997	0,471
	Temperatura mínima - alta	0,791	0,777	<b>0,999</b>
	Temperatura mínima - baixa	0,204	<b>0,891</b>	0,004
	Umidade relativa - alta	0,079	0,591	<b>0,951</b>
	Umidade relativa - baixa	<b>0,956</b>	0,910	0,255
5%	Freq. Precipitação (dia úmido)	0,000	0,000	<b>0,979</b>
	Freq. Precipitação (dia seco)	0,000	<b>0,999</b>	<b>0,999</b>
	Temperatura média - alta	0,974	<b>0,991</b>	0,769
	Temperatura média - baixa	<b>0,986</b>	0,903	0,000
	Temperatura máxima - alta	<b>0,999</b>	0,990	0,470
	Temperatura máxima - baixa	<b>0,908</b>	0,654	0,002
	Temperatura mínima - alta	0,990	<b>0,995</b>	0,762
	Temperatura mínima - baixa	0,113	<b>0,835</b>	0,002
	Umidade relativa - alta	0,006	0,793	<b>0,984</b>
	Umidade relativa - baixa	<b>0,862</b>	0,783	0,130
10%	Freq. Precipitação (dia úmido)	0,000	0,000	<b>0,804</b>
	Freq. Precipitação (dia seco)	0,000	<b>0,999</b>	<b>0,999</b>
	Temperatura média - alta	<b>0,984</b>	0,992	0,684
	Temperatura média - baixa	0,545	<b>0,927</b>	0,000
	Temperatura máxima - alta	<b>0,999</b>	0,977	0,315

Continua na próxima página

Tabela 17 – Continuação da página anterior

Em comum	Variáveis climáticas	Métodos		
		Interpolação	Regressão	Vizinho
	Temperatura máxima - baixa	<b>0,943</b>	0,729	0,003
	Temperatura mínima - alta	0,992	<b>0,999</b>	0,614
	Temperatura mínima - baixa	0,077	<b>0,595</b>	0,000
	Umidade relativa - alta	0,005	0,693	<b>0,998</b>
	Umidade relativa - baixa	<b>0,658</b>	0,346	0,004
15%	Freq. Precipitação (dia úmido)	0,000	0,000	<b>0,600</b>
	Freq. Precipitação (dia seco)	0,000	<b>0,999</b>	<b>0,999</b>
	Temperatura média - alta	<b>0,999</b>	<b>0,999</b>	0,526
	Temperatura média - baixa	0,451	<b>0,908</b>	0,000
	Temperatura máxima - alta	<b>0,999</b>	0,975	0,516
	Temperatura máxima - baixa	<b>0,836</b>	0,680	0,000
	Temperatura mínima - alta	0,997	<b>0,999</b>	0,428
	Temperatura mínima - baixa	0,083	<b>0,752</b>	0,000
	Umidade relativa - alta	0,001	0,768	<b>0,999</b>
	Umidade relativa - baixa	<b>0,642</b>	0,520	0,014
20%	Freq. Precipitação (dia úmido)	0,000	0,000	<b>0,986</b>
	Freq. Precipitação (dia seco)	0,000	<b>0,999</b>	<b>0,999</b>
	Temperatura média - alta	0,976	<b>0,999</b>	0,640
	Temperatura média - baixa	0,546	<b>0,882</b>	0,000
	Temperatura máxima - alta	<b>0,999</b>	0,983	0,326
	Temperatura máxima - baixa	<b>0,955</b>	0,705	0,001
	Temperatura mínima - alta	<b>0,999</b>	<b>0,999</b>	0,452
	Temperatura mínima - baixa	0,109	<b>0,568</b>	0,000
	Umidade relativa - alta	0,001	0,710	<b>0,999</b>
	Umidade relativa - baixa	<b>0,634</b>	0,217	0,005

A hipótese para o teste de Tukey é

$$H_o : \mu = \mu_j, \text{ versus } H_a : \mu \neq \mu_j,$$

em que:  $\mu$  é a média de dados reais  $\{x_1, x_2, \dots, x_n\}$  onde  $X \subset \mathbb{R}^q$ ;  $\mu_j$  é a média dos dados completados pelo método  $j$  ( $j =$  interpolação, regressão e vizinhos). A veracidade da hipótese  $H_o$  é refutada com um nível de significância  $\alpha < 0,05$ .

## APÊNDICE B – TESTE DE ADERÊNCIA

Tabela 18 – Teste de aderência para os dados climáticos por estação do ano para região Centro-Oeste

Distribuição	Normal	Log-normal	Uniforme	Log-uniforme	Exponencial	Gamma	Weibull
$P_{ad}:1$	0,504	0,278	0,362	0,050	0,000	0,345	0,631
$P_{ad}:2$	0,906	0,611	0,110	0,000	0,000	0,841	0,933
$P_{ad}:3$	0,144	0,749	0,008	0,005	0,212	0,676	0,596
$P_{ad}:4$	0,288	0,469	0,014	0,319	0,000	0,470	0,240
$T_{med}:1$	0,490	0,497	0,150	0,177	0,000	0,499	0,637
$T_{med}:2$	0,878	0,899	0,283	0,329	0,000	0,893	0,440
$T_{med}:3$	0,724	0,703	0,001	0,001	0,000	0,720	0,191
$T_{med}:4$	0,678	0,723	0,000	0,000	0,000	0,700	0,201
$T_{min}:1$	0,973	0,977	0,067	0,032	0,000	0,979	0,712
$T_{min}:2$	0,526	0,504	0,852	0,697	0,000	0,513	0,467
$T_{min}:3$	0,421	0,328	0,602	0,456	0,000	0,331	0,469
$T_{min}:4$	0,332	0,394	0,206	0,356	0,000	0,383	0,227
$T_{max}:1$	0,933	0,953	0,704	0,801	0,000	0,948	0,657
$T_{max}:2$	0,507	0,539	0,070	0,090	0,000	0,533	0,318
$T_{max}:3$	0,554	0,518	0,820	0,878	0,000	0,527	0,625
$T_{max}:4$	0,832	0,829	0,237	0,329	0,000	0,834	0,723
$T_{dtr}:1$	0,589	0,457	0,435	0,179	0,000	0,486	0,650
$T_{dtr}:2$	0,133	0,114	0,392	0,326	0,000	0,112	0,233
$T_{dtr}:3$	0,356	0,259	0,161	0,051	0,000	0,269	0,484
$T_{dtr}:4$	0,256	0,166	0,140	0,043	0,000	0,184	0,300
$UR_{med}:1$	0,959	0,934	0,017	0,009	0,000	0,947	0,920
$UR_{med}:2$	0,922	0,881	0,237	0,151	0,000	0,894	0,977
$UR_{med}:3$	0,966	0,916	0,305	0,168	0,000	0,938	0,864
$UR_{med}:4$	0,742	0,660	0,041	0,019	0,000	0,698	0,971
$UR_{min}:1$	0,602	0,515	0,926	0,662	0,000	0,537	0,585
$UR_{min}:2$	0,331	0,360	0,604	0,367	0,000	0,384	0,127
$UR_{min}:3$	0,571	0,533	0,665	0,451	0,000	0,538	0,428
$UR_{min}:4$	0,456	0,394	0,397	0,515	0,000	0,409	0,648
$UR_{max}:1$	0,100	0,104	0,089	0,109	0,000	0,104	0,078
$UR_{max}:2$	0,450	0,455	0,405	0,476	0,000	0,455	0,368
$UR_{max}:3$	0,303	0,271	0,383	0,244	0,000	0,278	0,126
$UR_{max}:4$	0,231	0,225	0,386	0,304	0,000	0,224	0,200
$UR_{dtr}:1$	0,726	0,778	0,328	0,513	0,000	0,770	0,456
$UR_{dtr}:2$	0,798	0,849	0,031	0,050	0,000	0,831	0,315
$UR_{dtr}:3$	0,853	0,842	0,252	0,130	0,000	0,833	0,901
$UR_{dtr}:4$	0,369	0,430	0,000	0,001	0,000	0,419	0,364
$ET_o:1$	0,758	0,783	0,657	0,898	0,000	0,796	0,531
$ET_o:2$	0,622	0,591	0,406	0,222	0,000	0,600	0,410
$ET_o:3$	0,697	0,784	0,026	0,061	0,000	0,757	0,353
$ET_o:4$	0,384	0,328	0,458	0,756	0,000	0,348	0,457
$CO_2:1$	0,900	0,938	0,767	0,932	0,000	0,930	0,684
$CO_2:2$	0,891	0,931	0,734	0,912	0,000	0,922	0,682
$CO_2:3$	0,869	0,905	0,635	0,835	0,000	0,899	0,640
$CO_2:4$	0,869	0,906	0,715	0,901	0,000	0,899	0,678

Tabela 19 – Teste de aderência para os dados climáticos por estação do ano para região Sul

Distribuição	Normal	Log-normal	Uniforme	Log-uniforme	Exponencial	Gamma	Weibull
$P_{ad}:1$	0,225	0,561	0,000	0,070	0,000	0,460	0,149
$P_{ad}:2$	0,894	0,927	0,046	0,316	0,000	0,927	0,944
$P_{ad}:3$	0,766	0,336	0,331	0,000	0,001	0,602	0,889
$P_{ad}:4$	0,405	0,795	0,120	0,001	0,000	0,826	0,414
$T_{med}:1$	0,905	0,910	0,539	0,617	0,000	0,910	0,596
$T_{med}:2$	0,516	0,564	0,000	0,000	0,000	0,529	0,078
$T_{med}:3$	0,977	0,991	0,043	0,077	0,000	0,987	0,534
$T_{med}:4$	0,909	0,924	0,034	0,049	0,000	0,927	0,680
$T_{min}:1$	0,629	0,478	0,005	0,000	0,000	0,585	0,316
$T_{min}:2$	0,806	0,003	0,010	0,000	0,138	0,028	0,058
$T_{min}:3$	0,310	0,002	0,003	0,000	0,003	0,032	0,023
$T_{min}:4$	0,680	0,961	0,006	0,130	0,000	0,904	0,416
$T_{max}:1$	0,998	0,997	0,547	0,413	0,000	0,998	0,951
$T_{max}:2$	0,827	0,819	0,362	0,253	0,000	0,818	0,502
$T_{max}:3$	0,702	0,634	0,116	0,059	0,000	0,638	0,899
$T_{max}:4$	0,589	0,542	0,309	0,404	0,000	0,553	0,918
$T_{dtr}:1$	0,244	0,121	0,021	0,004	0,000	0,159	0,189
$T_{dtr}:2$	0,734	0,496	0,002	0,000	0,000	0,594	0,981
$T_{dtr}:3$	0,356	0,203	0,004	0,000	0,000	0,244	0,867
$T_{dtr}:4$	0,104	0,065	0,000	0,000	0,000	0,072	0,466
$UR_{med}:1$	0,413	0,367	0,003	0,001	0,000	0,384	0,563
$UR_{med}:2$	0,786	0,750	0,009	0,014	0,000	0,763	0,391
$UR_{med}:3$	0,454	0,490	0,000	0,000	0,000	0,466	0,078
$UR_{med}:4$	0,279	0,320	0,000	0,001	0,000	0,304	0,066
$UR_{min}:1$	0,191	0,317	0,001	0,007	0,000	0,271	0,058
$UR_{min}:2$	0,054	0,105	0,000	0,002	0,000	0,082	0,017
$UR_{min}:3$	0,019	0,027	0,000	0,000	0,000	0,022	0,027
$UR_{min}:4$	0,142	0,270	0,000	0,000	0,000	0,232	0,028
$UR_{max}:1$	0,237	0,247	0,161	0,194	0,000	0,247	0,134
$UR_{max}:2$	0,195	0,206	0,156	0,187	0,000	0,205	0,101
$UR_{max}:3$	0,147	0,142	0,103	0,086	0,000	0,145	0,102
$UR_{max}:4$	0,145	0,151	0,151	0,188	0,000	0,150	0,111
$UR_{dtr}:1$	0,241	0,155	0,000	0,000	0,000	0,192	0,549
$UR_{dtr}:2$	0,193	0,115	0,000	0,000	0,000	0,155	0,768
$UR_{dtr}:3$	0,110	0,068	0,013	0,002	0,000	0,083	0,404
$UR_{dtr}:4$	0,486	0,366	0,002	0,000	0,000	0,428	0,905
$ET_o:1$	0,380	0,527	0,112	0,086	0,000	0,471	0,123
$ET_o:2$	0,524	0,579	0,220	0,100	0,000	0,573	0,335
$ET_o:3$	0,935	0,963	0,010	0,038	0,000	0,964	0,697
$ET_o:4$	0,132	0,220	0,001	0,010	0,000	0,187	0,049
$CO_2:1$	0,900	0,938	0,767	0,932	0,000	0,930	0,684
$CO_2:2$	0,891	0,931	0,734	0,912	0,000	0,922	0,682
$CO_2:3$	0,869	0,905	0,635	0,835	0,000	0,899	0,640
$CO_2:4$	0,869	0,906	0,715	0,901	0,000	0,899	0,678

Nota: “1” corresponde ao verão; “2” corresponde ao outono; “3” corresponde ao inverno; e “4” corresponde a primavera.



**APÊNDICE C – ARTIGOS PUBLICADOS**

DIAS, V. de S.; DANTAS, M. J. P.; MERELLES, L.; LUZ, M. Identificação de gargalos em sistema de manufatura aplicando modelagem e simulação. In: *XLVIII SBPO*. Campus da UFES: [s.n.], 2016.

ZALUSKI, P.R.S.; ROCHA, D.S.G.M.; MERELLES, L.R.O.; MACHADO, R.L. Diferenças entre lean manufacturing e green supply chain–revisão sistemática. *Revista Gestão, Inovação e Negócios*, v. 2, n. 2, p. 72-84, 2016.

MERELLES, L. R. de O.; DANTAS, M. J. P.; MENEZES, J. E. de; DIAS, V. de S. Aplicação do controle de qualidade para determinar o período de aquecimento em simulação de filas. In: *XLIX SBPO*. Campus da FURB: [s.n.], 2017.

MERELLES, L. R. de O.; DANTAS, M. J. P.; MENEZES, J. E. de; DIAS, V. de S. Teste de aderência e poder do teste pela simulação de monte carlo: Teste de distribuição nos dados climáticos do brasil. In: *XLIX SBPO*. Campus da FURB: [s.n.], 2017.

**APÊNDICE D – *SCRIPT* EM R**

# APÊNDICE A – *Script em R*

## *Modeling grains yield by time-series*

```

path <- c("F:/Users/Leonardo/Desktop/pls/dados")

# function to read and adjust data
#' @rdname load.dados
#' @export
load.dados <- function(idx) {
  if(missing(idx))
    stop("'idx' must be a non-empty numeric. it must be region code.")
  x <- read.table(file = paste0(path, "/R", idx, "/mud_cli.csv"),
                 header = TRUE, sep = ";")
  Y <- read.table(file = paste0(path, "/R", idx, "/prod.csv"),
                 header = TRUE, sep = ";")

  Y <- Y[, -1]
  X <- NULL
  k <- ncol(x)
  name <- colnames(x)
  for(k in seq_len(k)) {
    a <- matrix(x[,k], ncol = 4, byrow = TRUE)
    colnames(a) <- paste(name[k], 1:4, sep = ":")
    X <- cbind(X, a)
  }
  rownames(X) <- rownames(Y) <- c(1980:2016)
  return(list(X, Y))
}

# package of forecast
require(forecast)

## Loading required package: forecast

# pacote whit dickey-fuller test
require(urca)

## Loading required package: urca

# function for time series automation
#'@rdname fitting.ts
#'@export
#'@import tseries
#'@import forecast

```

```

#@import urca
#@importFrom stats shapiro.test
#@importFrom stats Box.test
#@importFrom tseries jarque.bera.test
fitting.ts <- function(y) {
  Yt = ts(y, start = 1980, freq = 1)
  y = diff(Yt)

  # dickey-fuller test
  dickey.fuller <- vector("list", 3)
  names(dickey.fuller) <- c("none", "drift", "trend")

  dickey.fuller$none <- summary(ur.df(y,type = "none", lags = 0))
  dickey.fuller$drift <- summary(ur.df(y, type = "drift", lags = 0))
  dickey.fuller$trend <- summary(ur.df(y, type = "trend", lags = 0))

  # time-series modeling
  time.series <- vector("list", 3)
  names(time.series) <- c("model", "confint", "forecast")

  time.series$model <- auto.arima(Yt)
  time.series$confint <- confint(time.series$model)
  time.series$forecast <- forecast::forecast(time.series$model, 5)

  # waste test
  waste.test <- vector("list", 4)
  names(waste.test) <- c("Box-Pierce", "Ljung-Box",
                        "Shapiro-Wilk", "jarque-bera")
  waste <- c(time.series$model$residuals)

  # Box-Pierce and Ljung-Box test
  waste.test$`Box-Pierce` <- Box.test(waste, type='Box-Pierce')
  waste.test$`Ljung-Box` <- Box.test(waste,lag=1,type='Ljung-Box')

  # Shapiro-Wilk and Jarque-Bera test
  waste.test$`Shapiro-Wilk` <- shapiro.test(waste)
  waste.test$`jarque-bera` <- tseries::jarque.bera.test(waste)

  out <- list(dickey.fuller, time.series, waste.test)
  names(out) <- c("dickey fuller", "time series", "waste test")
  return(out)
}

# reading data from the Midwest region

```

```

Y_co <- load.dados(5)[[2]]
Y_co <- as.matrix(Y_co[,c(1:2), drop = FALSE])

# reading data from the south region
Y_s = load.dados(4)[[2]]
Y_s <- as.matrix(Y_s[,c(1:2), drop = FALSE])

# reserved data to train and test models
train <- c(01:32)
teste <- c(33:37)

output.co <- output.sul <- vector("list", 2)
names(output.co) <- names(output.sul) <- c("soja", "milho")

# here, one refers to soy and two to corn
for(grain in 1:2) {
  output.co[[grain]] <- fitting.ts(Y_co[train, grain])
  output.sul[[grain]] <- fitting.ts(Y_s[train, grain])
}

```

### *Modeling grains yield by partial-least-square regression*

```

# function to remove the transformation of the data and make predictions
#' @rdname forecast
#' @export
forecast <- function(out, X, Y) {
  k <- ncol(X)
  n <- nrow(X)
  m <- ncol(Y)
  beta <- out$reg.coefs[1:k, ]
  cte <- out$reg.coefs[(k + 1), ]
  cte <- matrix(rep(cte, each = n), n, m)
  return(X %*% beta + cte)
}

# reading data from the south region
rg_s <- load.dados(4)
X <- rg_s[[1]]
Y <- rg_s[[2]]
Y <- as.matrix(Y[, c(1, 2), drop = FALSE])

# these rows correspond to the years 1992, 1995 and 1999
idx = c(13, 16, 20)

# south region modeling

```

```

out <- plsdepot::plsreg2(X[train[-idx], ], Y[train[-idx], ])
nc <- ncol(out$VIP)
inc <- out$VIP[,nc] >= 1
out <- plsdepot::plsreg2(X[train[-idx], inc], Y[train[-idx], ])
y.hat <- forecast(out, X[, inc], Y)
colnames(y.hat) <- paste(c("soja", "milho"), "hat", sep = "-")
y2 <- cbind(y.hat, Y)

# reading data from the Midwest region
rg_co <- load.dados(5)
X <- rg_co[[1]]
Y <- rg_co[[2]]
Y <- as.matrix(Y[, c(1, 2), drop = FALSE])

# these rows correspond to the years 1983, 1995 e 2011
idx <- c(4, 16, 32)

# Midwest region modeling
out <- plsdepot::plsreg2(X[train[-idx], ], Y[train[-idx], ])
nc <- ncol(out$VIP)
inc <- out$VIP[,nc] >= 1
out <- plsdepot::plsreg2(X[train[-idx], inc], Y[train[-idx], ])
y.hat <- forecast(out, X[,inc], Y)
colnames(y.hat) <- paste(c("soja", "milho"), "hat", sep = "-")
y1 <- cbind(y.hat, Y)

```

### Goodness-of-Fit test

```

#' @rdname dlunif
#' @export
plunif <- function(q, min, max, base=exp(1)) {
  if(mode(q) != "numeric")
    stop("'q' must be a non-empty numeric vector")
  if(any(missing(min), missing(max)))
    stop("'min' and 'max' not provided, without default.\n")
  return(1/(log(max,base)-log(min,base)) * (log(q,base)-log(min,base)))
}

#' @rdname rlunif
#' @export
rlunif <- function(n, min, max, base=exp(1)) {
  if(mode(n) != "numeric")
    stop("'n' must be a non-empty numeric vector")
  if(any(missing(min), missing(max)))
    stop("'min' and 'max' not provided, without default.\n")
}

```

```

    ifelse(base==exp(1), return(exp(runif(n, log(min, base), log(max, base)))),
           return(base^(runif(n, log(min, base), log(max, base)))))
  }

#' Goodness-Of-Fit Test - Kolmogorov-Smirnoff
#' @rdname GoF.ks
#' @export
#' @import mclust
#' @import MASS
#' @importFrom stats sd
#' @importFrom stats var
#' @importFrom stats cor
#' @importFrom stats pnorm
#' @importFrom stats rnorm
#' @importFrom stats rlnorm
#' @importFrom stats runif
#' @importFrom stats rweibull
#' @importFrom stats rexp
#' @importFrom stats rgamma
#' @importFrom stats ks.test
GoF.ks <- function(x, nreps = 249) {
  if (missing(x) || length(x) == 0L || mode(x) != "numeric")
    stop("'x' must be a non-empty numeric vector")
  if (any(!is.finite(x)))
    stop("'x' contains missing or infinite values")
  if (nreps < 1)
    stop("'nreps' must be a positive integer")
  allowed <- c("pnorm", "plnorm", "punif",
              "plunif", "pexp", "pgamma", "pweibull")
  n <- length(x)
  D.sim <- D.obs <- p.value <- p.sim <- vector("list", length(allowed))
  names(p.value) <- names(D.obs) <- names(D.sim) <- names(p.sim) <- allowed

  # normal distribution
  mean.x <- mean(x)
  sd.x <- sd(x)
  test <- ks.test(x, "pnorm", mean=mean.x, sd=sd.x)
  p.value$pnorm <- as.vector(test$p.value)
  D.obs$pnorm <- as.vector(test$statistic)
  x.sim <- matrix(rnorm(n*nreps, mean=mean.x, sd=sd.x), nc=nreps)
  D.sim$pnorm <- sapply(1:nreps, function(i)
    as.vector(ks.test(x.sim[,i], "pnorm", mean=mean.x, sd=sd.x)$statistic) )
  if(mode(D.obs$pnorm) == "numeric" | mode(p.value$pnorm) == "numeric" |
     mode(D.sim$pnorm) == "numeric")

```

```

p.sim$pnorm <- (sum(D.sim$pnorm > D.obs$pnorm) + 1) / (nreps + 1)

# log-normal distribution
if (!any(x <= 0)){
  meanlog.x <- mean(log(x))
  sdlog.x <- sd(log(x))
  test <- ks.test(x, "plnorm", meanlog=meanlog.x, sdlog=sdlog.x)
  p.value$plnorm <- as.vector(test$p.value)
  D.obs$plnorm <- as.vector(test$statistic)
  x.sim <- matrix(rlnorm(n*nreps, meanlog=meanlog.x,
                        sdlog=sdlog.x), nc=nreps)
  D.sim$plnorm <- sapply(1:nreps, function(i)
    as.vector(ks.test(x.sim[,i], "plnorm", meanlog=meanlog.x,
                    sdlog=sdlog.x)$statistic) )
  if(mode(D.obs$plnorm) == "numeric" | mode(p.value$plnorm) == "numeric" |
      mode(D.sim$plnorm) == "numeric")
    p.sim$plnorm <- (sum(D.sim$plnorm > D.obs$plnorm) + 1) / (nreps + 1)
} else {
  cat("'x' values must be > 0 for lognormal distributions.\n")
  D.obs$plnorm <- D.sim$plnorm <- p.sim$plnorm <- p.value$plnorm <-
  c("insignificant")
}

# uniform distribution
min.x <- min(x)
max.x <- max(x)
test <- ks.test(x, "punif", min=min.x, max=max.x)
p.value$punif <- as.vector(test$p.value)
D.obs$punif <- as.vector(test$statistic)
x.sim <- matrix(runif(n*nreps, min=min.x, max=max.x), nc=nreps)
D.sim$punif <- sapply(1:nreps, function(i)
  as.vector(ks.test(x.sim[,i], "punif", min=min.x, max=max.x)$statistic) )
if(mode(D.obs$punif) == "numeric" | mode(p.value$punif) == "numeric" |
    mode(D.sim$punif) == "numeric")
  p.sim$punif <- (sum(D.sim$punif > D.obs$punif) + 1) / (nreps + 1)

# log-uniform distribution
if (!any(x <= 0)){
  min.x <- min(x)
  max.x <- max(x)
  test <- ks.test(x, "plunif", min=min.x, max=max.x)
  p.value$plunif <- as.vector(test$p.value)
  D.obs$plunif <- as.vector(test$statistic)
  x.sim <- matrix(rlnunif(n*nreps, min=min.x, max=max.x), nc=nreps)

```



```

D.sim$plunif <- sapply(1:nreps, function(i)
  as.vector(ks.test(x.sim[,i], "plunif", min=min.x, max=max.x)$statistic) )
if(mode(D.obs$plunif) == "numeric" | mode(p.value$plunif) == "numeric" |
  mode(D.sim$plunif) == "numeric")
  p.sim$plunif <- (sum(D.sim$pexp > D.obs$pexp) + 1) / (nreps + 1)
}else{
  cat("'x' values must be > 0 for loguniform distributions.\n")
  D.obs$plunif <- D.sim$plunif <- p.sim$plunif <- p.value$plunif <-
    c("insignificant")
}

# exponential distribution
if (!any(x <= 0)){
  rate.x <- 1 / mean(x)
  test <- ks.test(x, "pexp", rate=rate.x)
  p.value$pexp <- as.vector(test$p.value)
  D.obs$pexp <- as.vector(test$statistic)
  x.sim <- matrix(rexp(n*nreps, rate=rate.x), nc=nreps)
  D.sim$pexp <- sapply(1:nreps, function(i)
    as.vector(ks.test(x.sim[,i], "pexp", rate=rate.x)$statistic) )
  if(mode(D.obs$pexp) == "numeric" | mode(p.value$pexp) == "numeric" |
    mode(D.sim$pexp) == "numeric")
    p.sim$pexp <- (sum(D.sim$pexp > D.obs$pexp) + 1) / (nreps + 1)
}else{
  cat("'x' values must be > 0 for exponential distributions.\n")
  D.obs$pexp <- D.sim$pexp <- p.sim$pexp <- p.value$pexp <-
    c("insignificant")
}

# gamma distribution
if (!any(x <= 0)){
  shape.x <- mean(x)^2 / var(x)
  scale.x <- var(x) / mean(x)
  # Re-scale in case the scale is much larger than 1:
  param <- as.vector(suppressWarnings(
    MASS::fitdistr(x/scale.x, densfun="gamma",
      start = list(shape=shape.x, scale=scale.x/scale.x),
      control=list(maxit=25000))$estimate))
  param[2] <- param[2] * scale.x # Re-scale back
  test <- ks.test(x, "pgamma", shape=param[1], scale=param[2])
  p.value$pgamma <- as.vector(test$p.value)
  D.obs$pgamma <- as.vector(test$statistic)
  x.sim <- matrix(rgamma(n*nreps, shape=param[1], scale=param[2]), nc=nreps)
  D.sim$pgamma <- sapply(1:nreps, function(i)

```

```

    as.vector(ks.test(x.sim[,i], "pgamma",
                     shape=param[1], scale=param[2])$statistic) )
  if(mode(D.obs$pgamma) == "numeric" | mode(p.value$pgamma) == "numeric" |
      mode(D.sim$pgamma) == "numeric")
    p.sim$pgamma <- (sum(D.sim$pgamma > D.obs$pgamma) + 1) / (nreps + 1)
  }else{
    cat("'x' values must be > 0 for gamma distributions.\n")
    D.obs$pgamma <- D.sim$pgamma <- p.sim$pgamma <- p.value$pgamma <-
      c("insignificant")
  }

# weibull distribution
if (!any(x <= 0)){
  shape.x <- 1.2 / sd(log(x))
  scale.x <- exp(mean(log(x)) + 0.572/shape.x)
  # Re-scale in case the scale is much larger than 1:
  param <- as.vector(suppressWarnings(
    MASS::fitdistr(x/scale.x, densfun="weibull",
                  start = list(shape=shape.x, scale=scale.x/scale.x),
                  control=list(maxit=25000))$estimate))
  param[2] <- param[2] * scale.x      # Re-scale back
  test <- ks.test(x, "pweibull", shape=param[1], scale=param[2])
  p.value$pweibull <- as.vector(test$p.value)
  D.obs$pweibull <- as.vector(test$statistic)
  x.sim <- matrix(rweibull(n*nreps, shape=param[1], scale=param[2]),nc=nreps)
  D.sim$pweibull <- sapply(1:nreps, function(i)
    as.vector(ks.test(x.sim[,i], "pweibull",
                     shape=param[1], scale=param[2])$statistic) )
  if(mode(D.obs$pweibull) == "numeric" | mode(p.value$pweibull) == "numeric" |
      mode(D.sim$pweibull) == "numeric")
    p.sim$pweibull <- (sum(D.sim$pweibull > D.obs$pweibull) + 1) / (nreps + 1)
  }else{
    cat("'x' values must be > 0 for Weibull distributions.\n")
    D.obs$pweibull <- D.sim$pweibull <- p.sim$pweibull <- p.value$pweibull <-
      c("insignificant")
  }
}
out <- list("D.obs"=D.obs, "D.sim"=D.sim, "p.value"=p.value, "p.sim"=p.sim)
return(out)
}

# selecting the midwest region 'five'
# here, 'one' refers to data matrix from climate change
x = load.dados(5)[[1]]
teste_aderencia = matrix(NA,nr=7,nc=44)

```

```
colnames(teste_aderencia) = colnames(x)

# loop for automation of the 'Goodness-of-Fit' test for all samples
for(i in 1:ncol(x)) {
  out = GoF.ks(x[, i])
  teste_aderencia[,i] = simplify2array(out$p.value)
  if(i == 1)
    rownames(teste_aderencia) <- names(out$p.value)
}

# saving the results
teste_aderencia <- round(teste_aderencia, 3)
ofname <- paste0(path, "/teste de aderencia.csv")
write.table(teste_aderencia, ofname, sep = ";")
```