UNIVERSIDADE DE SÃO PAULO
Escola de Engenharia de São Carlos

# Development of index insurance under the dynamics of multi-hazard risk in the context of climate change

Aluno: Marcos Roberto Benso

Orientador: Prof. Dr. Eduardo Mario Mendiondo

EESC • USP

**Marcos Roberto Benso**

# Development of index insurance under the dynamics of multi-hazard risk in the context of climate change

Tese apresentada ao Departamento de Hidráulica e Saneamento da Escola de Engenharia de São Carlos da Universidade de São Paulo como requisito parcial para a obtenção do título de Doutor em Ciências.

Área de Concentração: Hidráulica e Saneamento

Orientador:
Prof. Dr. **EDUARDO MARIO MENDIONDO**

CORRECTED VERSION

**São Carlos**
**2023**

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da EESC/USP com os dados inseridos pelo(a) autor(a).

Benso, Marcos Roberto
B474d    Desenvolvimento de seguro indexado sob a dinâmica do risco multi-ameaça no contexto das mudanças climáticas / Marcos Roberto Benso; orientador Eduardo Mario Mendiondo. São Carlos, 2023.

Tese (Doutorado) - Programa de Pós-Graduação em Engenharia Hidráulica e Saneamento e Área de Concentração em Hidráulica e Saneamento -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2023.

1. Adaptação climática. 2. Gestão de risco. 3. Eventos extremos. 4. Produção de alimentos. I. Título.

Eduardo Graziosi Silva - CRB - 8/8907

## FOLHA DE JULGAMENTO

Candidato: Engenheiro **MARCOS ROBERTO BENSO**.

Título da tese: "Desenvolvimento de seguro indexado sob a dinâmica do risco multiameaça no contexto das mudanças climáticas. ".

Data da defesa: 01/12/2023.

| Comissão Julgadora | Resultado |
| --- | --- |
| **Prof. Dr. Eduardo Mario Mendiondo** **(Orientador)** (Escola de Engenharia de São Carlos/EESC-USP) | Aprovado |
| **Dra. Ana Paula Martins do Amaral Cunha** (Centro Nacional de Monitoramento e Alertas de Desastres Naturais/CEMADEN) | Aprovado |
| **Profa. Dra. Silvia Helena Galvão de Miranda** (Escola Superior de Agricultura Luiz de Queiroz/ESALQ-USP) | Aprovado |
| **Dr. Eduardo Monteiro Gelcer** (IRB Brasil RE) | Aprovado |
| **Profa. Dra. Conceição de Maria Albuquerque Alves** (Universidade de Brasilia/UnB) | Aprovado |

Coordenador do Programa de Pós-Graduação em Engenharia Hidráulica e Saneamento:
Prof. Assoc. **Juliano Jose Corbi**

Presidente da Comissão de Pós-Graduação:
Prof. Titular **Carlos De Marqui Junior**

## ACKNOWLEDGEMENTS

First, I would like to express my sincere gratitude to my supervisor, Prof. Dr. Eduardo Mario Mendiondo, for his support, guidance, and encouragement during the course of my Ph.D. studies. His expertise and willingness to explore new topics of research that are highly relevant for society continuously inspired me and shaped my research and academic growth.

I am also grateful for the support of Prof. Dr. Angelica Patricia Marques of Luiz de Queiroz College of Agriculture, University of São Paulo for mentoring me and answering all my questions about agriculture and crop production. In this regard, I am also grateful to all professors and researchers of the Center for Artificial Intelligence (C4AI) who were of fundamental influence in my development, especially Dr. Roberto Fray da Silva, Prof. Dr. Alexandre Delbem, and Prof. Dr. António Mauro Saraiva. I am also deeply grateful for the support of Dr. José Marengo from the National Center for Monitoring and Early Warning of Natural Disasters (CEMADEN) for accepting my invitation to participate in the first chapter of my thesis and for giving me valuable feedback.

To my friends Cibele Zeni, Denise Parizotto, Gabriela Chiquito Gesualdo, Greicelene Jesus da Silva, Pedro Gustavo Câmara da Silva, who walked with me during my academic journey, they were a source of inspiration and provided me many moments of joy and fun. To Marcos Vinicios Marcolino for all his love and support.

I would like to extend my thanks to my dear colleagues from the Water-Adaptive Design and Innovation Laboratory (WADI Lab), professors of the Department of Hydraulic Engineering and Sanitation (SHS), and the amazing staff, particularly Sá, Rose, and Fernanda.

This study was financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior do Brasil (CAPES) (finance code 001, grant no. 88888.057913/2013-00) and regular funding to the postgraduate program in Hydraulics and Sanitation of the São Carlos School of Engineering of the University of Sao Paulo by the Brazilian National Council for Scientific and Technological Development (CNPq). This study was additionally supported by the National Institute of Science and Technology for Climate Change Phase 2 (INCT-II) under CNPq grant no. 465501/2014-1, the São Paulo Research Foundation (FAPESP) under grant no. 2014/50848-9, the Center for Artificial Intelligence (C4AI-USP) under FAPESP grant no. 2019/07665-4, the IBM Corporation, CAPES under grant no. 16/2014, and the National Institute of Science and Technology (INCT) for the Fight Against Hunger under CNPq grant no. 406774/2022-6. I would like to extend my aknowledgements to the National Observatory on Water Security and Adaptive Management (ONSEADAdapta) financed by CNPq, which is an international network of researches and provided me contact with researches and novel research ideias that were fundemental for my next academic steps.

Finally, I would like to express my gratitude and appreciation to my parents (Carlos and Lenir) and to my sisters (Aline and Camila) for their love and support.

# EPIGRAPH

Nada a temer, senão o correr da luta
Nada a fazer, senão esquecer o medo
Abrir o peito à força, numa procura
Fugir às armadilhas da mata escura
Longe se vai sonhando demais
Mas onde se chega assim?
Vou descobrir o que me faz sentir
Eu, caçador de mim

*Luís Carlos Sá / Sérgio Magrão*

# ABSTRACT

BENSO, M. R. **Development of index insurance under the dynamics of multi-hazard risk in the context of climate change**. 2023. Tese (Doutorado) – São Carlos School of Engineering, University of São Paulo, 2023.

The increase in losses and damage caused by climatic extremes, such as droughts, excessive rainfall, and temperature variability associated with climate change, represents a major challenge for risk management. In this way, new methodologies aimed at insurance development, which consider the dynamics of multiple hazards, that is, how their frequencies and magnitudes vary over time, play an important role in increasing society's financial resilience in the face of climate change. Based on this challenge, this thesis aims to develop a dynamic multi-hazards indexed insurance model as an effective climate risk management tool. This work focuses on the development of insurance aimed at protecting food production, with generalizable results and methods for different sectors susceptible to losses and damages related to climate extremes. With this, the main challenges and trends, in addition to approaches traditionally used in insurance development, were explored in a systematic literature review, through which a conceptual model of indexed insurance is proposed based on the identification of hazards, the assessment of vulnerability, in financial methods, and risk pricing. Climatic impact-drivers are defined as climate means, events, and extremes that affect society and ecology. The threat these drivers pose to food production is increasing, and further research is needed to understand their relationship and how climate change influences them in Brazil to design better risk mitigation strategies. Despite the multi-hazard nature of climate threats to food production, research in index insurance has mostly focused on single index insurance. This paper presents a multi-hazard risk approach for index insurance design for food production and climate change adaptation. We used municipal-level crop yield data for soybean (1980-2022) in 452 municipalities and for maize (2003-2022) in 216 municipalities and climate data (1980-2022). We employ a multi-hazard insurance model to evaluate the effectiveness of risk transfer strategies to improve resilience and increase food production under current climate conditions and also to simulate the impacts of projected climate change (1980-2100) under moderate (SSP 5 8.5) and high emission (SSP 5 8.5) scenarios. The results showed that the inclusion of climate extreme indices improves the risk reduction potential of soybeans and maize, and these improvements were statistically significant ($p \leq 0.01$) under current conditions for soybeans and maize. The projected climate changes in both scenarios suggest that the monthly mean temperatures could rise to 5 °C at the end of the century, while there is no evidence of any trend in the monthly precipitation averages. Droughts (SPEI less than -1.5) are projected to increase 2 to 5 times the probability; however, they are located in the northwest of Paraná and in the states of São Paulo, Mato Grosso do Sul, Minas Gerais and Goias. Climate change projections also indicate an increase in high-temperature indices and an increase in heavy rainfall, especially in South Brazil. The dynamics of risk represents a potential increased pressure for the insurance market; on average, projected climate extremes represent an increase in 26.5 % risk premiums for soybeans, and temperature means an increase of 24.6 % for maize. Our study highlights the importance of considering climate extremes to reduce the risk of financial losses for farmers and provides an overview of how the risk dynamic caused by climate change is projected to affect risk premium prices and insurance design.

**Keywords:** Climate adaptation, Risk management, Extreme events, Food production

# RESUMO

BENSO, M. R. **Desenvolvimento de seguro indexado sob a dinâmica do risco multi-ameaça no contexto das mudanças climáticas**. 2023. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, 2023.

O aumento das perdas e danos causados por extremos climáticos, como secas, chuvas excessivas e variabilidade de temperatura associada às mudanças climáticas, representa um grande desafio para a gestão de riscos. Desta forma, novas metodologias destinadas ao desenvolvimento de seguros, que considerem a dinâmica dos riscos multi-ameaça, ou seja, como as suas frequências e magnitudes variam ao longo do tempo, desempenham um papel importante no aumento da resiliência financeira da sociedade face às alterações climáticas. Com base neste desafio, esta tese visa desenvolver um modelo dinâmico de seguro indexado a riscos multi-ameaça como uma ferramenta eficaz de gestão de riscos climáticos. Este trabalho tem como foco o desenvolvimento de seguros voltados à proteção da produção de alimentos, com resultados e métodos generalizáveis para diferentes setores suscetíveis a perdas e danos relacionados a extremos climáticos. Com isso, os principais desafios e tendências, além das abordagens tradicionalmente utilizadas no desenvolvimento de seguros, foram explorados em uma revisão sistemática da literatura, por meio da qual é proposto um modelo conceitual de seguro indexado baseado na identificação de perigos, na avaliação de vulnerabilidade, em métodos financeiros e precificação de risco. Os fatores de impacto climático são definidos como meios, eventos e extremos climáticos que afetam a sociedade e a ecologia. A ameaça que esses fatores representam para a produção de alimentos está aumentando, e mais pesquisas são necessárias para compreender sua relação e como as mudanças climáticas os influenciam no Brasil para projetar melhores estratégias de mitigação de riscos. Apesar da natureza multirriscos das ameaças climáticas à produção alimentar, a investigação em seguros de índices tem-se concentrado principalmente em seguros de índices únicos. Este artigo apresenta uma abordagem de riscos múltiplos para a concepção de seguros indexados para a produção de alimentos e adaptação às alterações climáticas. Utilizamos dados de produtividade agrícola em nível municipal para soja (1980-2022) em 452 municípios e para milho (2003-2022) em 216 municípios e dados climáticos (1980-2022). Empregamos um modelo de seguro contra riscos múltiplos para avaliar a eficácia das estratégias de transferência de riscos para melhorar a resiliência e aumentar a produção de alimentos nas condições climáticas atuais e também para simular os impactos das alterações climáticas projetadas (1980-2100) sob condições moderadas (SSP 5 8.5) e cenários de alta emissão (SSP 5 8,5). Os resultados mostraram que a inclusão de índices climáticos extremos melhora o potencial de redução de risco da soja e do milho, e essas melhorias foram estatisticamente significativas ($p \leq 0,01$) nas condições atuais para a soja e o milho. As alterações climáticas projectadas em ambos os cenários sugerem que as temperaturas médias mensais poderão subir para 5 °C no final do século, embora não haja evidência de qualquer tendência nas médias mensais de precipitação. Prevê-se que as secas (SPEI inferior a -1,5) aumentem 2 a 5 vezes a probabilidade; porém, estão localizados no noroeste do Paraná e nos estados de São Paulo, Mato Grosso do Sul, Minas Gerais e Goiás. As projeções de mudanças climáticas também indicam aumento nos índices de altas temperaturas e aumento de chuvas intensas, especialmente no Sul do Brasil. A dinâmica do risco representa um potencial aumento de pressão para o mercado segurador; em média, os extremos climáticos projectados representam um aumento de 26,5 % nos prémios de risco para a soja, e a temperatura significa um aumento de 24,6 % para o milho. O nosso estudo destaca a importância de considerar os extremos climáticos para reduzir o risco de perdas financeiras para os agricultores e fornece uma visão geral de como se prevê que a dinâmica do risco causada pelas alterações climáticas afecte os preços dos prémios de risco e a concepção dos seguros.

**Palavras-chave:** Adaptação climática, Gestão de risco, Eventos extremos, Produção de alimentos

## LIST OF FIGURES

1.1 Petal plot of the overview of the complex interactions of risk to food security & nutrition considering the aspects of hazard for multiple climate **extremes** (Hot and cold, and wet and dry), **exposure** to different sectoral boundaries, **vulnerability** to different aspects of food production, and **response** measures. The main focus of this research demonstrated in the red circles, which emphasize the multi-hazard climate risks on agro-ecosystems, the quantity of food produced, and the effects on food security. . . . . . 19
1.2 The layers of risk management adapted from **?** and **?** . . . . . 21
1.3 Flowchart representing the structure of the dissertation . . . . . 24

2.1 Methodological steps of PRISMA statement . . . . . 32
2.2 Weather index insurance studies. (a) Temporal distribution from papers collected from 01/01/2010 to 19/02/2022. (b) Thematic map representing the global collaboration network, where the countries in blue represent the number of studies produced by scientists. The global map indicates the country that the affiliation is located. The darker the color, the more affiliations . . . . . 35
2.3 The framework illustrates the process of selecting and prioritizing hazards, defining index thresholds, modeling losses, and optimizing insurance risk premiums. The vulnerability assessment presents two types of systems: (a) Stationary System State, where both thresholds and hazards are stationary and represent an analysis based on observed historical information, and (b) Non-stationary System State, where both indices and hazards are non-stationary and reflect a combination of observed historical and projected data. The non-stationary system anticipates a potential increase in risk and optimizes risk premiums. . . . . . 45
2.4 Graphical representation of static and dynamic state resilience systems . . . . . 46

3.1 Flowchart illustrating the systematic methodology for analyzing the impact of climate extreme drivers (CID) on crop yield. . . . . . 59
3.2 Demonstration of SHAP Explanations. Adapted from **?** . . . . . 61
3.3 Location of selected study municipalities with respect with observed cropland extent between from 2003 to 2019. The map show the percentage of cropland within the pixel 62
3.4 Spatial correlation of municipal soybean crop yields for each dataset, IBGE, Deral and GDHY . . . . . 67
3.5 Spatial correlation of municipal maize crop yields for earth dataset, IBGE, Deral and GDHY. They grey areas represent no data. . . . . . 67
3.6 Correlation analysis of three different datasets for soybean crop yields: (Global dataset of historical yields ($yield_{gdhy}$), Brazilian Institute for Geography and Statistics ($yield_{IBGE}$), and Paraná Department of Rural Economy ($yield_{PR}$) . . . . . 68
3.7 Correlation analysis of three different datasets for maize second cycle crop yields: (Global dataset of historical yields ($yield_{gdhy}$), Brazilian Institute for Geography and Statistics ($yield_{IBGE}$), and Paraná Department of Rural Economy ($yield_{PR}$) . . . . 69
3.8 Heteroskedasticity Test Results for soybean crop yields in Brazilian Municipalities . 70

3.9 Heteroskedasticity Test Results for maize crop yields in Brazilian Municipalities . . . 70
3.10 Performance evaluation of regression model for (a) Soybean and (b) Maize. Considering data from Department of Rural Economics (Deral) with data only for Paraná, Global Dataset on Historical Yields (GDHY) and Brazilian Institute of Geography and Statistics (IBGE) for the major historical agricultural producing municipalities . 71
3.11 SHAP waterfall plot visualizing the key climatic impact-drivers contributions to crop yield losses for the state of Paraná (PR) in 2019, a drought year.Monthly Maximum Value of Daily Minimum Temperature (tnx), Minimum Value of Daily Maximum Temperature (txn), Minimum Value of Daily Minimum (tnn), Maximum Value of Daily Maximum (txx), Percentage of days when TX > 90th percentile, Standardized Precipitation Evapotranspiration Index (SPEI), Total Precipitation (prctot), Daily temperature range (DTR) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 74
3.12 SHAP waterfall plot visualizing the key climatic impact-drivers contributions to crop yield losses for the state of São Paulo (SP) in 2015, a drought year. Monthly Maximum Value of Daily Minimum Temperature (tnx), Minimum Value of Daily Maximum Temperature (txn), Minimum Value of Daily Minimum (tnn), Maximum Value of Daily Maximum (txx), Percentage of days when TX > 90th percentile, Standardized Precipitation Evapotranspiration Index (SPEI), Total Precipitation (prctot), Daily temperature range (DTR) . . . . . . . . . . . . . . . . . . . . . . . . . . . 75
3.13 A comparison of the key climatic impact-drivers derived from the Brazilian Institute of Geography and Statistics (IBGE) from 2013 to 2021 and the Global Dataset on Historical Yields (GDHY) from 2009 to 2016 annual data aggregated at the municipal level was used to create a dependence plot for the soybean explanation model for validation data in RS. The data spanned from 2013 to 2021 for IBGE and from 2008 to 2016 for GDHY.. Considering 3-month SPEI in February (spei_3month_Feb) for (a) IBGE and (b) GDHY; precipitation accumulated in February (prcptot_Feb) for (c) IBGE and (d) GDHY, and precipitation accumulated in December (prcptot_Dec) for (e) IBGE and (f) GDHY . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 76
3.14 2D partial dependence derived from from the Department of Rural Economics (Deral) of the state of Parana from 2016 to 2021 of the key climatic impact-driver total precipitation in December (prctot_Dec) with the most correlated CID and their combined impact on SHAP values (ton/ha): daily temperature range in October (dtr_Oct); Considering 3-month SPEI in Jan (spei_3month_Jan); Daily Minimum Temperature in February (tnx_Feb), Minimum Value of Daily Minimum (tnn_Feb) 78

4.1 Description of the methodological steps presented in this study . . . . . . . . . . . . 90
4.2 Theoretical definition of risk reduction potential based on the impact-probability curve that is formed by an indicator of financial impact on the sector (x-axis) and the exceedance probability (y-axis). The concept is adapted from **?** . . . . . . . . . . 92
4.3 Comparison between seasonality maps for Brazil considering observed data (Era5 Land) and the three climate models considering historical simulations of the climate models K.ACE, UKESM1.0, and HadGEM3.GC3.1 considering the data from 1980 to 2010 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 95

4.4 Definition of Sowing Day of the Year estimated using deterministic rules and K.ACE, UKESM1.0, and HadGEM3.GC3.1. climate change historic simulations and projection data for historic (1980-2010), short-term (2015-2044), mid-term (2045-2074) and log-term (2075-2100) considering moderate (SSP2 4.5) and high (SSP5 8.5) emission scenarios . . . . . 96
4.5 Observed climate from ERA5 Land for the period of 1980 to 2022 and historic simulated period from 1980 to 2013 and projected period from 2015 to 2100 from climate models K.ACE, UKESM1.0, and HadGEM3.GC3.1. Fig. (a) is the observed and simulated changes in the temperature anomaly, the anomaly, and (b) is the monthly accumulated precipitation . . . . . 97
4.6 Map of the Probability Ratio (PR) $PR = \frac{P_{fut}}{P_0}$ of each municipality which is calculated by designating 3-month SPEI dips less than -1.5 as drought events, comparing historic SPEI $P_0$ to (a) moderate emission scenario (SSP2 4.5) and (b) high emission scenario (SSP5 8.5). Values above 1 indicate higher risks of drought events in forced conditions, while values below 1 indicate lower risks of drought as compared to observed values. The models used were K.ACE, UKESM1.0, and HadGEM3.GC3.1. 98
4.7 Spatial distribution of trends in climate extreme indices of at monthly time scaale for the historic period (1980 to 2010), and future period (2015 to 2100) for moderate (SSP2 4.5) and high emission (SSP5 8.5) scenarios: monthly count of days when PRCP 20mm (R20mm), monthly maximum value of daily maximum temperature (TXX), and Monthly minimum value of daily minimum temperature (TNN). The models used were K.ACE, UKESM1.0, and HadGEM3.GC3.1. . . . . . 100
4.8 Analysis of Risk Reduction Potential for Soybean (a) and Maize (b) in Response to Four Policy Scenarios: precipitation means (Prec), temperature means (Temp), coupled precipitation and temperature means (Prec-Temp), and means and extreme events (Complex) . . . . . 101
4.9 Projected additional impact of future global warming scenarios on Soybean and Maize due to increased precipitation means (Prec), temperature means (Temp), coupled precipitation and temperature means (Prec-Temp), and means and extreme events (Complex). Ensemble additional impact of climate change on Soybean as a percentage of mean of historic premiums (1980-2010), over short-term (2015-2044), mid-term (2045-2074) and long-term(2075-2100) under moderate-emissions scenario (SSP2-4.5) and high-emission scenario (SSP5-8.5) . . . . . 103

A.1 Risk analysis module applied to the city of Toledo-PR, Brazil (1,197.016 km$^2$) for demonstrating the multi-hazard model M1 (SPI,TX90p); multi-hazard model M2 (SPI,pmax,TX90p), the sum of the two models; and observed crop losses in the studied period; (a) simulation of crop losses and (b) crop loss probability in the studied period including the sum of the losses estimated with two loss models . . . . 120
A.2 Technological trends for Soybean considering different datasets . . . . . 121
A.3 Technological trends for Maize considering different datasets . . . . . 121

A.4 Variable Importance Analysis in Random Forest for Soybean Yield Prediction Across Brazilian States. The figure illustrates the variable importance scores obtained from Random Forest models applied to three distinct datasets: Deral, IBGE, and GDHY, encompassing seven Brazilian states - RS, SC, PR, SP, MS, MG, and GO . . . . . . 127
A.5 Variable Importance Analysis in Random Forest Models for Maize Yield Prediction Across Brazilian States. The figure illustrates the variable importance scores obtained from Random Forest models applied to three distinct datasets: Deral, IBGE, and GDHY, encompassing seven Brazilian states - PR, SP, MS, MG, and GO . . . . . . 128
B.6 Plant dates in Brazil for soybean. Values represent the day of the year according to Sacks et al. (2010) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 133
B.7 Harvest dates in Brazil for soybean. Values represent the day of the year according to Sacks et al. (2010) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 134
B.8 Plant dates in Brazil for maize second cycle. Values represent the day of the year according to Sacks et al. (2010) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 134
B.9 Harvest dates in Brazil for maize second cycle. Values represent the day of the year according to Sacks et al. (2010) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 135

## LIST OF TABLES

2.1 Thematic mapping of the documents based on the conceptual structure of the author's keywords divided into seven clusters with word frequency higher than 40 words according to centrality (the relevance of the theme in the development of the field) and density (the development of the field) . . . 36
2.2 Main categories for index-based insurance and specific application. Indices: Cumulative Precipitation Index (CPI); Water Storage (WS); Water Deficit (WD); Normalized Difference Vegetation Index (NDVI); Soil Moisture Index (SMI); Standardized precipitation evapotranspiration index (SPEI); Standardized Precipitation Index (SPI); El Niño Southern Oscillation (ENSO); Evaporative Stress Index (ESI); Ped Drought Index (PDI); Ribéreau-Gayon and Peynaud hydrothermal scale (RGP); R2mm; Berman and Levadoux (BBL); High Temperature (HT); Low temperature (LT); Ribéreau-Gayon and peynaud hydrothermal scale (RGP); Visible Infrared Imaging Radiometer Suite (VIIRS); Solar Radiation (SR); Wind Speed (WSpeed) . 38
2.3 Summary of Expected Loss Amount (ELA) and Expected Annual Damage (EAD) models . . . 41

3.1 Description of datasets used in the methods . . . 63
3.2 Description of the Climatic Impact-Drivers (CID) considered in this study and their respective Indices . . . 66

A.1 Description of each cluster identified . . . 119
A.2 Summary of pure risk premiums in terms of percentage of expected crop yields . . . 120
A.3 T-test for Randon Forest models for Soybean . . . 125
A.4 T-test for Randon Forest models for Maize . . . 126
B.5 Crop production of soybean in 2021 . . . 131
B.6 Crop production of maize off-season in 2021 . . . 132

# SUMMARY

Abstract . . . . . . . . . . 4
Resumo . . . . . . . . . . 6
List of Figures . . . . . . . . . . 8
List of Tables . . . . . . . . . . 12
1 General Introduction . . . . . . . . . . 17
1.1 Hypothesis . . . . . . . . . . 23
1.2 Objectives . . . . . . . . . . 23
1.2.1 General objective . . . . . . . . . . 23
1.2.2 Specific objectives . . . . . . . . . . 23
1.3 Text organization . . . . . . . . . . 23
References . . . . . . . . . . 24
2 Design and Evaluation of Weather Index Insurance for Multi-Hazard Resilience and Food Insecurity . . . . . . . . . . 29
Abstract . . . . . . . . . . 29
2.1 Introduction . . . . . . . . . . 29
2.2 Methodology . . . . . . . . . . 31
2.3 Results and Discussion . . . . . . . . . . 33
2.3.1 Bibliometric analysis . . . . . . . . . . 34
2.3.2 Systematic literature review . . . . . . . . . . 36
2.3.2.1 Hazard assessment . . . . . . . . . . 37
2.3.2.2 Vulnerability analysis . . . . . . . . . . 40
2.3.2.3 Financial methods and risk pricing . . . . . . . . . . 42
2.3.3 Conceptual framework . . . . . . . . . . 44
2.4 Conclusions . . . . . . . . . . 47
References . . . . . . . . . . 48
3 A data-driven framework for assessing climatic impact-drivers in the context of food production . . . . . . . . . . 55
Abstract . . . . . . . . . . 55
3.1 Introduction . . . . . . . . . . 55
3.2 Methodology . . . . . . . . . . 58
3.2.1 Modeling Framework . . . . . . . . . . 58
3.2.2 Study Area . . . . . . . . . . 61
3.2.3 Data collection and Processing . . . . . . . . . . 62
3.2.3.1 Crop yield data . . . . . . . . . . 63
3.2.3.2 ERA5 Land reanalysis dataset . . . . . . . . . . 65
3.2.3.3 Indices for Climatic Impact-Drivers . . . . . . . . . . 65
3.2.3.4 Soil data . . . . . . . . . . 66
3.3 Results and Discussion . . . . . . . . . . 66
3.3.1 Spatiotemporal analysis of crop yield data . . . . . . . . . . 66
3.3.2 Identifying Key Climate Impact-Drivers . . . . . . . . . . 70

3.3.2.1 Wet and dry . . . . . . . . . . . . . . . . . . . . . . . . . . . 72
3.3.2.2 Hot and cold . . . . . . . . . . . . . . . . . . . . . . . . . . 73
3.3.3 Determining Thresholds and Significance . . . . . . . . . . . . . . . . . . . 73
3.3.4 Evaluating Combined Hazards . . . . . . . . . . . . . . . . . . . . . . . 77
3.4 Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 77
References . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 79
4 A dynamic risk approach using multi-hazard index insurance design for climate extremes 87
Abstract . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 87
4.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 87
4.2 Material and methods . . . . . . . . . . . . . . . . . . . . . . . . . . . . 89
4.2.1 Index insurance model . . . . . . . . . . . . . . . . . . . . . . . . . . 89
4.2.2 Data collection and processing . . . . . . . . . . . . . . . . . . . . . . 92
4.2.3 Selection of climate models and crop calendars . . . . . . . . . . . . . . . 93
4.2.4 Climate-based multi-hazard indices . . . . . . . . . . . . . . . . . . . . 93
4.3 Results and discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . 94
4.3.1 Selection of climate change models . . . . . . . . . . . . . . . . . . . . 94
4.3.2 Projected trends of Climate Means and Extremes from 1980 to 2100 . . . . 96
4.3.3 Effectiveness of multi-hazard index insurance . . . . . . . . . . . . . . . 101
4.3.4 Increased pressure in premiums . . . . . . . . . . . . . . . . . . . . . . 102
4.4 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 104
References . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 105
5 Conclusions and recommendations . . . . . . . . . . . . . . . . . . . . . . . . 113
5.1 Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 113
5.2 Recommendations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 115
Referências . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 115
Appendix . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 117
Appendix I . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 117
Appendix II . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 121
Appendix III . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 122
Appendix IV . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 123
Appendix V . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 125
Appendix VI . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 127
References . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 128
Annex . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 131
Annex I . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 131
Annex II . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 133

# LIST OF ABBREVIATIONS

**IBGE** - Brazilian Institute of Geography and Statistics
**Deral** - Department of Rural Economics of Paraná State
**GDHY** - Global Dataset on Historical Yields
**IPCC** - Intergovernmental Panel on Climate Change
**CMIP6** - Coupled Model Intercomparison Project Phase 6
**SSP** - Shared Socioeconomic Pathways
**RCP** - Representative Concentration Pathways
**AR6** - Sixth Assessment Report
**WG1** - Working Group 1
**ETCCDI** - Expert Team on Climate Change Detection and Indices
**CDI** - Climatic impact-driver
**SPI** - Standardizes Precipitation Index
**SPEI** - Standardizes Precipitation Evapotranspiration Index
**FD** - Number of frost days
**SU** - Number of summer days
**TXx** - Monthly maximum value of daily maximum temperature
**TNx** - Monthly maximum value of daily minimum temperature
**TXn** - Monthly minimum value of daily maximum temperature
**TNn** - Monthly minimum value of daily minimum temperature
**TN10p** - Percentage of days when TN $<$ 10th percentile
**TX10p** - Percentage of days when TX $<$ 10th percentile
**TN90p** - Percentage of days when TN $>$ 90th percentile
**TX90p** - Percentage of days when TX $>$ 90th percentile
**DTR** - Daily temperature range
**Rx1day** - Monthly maximum 1-day precipitation
**Rx5day** - Monthly maximum consecutive 5-day precipitation
**R10mm** - Monthly count of days when PRCP $\geq$ 10mm
**R20mm** - Monthly count of days when PRCP $\geq$ 20mm
**Rnnmm** - Monthly count of days when PRCP $\geq$ nnmm
**PRCPTOT** - Monthly total precipitation in wet days
**SHAP** - SHapley Additive exPlanations
**RF** - Random Forest
**XGBoost** - eXtreme Gradient Boosting

# 1 GENERAL INTRODUCTION

As suggested by numerous scientists and specialists, we are currently living in an era known as the Anthropocene, (Waters et al., 2016). In a nutshell, this means that the influence that humans exert on the planet has significantly changed the Earth System (ES) that is composed of the following subsystems, Geosphere, Biosphere, Hydrosphere, Lithosphere, and Cryosphere. As reported by Rockström et al. (2009), humans have pushed the boundaries of the planet. The planet's boundaries are states of the ES that can be directly associated with the sub-parts of the ES and provide relevant services and resources to humanity. The boundaries include novel entities related to synthetic chemicals released into the environment, climate change, land system change, biodiversity integrity, freshwater change, biogeochemical flows, ocean acidification, atmospheric aerosol loading, and stratospheric ozone depletion. Transgressing beyond these boundaries represents that the Earth is operating outside its safe space, threatening Earth's stability and resilience.

Recent assessment of the planet's boundaries have shown that six of them have been transgressed (Richardson et al., 2023), this means that the development and application of resilience strategies are urgent in the context of global changes. In this context, climate change has been pointed out as one of the biggest challenges faced by our society (Stocker, 2014). The physical conditions of the climate system, that is, average climatic conditions, droughts, heat waves, and excessive rainfall, are drivers of losses and damage to society and ecosystems. Climate extremes can contribute to social and environmental risk, however, the interaction between drivers and how they compound or cascade is still poorly understood (Zscheischler et al., 2020).

Given the urgent need to find a community framework to assess and propose mitigation strategies considering intercomparable models, the Working Group I (WGI) contribution to the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR 6) has adopted a new approach called climatic impact-drivers (CID). The term refers to climate means, extremes, and events are the physical climate system conditions that can cause loss and damage to any sector of society (Ruane et al., 2022). The CID approach is focused on providing climate information in the form of numerically computable indices and identifying thresholds, that is, values of the index in which, if exceeded, interacts with the sector increasing the risk of loss and damage (Ranasinghe et al., 2021). Furthermore, the interaction between multiple sectoral boundaries must be taken into account to understand conditions that have the potential to increase or reduce risk (Simpson et al., 2021). The projected impacts and risks under the scenarios of the climate change path indicate that sectors such as health and human well-being, energy and infrastructure, water and the agroecosystem are overarching risks (Pörtner et al., 2022).

In the context of the dynamics of climate's effect on food production, learning from observations of the real world plays a role in creating efficient methods for promoting efficient mitigation and adaptation measures. Learning through research, data analysis, and scientific exploration enables us to deepen our knowledge of these impacts, from crop yield variations to disruptions in supply chains. Furthermore, this knowledge helps to advance climate science tools, such as predictive models and simulations, that aid in assessing and mitigating risks associated with climate extremes.

Despite increasing attention, studies on the impact of climate on food production go back to the beginning of modern experimentation. As an example, in the $20^{th}$ century, the botanist William Balls conducted series of cotton experiments from 1909 to 1913 in Egypt that led to a detailed statistical understanding of optimal spacing (Balls e Holton, 1914), the factors driving the sowing dates (Balls e Holton, 1915), and environmental factors that influence crop yield and quality (Balls, 1918). Balls' in-depth statistical description provided sufficient understanding of the Egyptian cotton growth curves and environmental interactions, however, was not sufficient to be generalized elsewhere, as Balls himself suggested in one of his famous lectures in 1917 (Balls, 1917).

Recently, more attention has been paid to learning about the long-term interaction between climate variability and crop yields, which is fundamental knowledge for risk management (Vogel et al., 2019) and global food security (Alpino et al., 2022). It is possible to state that, currently, further scientific advances gravitate toward generalized and reproducible models that can be used for mitigation and adaptation studies. The need for climate-smart food systems has been reported with special attention to the fact that rural poor areas will be the most affected by ongoing and projected climate changes (Fanzo et al., 2018). However, despite the growing relevance of the topic and recent scientific advances, Leal Filho et al. (2022) indicates that studies on the impact of climate change on food production are still lacking and that there is a need for food-oriented mitigating strategies. When it comes to regional disparities, climate impact studies are understudied in Central and South America and Australasia (Wakatsuki et al., 2023). This underlies the pressing need for more studies in these regions, however, there are many challenges that must be overcome.

The concept of food security implies that all people, all the time, have the physical access to food in quantity and quality required to meet their dietary needs and food preferences for a healthy and active life (Pinstrup-Andersen, 2009). In this research, we offer a perspective of climate risk to food security and nutrition considering crop yields as a proxy for food production. The diagram in Figure 1.1 illustrates our focus on the effects of climate extremes on agroecosystems and how to reduce their financial impact on agricultural production, with a particular emphasis on quantity. This is of great importance for the four pillars of food security and nutrition, which are availability, access, utilization, and stability (Godfray et al., 2010). We construct the risk concept adapted from Simpson et al. (2021) that incorporates the pillars of complex risk, namely hazard, exposure, vulnerability, and response. The hazard aspect adopted in this research considers a multihazard approach, specifically hot and cold and wet and dry CID. We acknowledge that multiple sectors are exposed to climate risk and interact with each other, often amplifying the risk to food security (Simpson et al., 2021). For example, health, water, energy, and agroecosystems are sectoral conceptual boundaries that have interactions; however, we propose a particular point of view for agroecosystems. The various aspects of these agroecosystems are vulnerable to climate extremes, as derived from Pörtner et al. (2022), climate extremes can affect crop yields in terms of crop yield quantity and quality, and food prices. In this framework, responses, such as insurance, adaptation plans, post-disaster payments, and early warning systems, are included in the risk analysis as a driver of risk.

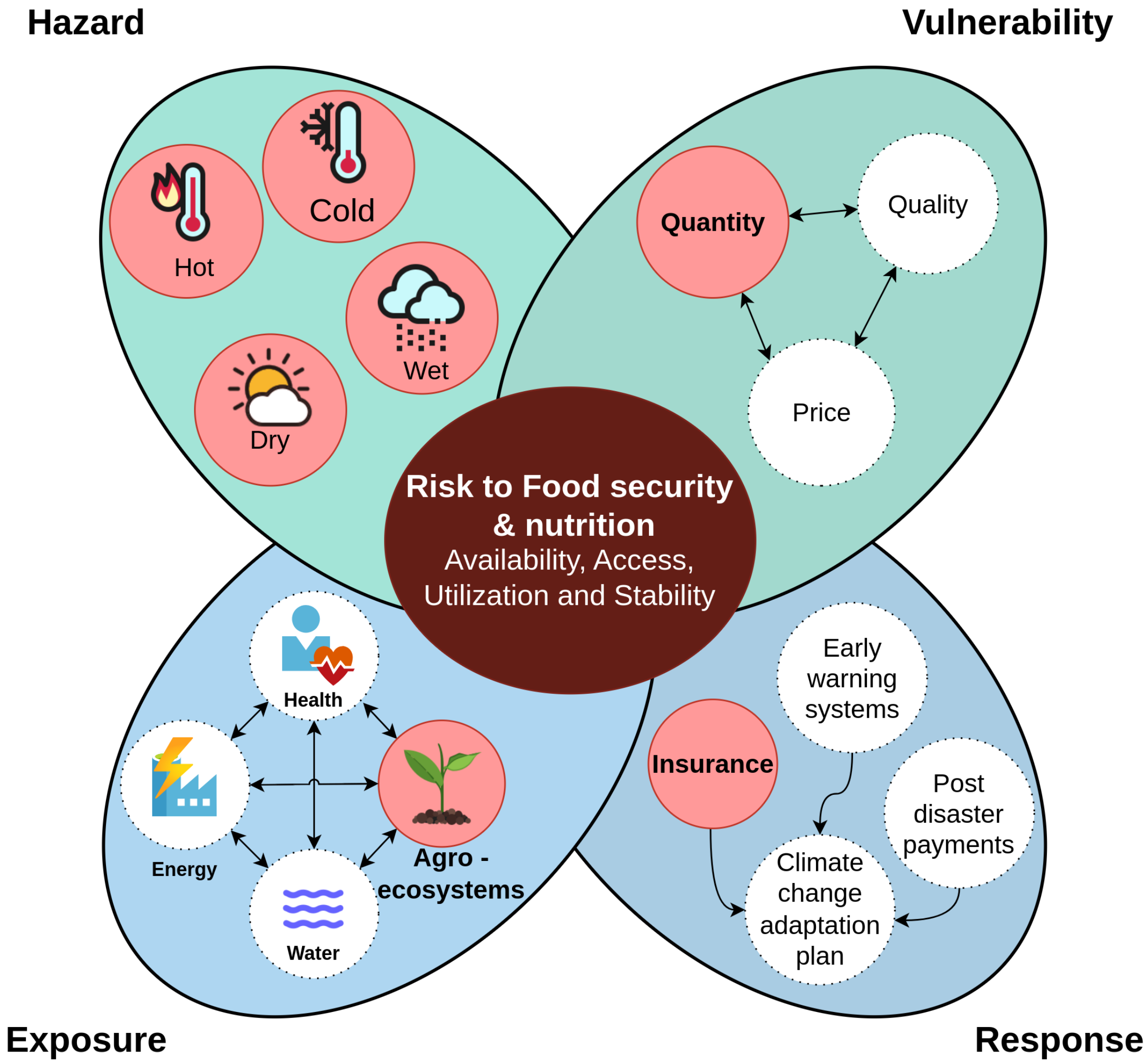


**Figure 1.1.** Petal plot of the overview of the complex interactions of risk to food security & nutrition considering the aspects of hazard for multiple climate **extremes** (Hot and cold, and wet and dry), **exposure** to different sectoral boundaries, **vulnerability** to different aspects of food production, and **response** measures. The main focus of this research demonstrated in the red circles, which emphasize the multi-hazard climate risks on agro-ecosystems, the quantity of food produced, and the effects on food security.

Climate variability is the main regulator of food production in the Earth system. Mean climate variables such as precipitation and temperature are key to determining crop suitability. Today, we need to produce more food for an increasing population (Godfray et al., 2010) with increased risk of climate change. Food production faces the limits of planetary boundaries and an increasing risk of climate change. An increase in the water consumption of crop production that can induce negative feedback to society, creating the paradigm of more food equals less water (Rockström et al., 2012). Extreme weather events are responsible for both the affect of yields and harvested areas around the world (Lesk et al., 2016), and their increasing frequency and magnitude have shown growing concern for future food security (Battisti e Naylor, 2009).

To address these issues, we need to have data and methods for designing appropriate risk management strategies. Risk management can be defined in different layers according to the probability and severity of an event (Mechler et al., 2014). Choudhary et al. (2016) proposes three layers of risk management (Figure 1.2). Risk mitigation (Layer 1) is usually the favored adaptation method for frequent, low-impact hazards (benefit-cost assessments have revealed great potential to reduce risks at this lower level). To transfer residual risk for medium-layer hazards (Layer 2), a mix of insurance, risk reduction, and other risk-financing tools might be used. If crucial stress thresholds are reached in the event of a rare or catastrophic incident (Layer 3), public and international aid will be required. Finally, managing the residual risk associated with the extremely high-level risk layer may pose difficulties for international aid organizations (Limit to adaptation). Layers 1 and 2 are considered as *ex-ante* risk management measures. Risk mitigation measures are measures taken to reduce the severity of losses and damages caused by an extreme event. Risk transfer is for impacts that cannot be prevented. Insurance is the benchmark example for risk transfer; threfore, the two terms are frequently used interchangeably. The core principle of insurance is to transfer the risk that cannot be avoided to a third party for a fee or premium. The third party, usually a (re)insurance company, pays an indemnization in case of a realization of an extreme event. Risk coping is a *ex post* measure, that is, it is performed after the event takes place. Usually, it is made up of emergency measures taken by governments to recover. However, there is limited capacity in which organizations or countries can absorb risk; this is also called the limits of adaptation (Mechler et al., 2014).

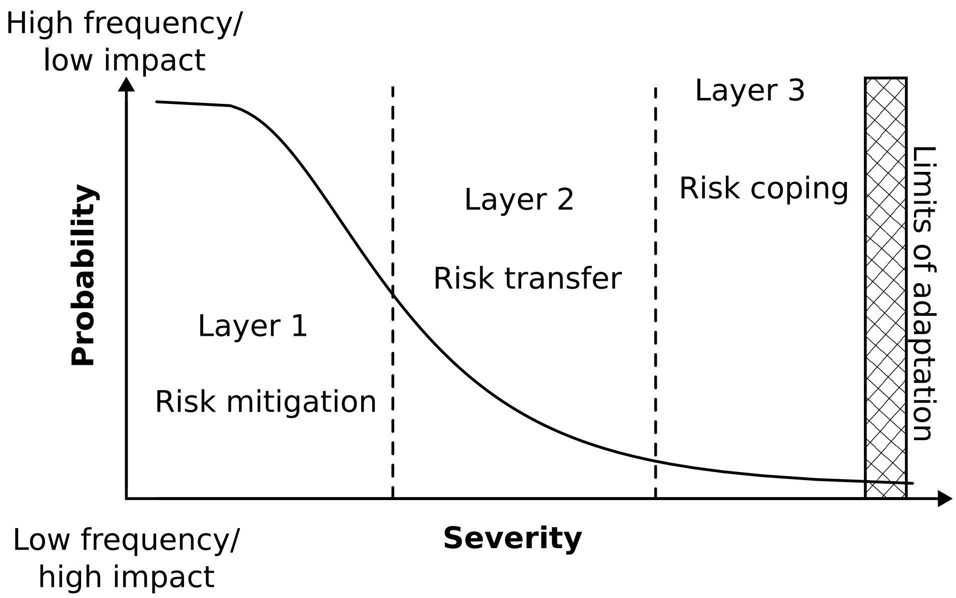


**Figure 1.2.** The layers of risk management adapted from Mechler et al. (2014) and Choudhary et al. (2016)

In the context of Brazil, there are two major programs that offer risk mitigation for agriculture (Souza et al., 2022). The first is the Agricultural Activity Guarantee Program (Programa de Garantia da Atividade Agropecuária – PROAGRO), which is designed for small and medium farmers and aims at safeguarding income for farmers in case of extreme events. The program can dismiss financial obligations of farmers in the case of rural credit for production costs acquired in a bank or reimburse them in case they used their own funds. The second is the Rural Insurance Premium Subsidy Program (PSR), a government initiative that provides monetary resources to (re)insurance companies, allowing them to reduce risk premiums for farmers. As reported by Souza et al. (2022), from 2013 to 2021, the PSR program experienced a substantial growth of 108%, achieving an insured capital of R$ 67 billion by 2021. However, the PROAGRO program has lost its place with a decrease of 39% in the same period and was able to cover a capital of R$ 17 billion in 2021.

So far, we have used the term (re)insurance to highlight the role of both insurance and reinsurance in agricultural risk management. Reinsurance includes the transfer of risk from an insurance company to another entity, the reinsurer. Reinsurance also denotes a collaboration with insurance providers, particularly in Brazil where a financial quota share is a crucial component of insurance contracts. One of the main functions of reinsurance is to increase the capacity of insurance companies to retain risk. To exemplify, when an insurance company underwrites a contract to protect a farmer against droughts, the company makes a commitment to make a payment for this farmer in case of losses. As the number of policies increases, the company faces the risk of not having sufficient funds to pay all the indemnization and keep operating. Therefore, the insurance company can enter into a contract with a reinsurance company and be protected against financial insolvencies. As presented by Contador (2014), in the demand direction, the demand for insurance also represents the demand for reinsurance; on the other hand, in the supply direction, reinsurance represents additional capital for insurance companies, which increases the capacity of insurance companies underwrite more insurance policies and increase insurance coverage.

Index insurance is a type of risk transfer tool that has gained attention in recent years as an innovative tool to improve climate resilience in the face of climate change (Collier et al., 2009), especially for poor and developing countries (Skees, 2008a). Index insurance is a type of contract based on an index that represents the physical conditions that drive losses and damages to a sector. Traditionally, claim payments are made every time the index is exceeded. An in-depth description is given in Chapter 2 of this research. In the context of the Water Adaptive Design and Innovation Laboratory (WADI-Lab), the topic index-insurance has been explored through the model Hydrologic Risk Transfer Model (MRTH) of the Department of Hydraulic Engineering and Sanitation of the University of São Paulo (SHS) (MTRH-SHS). The MTRH-SHS is an insurance fund simulator developed to insure financial losses related to hydrological extremes (droughts and floods) based on multiannual policies (Guzmán et al., 2020; Mohor e Mendiondo, 2017; Righetto et al., 2007).

The model is composed of three modules: hazard module; vulnerability module; and financial module. In the first step, the hazard module we perform stochastic modeling of the observed data to obtain a time series of low flows with a return period of up to 100 years. The vulnerability module focuses on building functions that relate economic loss with the damage caused by the low flows according to the water balance equation following a specific allocation schema. In the later financial module, a cashflow equation is run, and using optimization tools, the fund simulation will seek optimized premiums, here also mentioned as fair premium values.

Different types of hydrological index have been applied according to specific hydrological hazards, Righetto et al. (2007) used the stream level to assess the impacts of floods in a commercial setting. Mohor e Mendiondo (2017) have focused on the effects of droughts in larger areas using the 7-day, 10-year low-flow index (Q7,10). Guzmán et al. (2020) used water deficit from the Threshold level Method (TLM) for estimating the impacts of severity, duration, and frequency of droughts on a water supply utility. For the estimation of hazards, different hydrological modeling frameworks have been proposed. Righetto et al. (2007) used observed time series in a stochastic modeling framework. Mohor e Mendiondo (2017) used more complex semi-distributed conceptual hydrologic and hydrodynamic models, for example, the large basin MGB-IPH model, to simulate drought scenarios. Guzmán et al. (2020), on the other hand, used the lumped conceptual hydrological model, WEAP. Both Mohor e Mendiondo (2017) and Guzmán et al. (2020) incorporated climate projections from regional climate models into their modeling framework to assess the effects of climate change on optimal premium values. Here, we further develop the model that applies to multihazard climate risks and food production.

This doctoral dissertation aims to advance multihazard index insurance as a viable adaptation measure to provide long-term financial stability to farmers, foster climate resilience in food production systems, and ultimately contribute to food security in Brazilian regions under climate change. By proposing a conceptual framework for multihazard risk index insurance, this research seeks to address the challenges of changing climate risk, also known as dynamic risk. Furthermore, this doctoral dissertation is aligned with the United Nations Sustainable Development Goals (SDGs), with special regard to SDG 2 Zero Hunger and SDG 13 Climate Action. This research also contributes to research related to the 23 Unsolved Problems in Hydrology, when analyzing the variability of extremes and their interfaces with society in the nexus water-environment-food-health (Blöschl et al., 2019). This work is aligned with the UNESCO Intergovernmental Hydrological Programme (IHP) (en.unesco.org) and the scientific decade of the International Association of Hidrological Sciences (IAHS) Hydrology Engaging Local People IN one Global world (HELPING), more specifically Theme 1: HELPING with global and local interactions with contributions to the Working Group on Water for Biodiversity in a Changing World (iahs.info).

## 1.1 Hypothesis

This research is based on the hypothesis that the implementation of a multi-hazard risk index insurance strategy in Brazilian agro-producing regions can enhance the financial resilience of food production systems. For example, by promoting adaptation to the adverse effects of various climate extremes, including droughts, excessive rainfall, strong winds, cold spells, and extreme heat, which fall under the category of Climatic Impact-Drivers (CID), index insurance is considered an adaptive tool for risk management. This adaptive approach, when analyzed across different climate change scenarios scenarios, is expected to address current agricultural challenges while accommodating the non-stationarity of climate and environmental changes.

## 1.2 Objectives

### 1.2.1 General objective

Propose a multi-hazard risk transfer model as a tool to adapt the impacts of climate extremes on food production under climate changes in Brazil.

### 1.2.2 Specific objectives

- Review and analyze index insurance design for multihazard risk in the context of food production and climate extremes
- Propose a conceptual framework for multihazard risk transfer model for climate extremes
- Build a learning model to assess the impact of climate extremes on food production
- Assess the impacts of climate change on food production and index insurance as a mitigation strategy

## 1.3 Text organization

This dissertation explores the intersection between risk management, climate extremes, food production, and climate change. It adopts a multidisciplinary approach, incorporating systematic review, new modeling techniques for impact assessment, and the proposal of an index insurance model. Therefore, the dissertation is divided into five chapters (Fig. 1.3). A general introduction is presented in the **first chapter**. The chapter contains the motivation and justification for the work presented in the dissertation, the main hypothesis, and objectives. The **second chapter** is a systematic review of the existing literature on index insurance for food security in changing climate conditions, this research synthesizes current knowledge index insurance design focusing on multi-hazards risk and food security and addresses challenges and research opportunities. In this chapter, a conceptual framework for index insurance design is presented. Using learning methods, the **third chapter** presents the study that assesses the impact of Climatic Impact-Drivers on food production. Historical climate and agricultural data is used to train models that predict the effects of extreme weather events on crops. These models provide insight into areas of vulnerability and make it easier to predict potential production losses. In the **fourth chapter**, a case of study of multi-hazard index insurance for food production is proposed. This chapter presents the simulation of multi-year index insurance for maize and soybean considering the main agro-producing regions in Brazil. Different multi-hazard combinations are evaluated considering wet and dry and hot and cold CIDs under climate change scenarios. Finally, in the **fifth chapter** provides a general conclusion to this dissertation with a summary of the key findings and presenting recommendations for future work. This chapter explains how this research contributes to understanding risk management, climate extremes, food production, and climate change.

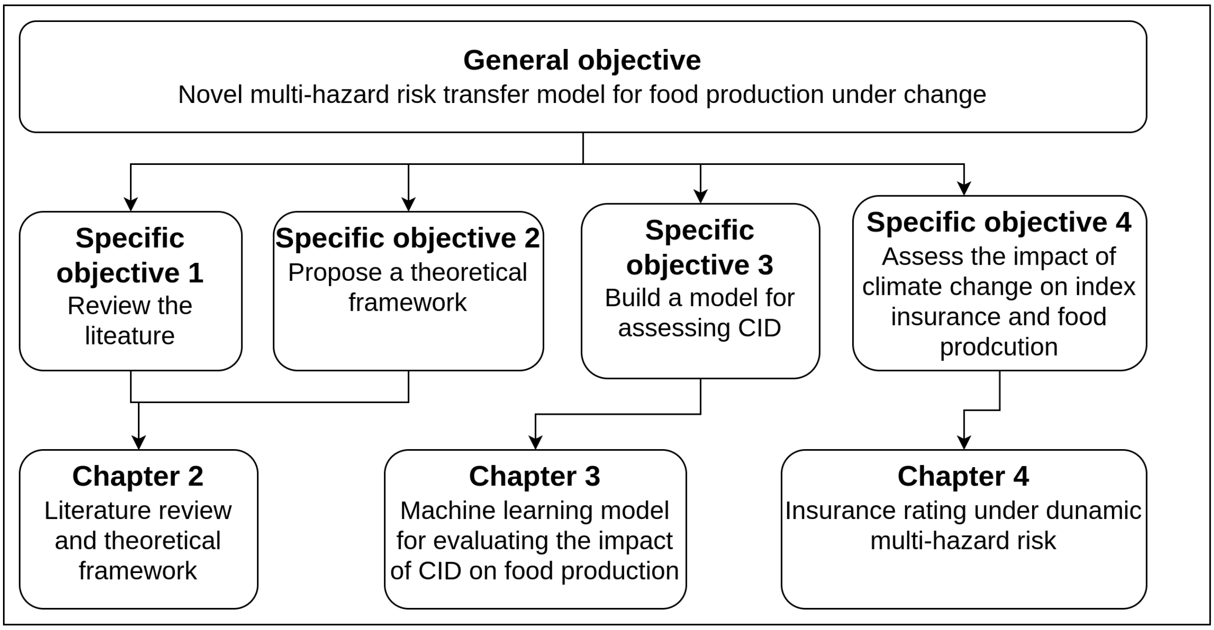


**Figure 1.3.** Flowchart representing the structure of the dissertation

## References

Alpino, T. d. M. A., Mazoto, M. L., Barros, D. C. d., e Freitas, C. M. d. (2022). The impacts of climate change on food and nutritional security: a literature review. *Ciência & Saúde Coletiva*, 27:273–286.

Balls, W. e Holton, F. (1914). Analyses of agricultural yield. part i, the spacing experiment with egyptian cotton. *Phil. Trans. Roy. Soc. B*, pages 103–180.

Balls, W. L. (1917). Crop reports and forecasts. means whereby precise reports and reliable fore-casts of the cotton crop may be prepared. *Journal of the Textile Institute*, 8(3):222–224.

Balls, W. L. (1918). V. analyses of agricultural yield. part iii.—the influence of natural environmental factors upon the yield of egyptian cotton. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 208(348-359):157–223.

Balls, W. L. e Holton, F. S. (1915). Ix. analyses of agricultural yield. part ii.—the showing-date experiment with egyptian cotton, 1913. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 206(325-334):403–480.

Battisti, D. S. e Naylor, R. L. (2009). Historical warnings of future food insecurity with unprecedented seasonal heat. *Science*, 323(5911):240–244.

Blöschl, G., Bierkens, M. F., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J. W., McDonnell, J. J., Savenije, H. H., Sivapalan, M., et al. (2019). Twenty-three unsolved problems in hydrology (uph)–a community perspective. *Hydrological sciences journal*, 64(10):1141–1158.

Choudhary, V., D'Alessandro, S., Giertz, A., Suit, K., Johnson, T., Baedeker, T., e Caballero, R. (2016). Agricultural sector risk assessment: methodological guidance for practitioners. *World Bank Group. http://documents. worldbank. org/curated/en/586561467994685817/Agricultural-sector-risk-assessmentmethodological-guidance-forpractitioners.*

Collier, B., Skees, J., e Barnett, B. (2009). Weather index insurance and climate change: Opportunities and challenges in lower income countries. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 34:401–424.

Contador, C. R. (2014). Reinsurance in brazil: challenges and opportunities of the opening of the market. *CEBRI-Brazilian Center for International Relations, Rio de Janeiro*, page 148.

Fanzo, J., Davis, C., McLaren, R., e Choufani, J. (2018). The effect of climate change across food systems: Implications for nutrition outcomes. *Global food security*, 18:12–19.

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., e Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. *science*, 327(5967):812–818.

Guzmán, D. A., Mohor, G. S., e Mendiondo, E. M. (2020). Multi-Year Index-Based Insurance for Adapting Water Utility Companies to Hydrological Drought: Case Study of a Water Supply System of the Sao Paulo Metropolitan Region, Brazil. *Water*, 12(11):2954.

Leal Filho, W., Setti, A. F. F., Azeiteiro, U. M., Lokupitiya, E., Donkor, F. K., Etim, N. N., Matandirotya, N., Olooto, F. M., Sharifi, A., Nagy, G. J., et al. (2022). An overview of the interactions between food production and climate change. *Science of the Total Environment*, 838:156438.

Lesk, C., Rowhani, P., e Ramankutty, N. (2016). Influence of extreme weather disasters on global crop production. *Nature*, 529(7584):84–87.

Mechler, R., Bouwer, L. M., Linnerooth-Bayer, J., Hochrainer-Stigler, S., Aerts, J. C., Surminski, S., e Williges, K. (2014). Managing unnatural disaster risk from climate extremes. *Nature Climate Change*, 4(4):235–237.

Mohor, G. S. e Mendiondo, E. M. (2017). Economic indicators of hydrologic drought insurance under water demand and climate change scenarios in a Brazilian context. *Ecological Economics*, 140:66–78.

Pinstrup-Andersen, P. (2009). Food security: definition and measurement. *Food security*, 1(1):5–7.

Pörtner, H.-O., Roberts, D. C., Poloczanska, E. S., Mintenbeck, K., Tignor, M., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V., et al. (2022). *IPCC, 2022: Summary for policymakers*. Cambridge University Pres.

Ranasinghe, R., Ruane, A. C., Vautard, R., Arnell, N., Coppola, E., Cruz, F. A., Dessai, S., Saiful Islam, A., Rahimi, M., Carrascal, D. R., et al. (2021). Chapter 12: Climate Change Information for Regional Impact and for Risk Assessment — ipcc.ch. `https://www.ipcc.ch/report/ar6/wg1/chapter/chapter-12/`. [Accessed 08-10-2023].

Richardson, K., Steffen, W., Lucht, W., Bendtsen, J., Cornell, S. E., Donges, J. F., Drüke, M., Fetzer, I., Bala, G., von Bloh, W., et al. (2023). Earth beyond six of nine planetary boundaries. *Science Advances*, 9(37):eadh2458.

Righetto, J., Mendiondo, E., e Righetto, A. (2007). Modelo de Seguro para Riscos Hidrológicos. *Revista Brasileira de Recursos Hídricos*, 12(2):107–113.

Rockström, J., Falkenmark, M., Lannerstad, M., e Karlberg, L. (2012). The planetary water drama: Dual task of feeding humanity and curbing climate change. *Geophysical Research Letters*, 39(15).

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F. S., Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., et al. (2009). A safe operating space for humanity. *nature*, 461(7263):472–475.

Ruane, A. C., Vautard, R., Ranasinghe, R., Sillmann, J., Coppola, E., Arnell, N., Cruz, F. A., Dessai, S., Iles, C. E., Islam, A. S., et al. (2022). The climatic impact-driver framework for assessment of risk-relevant climate information. *Earth's Future*, 10(11):e2022EF002803.

Simpson, N. P., Mach, K. J., Constable, A., Hess, J., Hogarth, R., Howden, M., Lawrence, J., Lempert, R. J., Muccione, V., Mackey, B., et al. (2021). A framework for complex climate change risk assessment. *One Earth*, 4(4):489–501.

Skees, J. R. (2008a). Challenges for use of index-based weather insurance in lower income countries. *Agricultural Finance Review*, 68(1):197–217.

Souza, P. Z., Pereira, L., e Stussi, M. (2022). Revealing Incentives: Implications of the Design of Public Rural Insurance Policies in Brazil - CPI — climatepolicyinitiative.org. [Accessed 23-10-2023].

Stocker, T. (2014). *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge university press.

Vogel, E., Donat, M. G., Alexander, L. V., Meinshausen, M., Ray, D. K., Karoly, D., Meinshausen, N., e Frieler, K. (2019). The effects of climate extremes on global agricultural yields. *Environmental Research Letters*, 14(5):054010.

Wakatsuki, H., Ju, H., Nelson, G. C., Farrell, A. D., Deryng, D., Meza, F., e Hasegawa, T. (2023). Research trends and gaps in climate change impacts and adaptation potentials in major crops. *Current Opinion in Environmental Sustainability*, 60:101249.

Waters, C. N., Zalasiewicz, J., Summerhayes, C., Barnosky, A. D., Poirier, C., Gałuszka, A., Cearreta, A., Edgeworth, M., Ellis, E. C., Ellis, M., et al. (2016). The anthropocene is functionally and stratigraphically distinct from the holocene. *Science*, 351(6269):aad2622.

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., van den Hurk, B., AghaKouchak, A., Jézéquel, A., Mahecha, M. D., et al. (2020). A typology of compound weather and climate events. *Nature reviews earth & environment*, 1(7):333–347.

# 2 DESIGN AND EVALUATION OF WEATHER INDEX INSURANCE FOR MULTI-HAZARD RESILIENCE AND FOOD INSECURITY

A modified version of this chapter has been published as: BENSO, Marcos Roberto; GESUALDO, Gabriela Chiquito; SILVA, Roberto Fray; SILVA, Greicelene Jesus; RÁPALO, Luis Miguel Castillo; NAVARRO, Fabricio Alonso Richmond; MARQUES, Patricia Angélica Alves; MARENGO, José Antônio; MENDIONDO, Eduardo Mario. Review article: design and evaluation of weather index insurance for multi-hazard resilience and food insecurity. **Natural Hazards And Earth System Sciences**, [S.L.], v. 23, n. 4, p. 1335-1354, 6 abr. 2023. Copernicus GmbH. http://dx.doi.org/10.5194/nhess-23-1335-2023.

#### Abstract

Ensuring food security against climate risks is attracting growing interest because of the need to meet current food demand for a growing world population. Index insurance has been referred to as a tool to increase the financial resilience of food production in the face of climate-related threats. Despite the growing attention, the area of developing multi-hazard insurance has not been explored in depth. This chapter aims to review the insurance design of the weather index for food security resilience, including the methodology for calculating the natural hazards indexes, the vulnerability assessment, and the risk pricing. We searched for relevant research papers in the Scopus database using the preferred reporting items for systematic reviews and meta-analysis (PRISMA) protocol. Initially, 364 peer-reviewed articles were selected from 01/01/2010 to 19/02/2022 for bibliometric analysis. The bibliometric analysis confirmed that the field is growing in terms of number of publications and citations. We highlight that climate change is a relevant theme associated with index insurance and is gaining more attention in the literature. The 26 most relevant articles of the last five years were then systematically analyzed. In terms of hazard assessment, most of the research is focused on droughts. Other hazards, such as extreme temperature variation, excessive rainfall, and wildfires, were poorly studied. Most studies considered only single-hazard risk, and the multi-hazard risk studies assumed independence between hazards, neglecting the synergy hypothesis between hazards. Despite considerable research efforts into index insurance, the majority of studies have concentrated on crop yields, i.e. the quantity of food produced. There is a lack of research looking at other aspects of food security, such as transportation, storage, and distribution. From the literature review, we proposed a conceptual framework that illustrates design paths for a generalized index insurance design and evaluation incorporating multi-hazard risk. Overall, these results broaden our understanding of the opportunities and challenges of index insurance design and provide guidance on how to design index insurance for food production and climate resilience.

## 2.1 Introduction

The increased frequency and magnitude of extreme weather and climate events have been evidenced in many regions of the world, widely attributed to climate change IPCC (2022).

In recent years, extreme weather events have caused significant losses and damage in many climate-sensitive sectors affecting urban and rural areas. Insurance is essential to provide economic sustainability to vulnerable sectors and improve recovery from catastrophic climate events.

Insurance has been pointed out as a tool for safeguarding populations and properties from climate change (UNEP, 2012). Nevertheless, Kraehnert et al. (2021) argue that insurance itself is not an adaptation measure and depends on several characteristics and factors. Some relevant factors are living standards, economic well-being, the availability of safety nets for poor people, the characteristics of the sector, and the type of risk to which sectors are exposed (FAO, 2014).

Swiss Re (2021a) predicted non-life insurance premiums to rise 10% above the prepandemic state and acknowledges that climate change might have an even more significant impact on the insurance industry. They propose that increasing underwriting policies against climate-related disasters is vital to tackle this problem.

However, the challenge might be more significant in developing countries with fewer insurance coverage. On the one hand, the premiums per capita in the US and Canada were 7,270 USD in 2020 much higher than the world average of 809 USD per capita and the Eurozone average of 2,723 USD. On the contrary, Latin America, the Caribbean, emerging Europe and Asia presented premiums of 203, 159, and 215 USD per capita, respectively. Africa and the emerging Middle East had a much lower number, of 45 and 93 USD per capita, respectively (Swiss Re, 2021b).

Index-based insurance policy is a solution to improve insurance coverage, especially in low-income areas (Raucci et al., 2019b). The term index insurance started to be used for crop yield insurance policies based on area-yield indices as firstly described by Halcrow (1949) and then further revisited by Miranda (1991). The area-yield insurance model was adopted in the US in the early 1990s, dividing agricultural areas in the crop domain into Group Risk Plans (GRP). Indemnities were triggered when forecasted crop yields would fall below a certain threshold within each GRP (Skees, 2008a).

Area-yield contracts depend on the availability of data and the technical capacity to evaluate and monitor group risk units, which can be costly and impractical in many poor and developing countries. To overcome this challenge, the researchers proposed contracts based on weather indices (Müller e Grandi, 2000).

In the financial and actuarial literature, weather derivatives have been used to associate the financial frustration of a business with a weather index (Müller e Grandi, 2000). Contracts based on weather indices have helped policyholders hedge against adverse conditions in the clothing industry (Štulec et al., 2019b), hydroelectric plants (Foster et al., 2015), and solar energy systems (Boyle et al., 2021). Crop yield contracts based on rainfall have been used due to their simplicity and availability of data (Yoshida et al., 2019). The method uses rain from nearby weather stations to predict losses and the threshold is usually defined based on an index in the growing season.

This type of contract almost eliminates the need for on-site verification of losses, reducing administrative costs, and improving the transparency of insurance products (Shirsath et al., 2019). Insurance companies also benefit from reducing moral hazards since crop losses are estimated from indices provided by third-party agencies (Ghosh et al., 2021). Furthermore, due to reduced costs, weather-indices-based contracts have been used for microinsurance contracts in poor rural areas to improve protection against adverse weather conditions and prevent smallholder farmers from falling into poverty traps (Skees, 2008a). Despite its advantages, index insurance has a particular side effect called basis risk, which is a mismatch between actual losses and predicted losses (Ghosh et al., 2021).

As expected from the relevance of agriculture in the insurance industry, most reviews in the literature focus on understanding index insurance and microinsurance for agriculture (Leblois et al., 2014; Sarris, 2013). Zara (2010) proposed a systematic review of the role of weather derivatives in the wine industry. Akter (2012) focused on reviewing microinsurance problems in Bangladesh, looking for evidence for insurance demand, how to approach the market, and designing challenges to improve the safety of the vulnerable population, especially smallholder farmers.

Several studies have been reported on the single-hazard risk insurance design. Considering only one hazard does not include the expression of risk due to interactions between different hazards (Gill e Malamud, 2014; Hillier et al., 2020). Insurance risk assessment and climate change impacts have recently been reviewed by Lyubchich et al. (2019). The authors review several adverse events such as floods, hail, and excessive wind, but the interaction effect between hazards could be explored further.

Sekhri et al. (2020) proposed a framework for multi-hazard risk management. However, it was specific to mountainous regions and a broader risk management strategy. Komendantova et al. (2014) introduced a participatory risk governance framework that allows feedback from stakeholders. Abdi et al. (2022) conducted an extensive review of possible index insurance applications for agriculture. The authors summarized the indices and methods for designing index insurance with possible applications for multi-hazard risks. However, multi-hazard implementation has not been nearly as thoroughly investigated as single-hazard problems.

Despite the importance of index insurance as a tool for mitigating impacts of climate extremes in several sectors, but with special regard to food security, recent developments must be synthesized. Taking into account this challenge, this study thoroughly analyzes the literature, further describes identified gaps, and proposes a framework to address the design of insurance based on multi-hazard indexes for agricultural purposes. The systematic review was designed to answer the following questions, considering the context of index insurance: 1) What indices are used to assess and monitor extreme weather events?; 2) What functions and methods are used to assess the vulnerability of food production to extreme weather events?; and 3) How to determine risk premiums?

## 2.2 Methodology

A systematic review was conducted to better identify the state-of-the-art in designing and implementing multi-hazard index-based insurance in agricultural environments and to identify the main gaps in the current techniques and models. The Preferred Reporting Items for Systematic Reviews and Meta-Analyzes (PRISMA) protocol (Liberati et al., 2009) was applied and the Scopus database was used for data collection.

This database was chosen because of its comprehensive coverage of relevant events and scientific journals related to climate change, agriculture, insurance design, and multi-hazard frameworks and techniques, among other relevant topics. It encompasses a wide range of topics in technology, science, social sciences, medicine, humanities, and the arts (Scopus, 2022).

We reviewed the literature following the PRISMA protocol (Liberati et al., 2009). First, a bibliometric analysis was performed on the selected articles from 01/01/2010 to 19/02/2022, using the Bibliometrix R package (Aria e Cuccurullo, 2017). Then a critical analysis of the most cited articles in the last five years (2018-2022) was performed to identify fundamental research topics, themes, keywords, and guidelines for index insurance design and evaluation and to identify the main literature gaps.

The systematic review process was divided into four steps (Figure 2.1). The first consisted of defining the search strings based on the three research questions described in Section 1. Our search string was made up of keywords in the English language extracted from an in-depth analysis of relevant literature reviews and papers on the topic. It was then used to search for terms in the title, abstract, and keywords of the documents in the Scopus database. The following criteria were considered:

- English keywords: multi-risk weather index insurance.
- English synonyms: multi-risk, risk, weather, climate, index, parametric, insurance, microinsurance, derivative.
- Search string: TITLE-ABS-KEY ( (risk (multi AND risk) OR portfolio) OR ( index OR parametric ) AND ( insurance OR microinsurance OR derivative ) AND ( weather OR climate ) ).

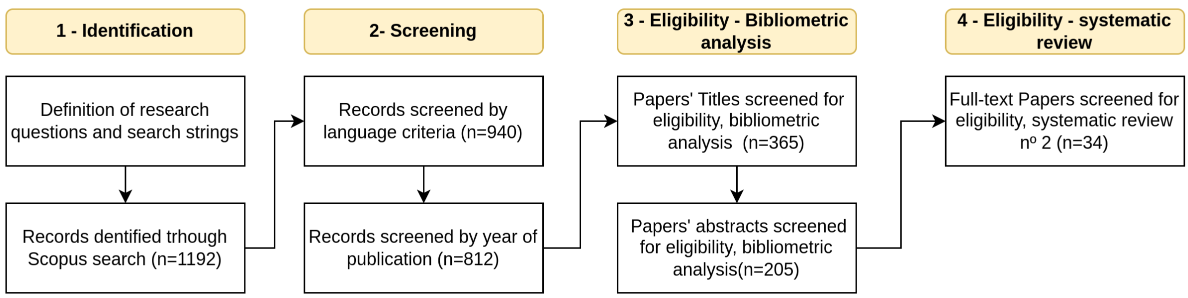


**Figure 2.1.** Methodological steps of PRISMA statement

The second step was the screening process. First, we selected only scientific papers published in peer-reviewed journals in English, Portuguese, or Spanish. Review articles, books, book chapters, and conference proceedings were excluded from the analysis, following the methodology used in other systematic reviews in the literature. This step resulted in 1192 documents.

In the third step, an analysis of the documents' titles and abstracts was conducted to filter only works that designed or implemented a complete application of an index insurance or weather derivative. Many studies on the evaluation of index insurance demand and traditional insurance models were excluded. This step resulted in 365 documents. Then, several tools used for bibliometric analysis were applied to this dataset.

The fourth step was to perform a critical review of the 26 most cited papers published in the last five years of the dataset (2018 to 2020). This evaluation excluded papers that did not provide information on index insurance design. This review was divided into (i) hazard identification; (ii) vulnerability analysis; and (iii) financial method and risk pricing analyses. These three modules were adapted from the frameworks developed by Guzmán et al. (2020), Mohor e Mendiondo (2017), and Righetto et al. (2007), which encompass the main aspects of weather-based insurance design.

Before analyzing the full papers, it is critical to specify the main concepts and definitions. Although there are many definitions for concepts such as hazard, multi-hazard, resilience, and food security, we chose to adopt the most broadly accepted ones. These were:

- Hazard: "A dangerous phenomenon, substance, human activity or condition that may cause loss of life, injury or other health impacts, property damage, loss of livelihoods and services, social and economic disruption, or environmental damage" (UNDRR, 2014). This paper explicitly refers to hazards derived from extreme weather and climate events.

- Multi-hazard: "[...] all possible and relevant hazards and the valid comparison of their contributions to hazard potential, including the contribution to hazard potential from hazard interactions and spatial/temporal coincidence of hazards, while also taking into account the dynamic nature of vulnerability to multiple stresses" (Gill e Malamud, 2014).

- Vulnerability: "The conditions determined by physical, social, economic and environmental factors or processes, which increase the susceptibility of a community to the impact of hazards" as defined by the Hyogo Framework for Action (UNDRR, 2014). For this paper, the concept of vulnerability was focused on physical damages and losses derived from the realization of an extreme weather event. Therefore, we use a classical approach to quantify the vulnerability of risk-averse individuals, which considers that the greater the losses, the greater the vulnerability. Although this traditional definition has been questioned as a reducer of only the economic sphere of an issue that permeates social, political and environmental dimensions, this is ultimately a practical approach of widespread use (Machado et al., 2005).

- Resilience: "The ability of a system, community or society exposed to hazards to resist, absorb, accommodate to and recover from the effects of a hazard in a timely and efficient manner, including through the preservation and restoration of its essential basic structures and functions" UNDRR (2009). In the context of this paper, and as described primarily by Mohor e Mendiondo (2017) and Guzmán et al. (2020), in the resilience module of an index insurance schema, the risk premium is an indicator of the resilience of a sector to cope with weather and climate extreme events.

- Food security: "exists when all people, at all times, have physical, social and economic access to sufficient, safe and nutritious food which meets their dietary needs and food preferences for an active and healthy life. Household food security is the application of this concept to the family level, with individuals within households as the focus of concern." (FAO, 2003). In summary, food security is rooted in the pillars of availability, access, and utilization (Barrett, 2010). This broadens the concept of food security to encompass different linkages in the supply chain, such as food production, transportation, storage and distribution.

## 2.3 Results and Discussion

This section describes the main results of this work. It also discusses important aspects related to applying these results in different scenarios and contexts. It is divided into the following subsections: 3.1 contains the main results of the bibliometric analysis; 3.2 presents an in-depth literature review of the most relevant papers identified, exploring the hazard assessment, vulnerability analysis, and financial methods and risk pricing modules; and 3.3 presents the proposed conceptual framework, encompassing both its description and an example to illustrate its main aspects and contributions to the field of multi-hazard weather-based insurance design.

### 2.3.1 Bibliometric analysis

First, it is vital to observe that around 50% of the works analyzed were published since 2018, denoting the increased interest in the topic. The average number of citations per year per article demonstrates an increasing impact of weather index insurance in the literature (Figure 2.2a). However, the global distribution is concentrated in Europe, USA/Canada, and Asia. The role of Latin America/Caribbean, Australia/New Zealand/Oceania, and Africa is much lower, representing less than 10% of the published papers each.

In addition, international collaboration is a critical factor for high-impact scientific studies. Two important countries to analyze in this aspect are Russia and China. In Russia, more than 90% highly cited articles were published in an international setting (Pislyakov e Shukshina, 2014). Similarly, in China, 47% of the highly cited papers were written collaboratively on an international basis.

In general, the international cooperative background of these countries opens the way to more innovative research in the field. These publications illustrate how collaborations with international scientists from centers of excellence enhance the dissemination of the study.

The scientific collaboration map (Figure 2.2b) shows strong collaboration networks between the United States, European countries, China, and India. European countries such as Germany, Switzerland, and the Netherlands have played a dominant role in integration and have promoted collaboration with Kenia, Ethiopia, Nigeria, and South Africa. Canada has collaborated with China, Indonesia, the United States and European countries. From this analysis, we conclude that the United States, China, and Germany play dominant positions in scientific collaboration and are the most influential countries.

Keyword analysis revealed that agricultural and crop insurance are well-developed subjects with a substantial impact on index insurance. Furthermore, drought is the most studied hazard, explained by the impacts of droughts on agriculture. It is important to note that since index insurance was designed to be used in agriculture (Miranda, 1991; Skees, 2008a) and, since the concept has gained attention, a wider range of applications could be shown to be feasible.

On the one hand, Latin American countries such as Brazil, Argentina, and Mexico are vital in global food production (Baldos et al., 2020). On the other hand, by conducting a bibliometric analysis of relevant studies from 2010 to 2022, we discovered a low academic engagement between them and the rest of the world. This is incoherent, as these countries suffer the most from extreme event losses due to their solid economic link with climate-dependent primary activities. These findings emphasize the importance of developing index insurance in tropical countries, notably Latin America, to adapt better to climate change. Furthermore, climate change and basis risk are critical in building index insurance. However, these themes need to be further developed in the literature that is analyzed.

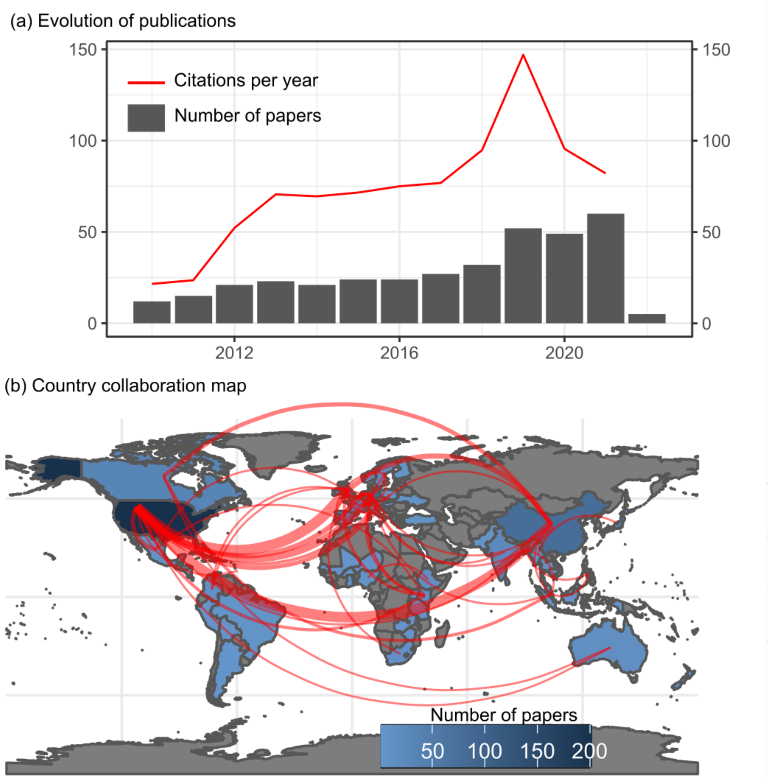


**Figure 2.2.** Weather index insurance studies. (a) Temporal distribution from papers collected from 01/01/2010 to 19/02/2022. (b) Thematic map representing the global collaboration network, where the countries in blue represent the number of studies produced by scientists. The global map indicates the country that the affiliation is located. The darker the color, the more affiliations

Table 2.1 presents a strategic diagram for analyzing clusters of keywords (referred to as themes) according to their centrality and density values. According to Cobo et al. (2011), the density is a measure of the development of the theme, and the centrality is the importance of the theme in the development of the whole field of research we analyzed. According to their densities and centralities, the themes are divided into four classes: basic, motor, niche, and emerging and declining themes. These will be analyzed in the following paragraphs.

The basic themes represent relevant keyword clusters for all the documents analyzed. These include the following clusters: "index-based insurance", "climate change", "index insurance", and "basis risk". Climate change has been a major concern for decision makers, especially

**Table 2.1.** Thematic mapping of the documents based on the conceptual structure of the author's keywords divided into seven clusters with word frequency higher than 40 words according to centrality (the relevance of the theme in the development of the field) and density (the development of the field)

| Theme | Cluster | Density | Centrality |
|---|---|---|---|
| Motor | Insurance | 9 | 7 |
| | Crop insurance | 7 | 5 |
| Niche | Weather derivative | 2 | 8 |
| Emerging and Declining | Weather index insurance | 3 | 4 |
| | Index-based insurance | 5 | 3 |
| Basic | Basis risk | 8 | 1 |
| | Climate change | 6 | 2 |

risk management. Climate change might lead some regions toward higher risk profiles, increasing their vulnerability and expected losses. Therefore, this theme represents an opportunity to develop index insurance for agriculture. Basis risk is a primary topic that requires more development. Although it is a well-known bottleneck in the field, and our analysis suggests room for improvement. More attention should be paid to this topic in future studies.

Motor themes encompass keywords that are relevant to the entire insurance theme and that are well developed. Since they present strong centrality and high density, the clusters "insurance", "agriculture", and "risk management" and the cluster "crop insurance" are conceptually related to almost all papers gathered in the bibliometric analysis. This result confirms that agricultural and crop insurance are the most explored topics in the index insurance field.

Emerging and declining themes are related to themes that represent the combination of low levels of development and are marginal to the entire field of research. This quadrant includes "weather index insurance", which is a critical issue in terms of impact, since extreme weather events trigger significant disasters worldwide. Given the need to manage weather extremes and their importance to a broader geophysical community, weather index insurance is an emerging topic that will gain more attention in the coming years.

Finally, niche themes encompassed the term "weather derivatives". It is considered a niche theme because it is well developed, but has a marginal impact in the field. The themes have fundamental distinctions and similarities. Derivatives are traded Over the Counter (OTC) or on the Chicago Mercantile Exchange (CME). Index insurance is a product offered by insurance and reinsurance companies. They theoretically have a similar principle: a risk-averse individual pays a premium for a risk-bearing individual.

### 2.3.2 Systematic literature review

The food security concept is rooted in the pillars of availability, access, and utilization (Barrett, 2010). Food production can affect availability, while access to renewable or sustainable energy can facilitate proper food transportation and storage. Not all research in the systematic review could be thoroughly examined due to a lack of information on their applications. Therefore, for the 26 articles with complete information, we conducted an overview of the application and most relevant characteristics of index insurance for food security in three main categories:

(i) agricultural; (ii) hydrological; and (iii) sustainable energy insurance. Table 2.2 presents the results of this analysis.

We observed that most studies evaluated insurance in different spatiotemporal aggregations, such as crop insurance, which was analyzed at the farm level by governmental agencies, insurance companies or surveys. In many countries, for example, agricultural data is aggregated at a regional scale, that is, municipality, department, state and country, without standardization. The size of the properties varied greatly, from 5 to 400 hectares, and the total coverage was up to 1.6 million hectares.

Forestry insurance covers larger areas and uses remote sensing data to assess risk. Therefore, spatial discretization is performed at a pixel level. The catchment level was the spatial unit of hydrological insurance and the coverage included all hydrological processes that occurred upstream of the reservoir. For the sustainable energy insurance, wind and solar power insurance, a unique point, representing the location of the windmills and solar panels, was evaluated.

The temporal scale in which the insurance was purchased varied from seasonal to annual. Crop insurance is typically contracted before the sowing period and reaches maturity at the end of the crop cycle. Sectors that are continuously exposed to natural hazards are operated on an annual basis.

The insurance premiums were represented using different units. However, most of the works focused on premiums per unit of area and unit of cost. The crop insurance premium varied from \$6.18 to \$55.26 USD per hectare. This value was affected mainly by the cost of production and farmers' degree of risk aversion. A value of \$187.29 USD per ton of crop and 3 to 7% of production costs was also found.

In contrast, the hydrological insurance for water supply represents values of \$10.48, and the irrigation insurance ranges from \$212.83 to \$333.07 USD per hectare. The prices for irrigation were inconsistent with crop insurance. This might be related to irrigation costs. Sustainable insurance presented premium rates ranging from 0.35 to 0.50 of production costs or a percentage of \$0.033144 per kWh. A detailed analysis of hazard identification, vulnerability analysis, and financial methods of the reviewed paper is presented in sequence.

#### 2.3.2.1 Hazard assessment

The index insurance literature contains a wider range of hazards and indices analyzed (Table 2.2). Drought is the most frequent hazard (77%), followed by excessive rainfall and flood (27%), temperature variation: heat and cold waves (4%), wildfires (4%), low wind speed (4%), and lack of solar radiation (4%). Furthermore, 27% of the studies have a multi-hazard interaction, with drought and excessive rainfall being the most common, followed by wildfire and excessive rainfall, and hydrological drought and floods.

Most studies focused on drought, which is consistent with the findings of Abdi et al. (2022), who identified drought as the leading risk when studying index insurance for crop production. This is due to the fact that drought is the most damaging threat in the agricultural sector, and the sector was the driving subject of the evaluated studies. Our finding is consistent

**Table 2.2.** Main categories for index-based insurance and specific application. Indices: Cumulative Precipitation Index (CPI); Water Storage (WS); Water Deficit (WD); Normalized Difference Vegetation Index (NDVI); Soil Moisture Index (SMI); Standardized precipitation evapotranspiration index (SPEI); Standardized Precipitation Index (SPI); El Niño Southern Oscillation (ENSO); Evaporative Stress Index (ESI); Ped Drought Index (PDI); Ribéreau-Gayon and Peynaud hydrothermal scale (RGP); R2mm; Berman and Levadoux (BBL); High Temperature (HT); Low temperature (LT); Ribéreau-Gayon and peynaud hydrothermal scale (RGP); Visible Infrared Imaging Radiometer Suite (VIIRS); Solar Radiation (SR); Wind Speed (WSpeed)

| Category | Hazard | Index | Threshold | Author |
|---|---|---|---|---|
| Agricultural | Drought | CPI | Expected yield | Turvey et al. (2019) |
| | | | Expected yield | Roznik et al. (2019) |
| | | | 30th percentile | Kath et al. (2019) |
| | | | 50-85th percentile | Awondo (2019) |
| | | | Expected yield | Ricome et al. (2017) |
| | | CPI, NDVI | Customized trigger | Eze et al. (2020) |
| | | CPI, SPI, SPEI, SMI, ESI | Expected yield | Bucheli et al. (2021) |
| | | ENSO | Expected yield | Mortensen e Block (2018) |
| | | PDI, CPI | Expected yield | Bokusheva (2018) |
| | | SMI | Expected yield | Vroege et al. (2021) |
| | | SPI, SPEI | Expected yield | Hohl et al. (2020) |
| | | WD | Expected yield | Gómez-Limón (2020) |
| | Drought and excessive rainfall | BBL, RGP | 50th-90th percentile | Martínez Salgueiro (2019) |
| | | CPI | 50th percentile | Martínez-Salgueiro and Tarrazon-Rodon (2020) |
| | | CPI, R2mm | Customized trigger | Shirsath et al. (2019) |
| | Drought and excessive rainfall | DOWKI | 10th and 90th percentile | Kapsambelis et al. (2019) |
| | Excessive rainfall | CPI | 94th percentile | Furuya et al. (2021) |
| | | | 70th - 95th percentile | Kath et al. (2018) |
| | Temperature variation (high and low) | HT, LT | 5-year moving average yield | Guo et al. (2019) |
| | Wildfires and excessive rainfall | WSpeed and VIIRS | Customized trigger | Sacchelli et al. (2018) |
| Hydrological | Drought | WD | Q710 | Mohor e Mendiondo (2017) |
| | | | 70th-90th percentile | Denaro et al. (2018) |
| | Hydrological drought | WS | Expected yield | Guerrero-Baena and Gómez-Limón (2019) |
| | Hydrological droughts and floods | WD, PF | Expected yield | Denaro et al. (2020) |
| Sustainable Energy | Lack of solar ratiation | SR | Customized trigger | Boyle et al. (2021) |
| | Low wind speed | WSpeed | Customized trigger | Rodríguez et al. (2021) |

with the fact that drought-induced yield losses occurred in three-fourths of the global harvested regions between 1983 and 2009 (Kim et al., 2019).

Index insurance is a promising methodology for designing insurance models because it avoids the high administrative costs, adverse selection, and moral hazard issues associated with traditional indemnity-based insurance. The behavior of an index can characterize the variability of a hazard. However, its performance in covering losses is highly dependent on the index chosen.

The index choice is a critical phase in index insurance modeling, since a mismatch between the index and the actual loss might increase basis risk. Another significant element that influences performance is geographic basis risk, which occurs when the insurance index is based on a location other than the insured location. This difficulty emerges, for example, in agricultural insurance when the index is derived from a meteorological station placed in a location that does not adequately represent the insured region.

Because it is challenging to provide specific index insurance contracts to small regions, geographical basis risk is generally unavoidable (Odening e Shen, 2014). Another option to mitigate this risk is to utilize decorrelation functions and an insurance portfolio composed of contracts for various regions (Norton et al., 2013). Figueiredo et al. (2018) proposes a probabilistic framework to tackle uncertainties in trigger selection and damage modeling, therefore improving basis risk quantification, evaluation and communication.

Approximately 50% of the analyzed studies used a rainfall-based drought indicator (i.e., CPI, SPI, CDD, and R2mm) to indicate drought and excessive rainfall. This straightforward strategy indicates drought conditions and only requires precipitation data. In addition, these indices have the advantage of representing both water deficit and water excess.

Rainfall indices in the agricultural sector might considerably represent low-yield occurrences, in both deficit and excess forms, which would correlate well with low-yields (Abdi et al., 2022). Indices such as the SPI have the advantage of being determinable using only Precipitation data. However, the SPI's application is limited when issues related to the water balance must be considered in the analysis of the problem.

As an alternative, the Standardized Precipitation Evapotranspiration Index (SPEI)is capable of reflecting the combined effect of rainfall and evapotranspiration on droughts and requires at least temperature data, in the case of the Hargreaves-Samani method (Droogers e Allen, 2002), and radiation, temperature, and relative humidity for reference methods such as FAO Penman-Monteith (Allen et al., 1998). These can be difficult to obtain in developing countries with low density of weather stations.

In addition to data availability, the index's simplicity of computation is a significant factor to consider while choosing it. As a result of their ease of calculation and data input, indices such as CPI, Consecutive dry days (CDD), Water Storage (WS), and Water Deficit (WD) are popular among the studies examined. While data for CPI, CDD, WS, and WD calculations are easily acquired, the spatial distribution of risk has yet to be fully known, even in places with long-established weather stations. This poses a difficulty in the index insurance market: strategically exploiting information from existing stations at geographic locations where the precise weather observations are unknown (Norton et al., 2013).

SPI and SPEI are more complicated indexes that add weather statistics to the analysis. In addition to the complexity of its calculation, these indices require a database of at least 30 years. Indices derived from form modeling techniques and remote sensing are much more challenging to acquire and determine.

In addition to that, complex indices, such as the Soil Moisture Index (SMI) and the Normalized Difference Vegetation Index (NDVI), as well as the use of El Nino Southern Oscillation (ENSO), have become more frequent as modeling techniques and remote sensing have advanced. This popularization is already seen in the current review, as they were used in 11% of the investigated studies.

Although some indices are more popular than others, it is crucial to realize that no index can be used for all contexts and situations. The available data, the degree of drought monitoring, and the time resources available for its determination all influence the index representation of the hazard, as long as the uncertainties (basis risks) are under control.

Other hazards include temperature variation, wildfire, low wind speed, and lack of solar radiation, accounting for 16% of the studies analyzed. The thermal hazard is a growing concern for human health, agriculture production, forestry, and the environment. Similarly to rainfall, temperature extremes can also explain satisfactory yield losses (Abdi et al., 2022).

Given the topic's relevance, we expected more work focusing on this subject. However, we have found only one study on temperature variation insurance. High-Temperature Index (HTI) and Low-Temperature Index (LTI) were proposed by Guo et al. (2019) that focused on computing the number of days the temperature was higher or lower than a certain threshold, representing the rice yield reduction. Regardless of the hazard, we had one study per index. This reflects the diversity of these issues and a need for in-depth research on the specific hazards - temperature variation, fire, wind, and solar radiation.

The hazards are treated as independent occurrences in the multi-hazard interaction. However, not all studies examined one index per hazard. More than half of the multi-hazard studies considered drought and excessive rainfall, and they used the same index to reflect both hazards (I.e., CPI, DOWKI, and R2mm). According to Kapsambelis et al. (2019), simple climatic water balance based on precipitation and evapotranspiration data can simulate both drought and excess rainfall globally.

The authors state that the DOWKI index was able to fit extreme yield anomalies. One example of employing different indexes to reflect different hazards was found on wildfires and excessive rainfall on forestry in Italy (Sacchelli et al., 2018). The authors described forest fires using the Visible Infrared Imaging Radiometer Suite (VIIRS) and explored the effects of strong winds through wind speed (Wspeed).

The advantage of using only one index to represent two hazards is the ease of calculating and implementing the insurance policy. However, not all hazards can be represented by a single index, such as wildfires and excessive rainfall. It is known that multi-hazards and compound events are increasingly intense and significant. The finding of only a few studies connected to the theme reflects a substantial gap in the literature.

#### 2.3.2.2 Vulnerability analysis

In finance and management, insurance is a product that intends to totally or partially reduce or eliminate the loss caused due to different risks (Ejiyi et al., 2022). In order to lower the basis risk and boost the acceptance of insurance products, the prediction models of the link with the index and the damages must be effectively chosen. The vulnerability analysis is focused on modeling the physical damages and losses from an extreme weather event. Thus, we presented a summary of the Expected Loss Amount (ELA) and Expected Annual Damage (EAD) models applied in the reviewed papers, available in Table 2.3. The deterministic models were applied for income reduction impacts, especially crop insurance.

Most of the models applied were related to a unique explanatory variable or index. Due to their simple application and understanding, these models are expected to be the most common, primarily linear regression models. However, these models present the disadvantage of contemplating only one hazard at a time. In contrast, Generalized Additive Linear Models (GALM) and stepwise regression added the possibility of evaluating more than one index. These can then be used in a multi-hazard approach.

**Table 2.3.** Summary of Expected Loss Amount (ELA) and Expected Annual Damage (EAD) models

| Type of impact | Type of loss model | Authors |
|---|---|---|
| Expected loss amount (ELA) | Cluster analysis | Eze et al. (2020) |
| | Linear regression | Bokusheva (2018)<br>Bucheli et al. (2021)<br>Furuya et al. (2021)<br>Gómez-Limón (2020)<br>Guerrero-Baena e Gómez-Limón (2019)<br>Guo et al. (2019)<br>Hohl et al. (2020)<br>Kath et al. (2019)<br>Mortensen e Block (2018)<br>Denaro et al. (2020) |
| | GALM | Awondo (2019)<br>Kath et al. (2018) |
| | Stepwise regression | Shirsath et al. (2019) |
| | Copula | Bokusheva (2018)<br>Kapsambelis et al. (2019)<br>Martínez Salgueiro (2019) |
| Expected annual damage (EAD) | Empirical curve | Mohor e Mendiondo (2017)<br>Sacchelli et al. (2018) |

A multi-hazard approach requires understanding the frequency and magnitude of multiple hazards and the possibility of them occurring simultaneously. Multi-hazard risk index insurance papers analyzed presented combinations of drought and excessive rainfall for crop insurance (Kapsambelis et al., 2019; Shirsath et al., 2019), fire and storms for forest insurance (Sacchelli et al., 2018), temperature variation and excessive rainfall for crop insurance (Martínez Salgueiro, 2019), and high and low temperatures for crop insurance (Guo et al., 2019). The assumption of independence was considered prior knowledge by Martínez Salgueiro (2019) and Guo et al. (2019). However, the authors did not provide a mathematical proof of this choice. Instead, they prioritized hazards according to their frequency and magnitude using preexistent risk maps.

Another possibility of incorporating hazards interactions is through copulas. This has been incorporated in the loss modeling by Kapsambelis et al. (2019) and Martínez Salgueiro (2019). The copula theory (Nelsen, 2006) is widely used for multi-hazard analysis since it derives joint probability distributions from marginal distributions. Briefly, marginal distributions are not required to follow the same probability distribution model, giving flexibility and robustness to analyze the interaction of more than two marginal distributions.

Additionally, losses and damages databases are not necessarily Gaussian-normally distributed in copula models, which is common in crop production. This way, a significant amount of skewed information can be well embodied, depicting better hydrological and meteorological extremes than linear regression models.

In addition, to embody loss and damages originating from multi-hazard events, another important aspect of vulnerability assessment is the representation of complex patterns in the loss and damage series modeling. In this regard, linear regressions are commonly applied in loss forecasting due to their simplicity.

Generalized additive linear models (GALM) add a link function to express a linear relationship between more than one variable (Blier-Wong et al., 2020), making it possible to express both multi-hazard risk phenomena and simple non-linear effects. While simple and easily explainable, such models may be ineffective in learning complex patterns in the data, which are common in food production. Machine learning (ML) can enable optimal formulation of insurance policies when applied in the insurance field. These models help capture complex, non-linear, and high-dimensional interactions between indices and losses (Blier-Wong et al., 2020).

ML techniques are still emerging in loss models, and here we have reviewed only a very recent paper (from 2020) that used cluster analysis. The article provides evidence that ML techniques can improve loss modeling from different sources and present different time and spatial scales (Eze et al., 2020). Blier-Wong et al. (2020) emphasized that ML applications in actuarial science are expanding rapidly and show great promise.

Frees et al. (2014) affirm that, with greater data availability and robustness of data in ML algorithms, more heterogeneity represented by insured individuals can be captured, accurately representing their vulnerability. With these models becoming more common in future studies, the multi-hazard assessment could be better incorporated. Although promising, applying ML techniques to model loss and damages in the insurance sector may be bottlenecked, especially in developing countries, by the need for qualified personal and powerful GPUs.

When vulnerability studies or data sets for the quantification of losses and damages in specific sites are not available, the insurer can use empirical functions or crop modeling techniques. Monteleone et al. (2022) reviewed methods used to model the functional relationship between a given extreme event and crop losses. They also highlight the need to study crop vulnerability to other less studied climate-related hazards, such as extreme temperatures. Regarding index insurance design, we found a paper that presented empirical functions based on the assumption of linearity between water deficit and losses. Mohor e Mendiondo (2017) used the water deficit volume to the $Q_{7,10}$ reference flow for predicting the impact of the water shortage on the water supply, irrigation, livestock and ecological sectors. In crop insurance, if the historical yield losses database is non-nexistent or available at a high level of aggregation, crop modeling shows promise in estimating yield while utilizing different explainable variables.

#### 2.3.2.3 Financial methods and risk pricing

The impact provided in the vulnerability analysis module can be translated as expected values of damage, income reduction, or business interruption by financial methods. The reviewed articles presented burning rate, probabilistic fit, and index modeling as the most prevalent risk pricing models. They commonly use mean historical losses to estimate expected future losses for similar sectors (Sant, 1980).

The expected losses are called pure risk premiums and are the primary concern in index insurance papers. Historical losses are converted into payouts considering two critical variables (i) strike value K and (ii) degree of coverage dc. The K is the index value that triggers payouts, proportional to risk aversion and the degree of coverage. The risk aversion is reflected in the degree of coverage, for example, dc ranging from 0 to 1, being 0 with no protection and 1 with complete protection. These variables represent the behavior and aversion of policyholders towards a particular risk and will be the key to define the premium and indemnity values.

The loss expectation can be determined using the historical burn rate method (HBR), which is based on the observation of historical losses (Guerrero-Baena e Gómez-Limón, 2019; Hohl et al., 2020; Mortensen e Block, 2018; Shirsath et al., 2019). This method is widely applied in the insurance industry. However, sufficient data is required to be accurate. For smaller datasets considering uncertainty, expected values can be evaluated by fitting loss data to a probability density function (Bokusheva, 2018; Bucheli et al., 2021; Kath et al., 2019; Eze et al., 2020; Kath et al., 2019; Martínez Salgueiro, 2019; Sacchelli et al., 2018; Vroege et al., 2021). This procedure helps to improve pure risk premium rates by accounting for the probability of extreme events that have not been recorded. The probability distribution of loss data presents distortions in the tails, leading to underestimating pure risk premiums.

In addition, insurance companies present non-traded assets that add costs to final premium rates. The transformation method proposed by Wang (2002), also referred to as Wang Transform, takes into account the impact of non-traded assets on premium rates, and the method was applied by Boyle et al. (2021) and Denaro et al. (2018).

Other approaches for defining contract payments are based on a probabilistic fit. Bokusheva (2018) applied the Marginal Expected Shortfall (MES) method, which is a conditional probability modeling in which payouts are given when the target variable exceeds the strike value. On the contrary, Eze et al. (2020) used a cluster analysis that associates NDVI and weather variables with higher-results observations.

It is well known that climate variables present a certain degree of uncertainty when predicted. Therefore, this aspect must be considered when estimating losses caused by climate-related losses Smith e Matthews (2015). In the literature, a stochastic approach based on Monte Carlo simulations is used to address the problem. A Monte Carlo simulation is the basis of the index modeling method applied by Gómez-Limón (2020), Guo et al. (2019), Kapsambelis et al. (2019), Gómez-Limón (2020), Mohor e Mendiondo (2017), and Rodríguez et al. (2021). The generation of synthetic weather time series improves the understanding of climate uncertainty in terms of confidence intervals. A summary of risk pricing methods is described in Supplementary Information S1.

Econometric models provide values that guide decision makers in understanding the price of the risk. However, it is fundamental to evaluate the risk reduction performance of index insurance. The simulation of cash flows allows us to understand the effectiveness of the insurance policy hedging. Nevertheless, this efficiency depends on the point of reference adopted by the modeler.

The effectiveness problem arises when policyholders and insurance companies have different and often competing objectives. On the one hand, policyholders want to protect their assets at risk to prevent going out of business. On the other hand, insurance companies want to maximize profits to comply with the interests of their investors and shareholders. Since information asymmetry and moral hazards are allegedly minimized in the case of index insurance (Barnett et al., 2008; Mußhoff et al., 2018), the costs associated with moral hazards can be neglected from the pricing of premium rates.

The cash flow equation is a standard tool for evaluating the capital of companies and people. Simulation of cash flows using expected revenue and payouts as assets and premiums as liability for policyholders is used for evaluating the effectiveness of the index insurance policy (Bokusheva, 2018; Boyle et al., 2021; Kath et al., 2019; Martínez Salgueiro, 2019). For insurance companies, the cash flow changes direction, i.e., premiums are considered assets and payouts are a liability. This was used to calculate the loss ratio by (Mohor e Mendiondo, 2017).

Other authors have applied utility theory to evaluate insurance policies. Utility theory accounts for individual behavior and preferences in economic analysis and is based on several assumptions that apply to a group of individuals (Kahneman e Tversky, 1979). Some authors (Bucheli et al., 2021; Eze et al., 2020; Furuya et al., 2021; Vroege et al., 2021) used the concept of risk-averse utility functions for policyholders, where the asset's utility at risk is concave or diminishing. Detailed information on insurance policy evaluation methods is given in Supplementary Information S2.

It is worth mentioning that either the method employed to estimate fair premium values and to tackle future risk increases due to climate change (CC) scenarios, the insurance market should consider some of the following points to reduce their weaknesses or uncertainty sources. First, as mentioned before, to calculate the premium value, it converges into a multi-objective problem.

The insured could contract a long-term insurance policy with an established premium. However, due to the uncertainty of the CC, some extreme events could not occur, and the insured paid too much for unnecessary coverage, resulting in more profit for the insurer. However, the opposite case could occur. In that case, the insured pays less and extreme events happen, and the insurer does not have the liquidity to pay losses.

Second, a layered insurance scheme that includes the private and public sectors (PP) (Keskitalo et al., 2014; Paudel et al., 2015) to cope with extreme losses. This means that when a certain threshold of loss is reached, a second partner will pay the difference in the indemnities. However, the definition of that threshold is another literature gap, similar to the strike value K.

Third, according to the spatial scale, a pool risk scheme is preferable to reduce premium values, which requires cooperation among stakeholders. However, the main issue is to achieve complete diversification of the portfolio (Porth et al., 2016). Fourth, this induces risk reduction proposals to increase resilience and promote adaptation within the sector. The latter could be reached through financial incentives, such as premium discounts offered to stakeholders when they adopt some mitigation measure, as shown in (Hudson et al., 2016).

Finally, for a multi-hazard scheme, the schemes mentioned earlier should be calculated for each hazard. However, the premium values will be different and a weighted procedure will be required, such as that done with the hazard frequency by Martínez Salgueiro (2019) and Guo et al. (2019) as mentioned in Section 3.2.2.

### 2.3.3 Conceptual framework

Based on the results discussed, we present a conceptual framework for multi-hazard risk index-based insurance design delineating various paths through which insurance policies can be developed. As shown in Figure 2.3, the index-based insurance design can be divided into three modules: (i) hazard identification; (ii) vulnerability assessment; and (iii) financial methods and risk pricing.

Hazard identification is the process of analyzing and selecting the most critical threats and their respective indices. Vulnerability analysis refers to the process of selecting loss models and thresholds. Finally, financial methods and risk pricing refer to methods to estimate long-term loss expectations and risk premium rates.

This framework (Figure 2.3) generates paths for the design of an insurance policy, and each step indicates a design option supported by the literature. Turvey et al. (2019), for example, followed a path A1-C1-D2-E1-F1, i.e., drought insurance (A1) with a static threshold (C1), a loss model based on losses projected by an index (D2), for single policyholders (E1) for many farmers in a region (F1). Sacchelli et al. (2018) provides another example, presenting a design path A2-B1-C1-D1-E1-F1, that is, multi-hazard risk insurance for wildfires and excessive rainfall (A2), considering the hypothesis of independent events (B1), with a static threshold (C1), using the index value to model losses (D1), and premium rates for a single policyholder (E1) without considering risk pooling (F1).

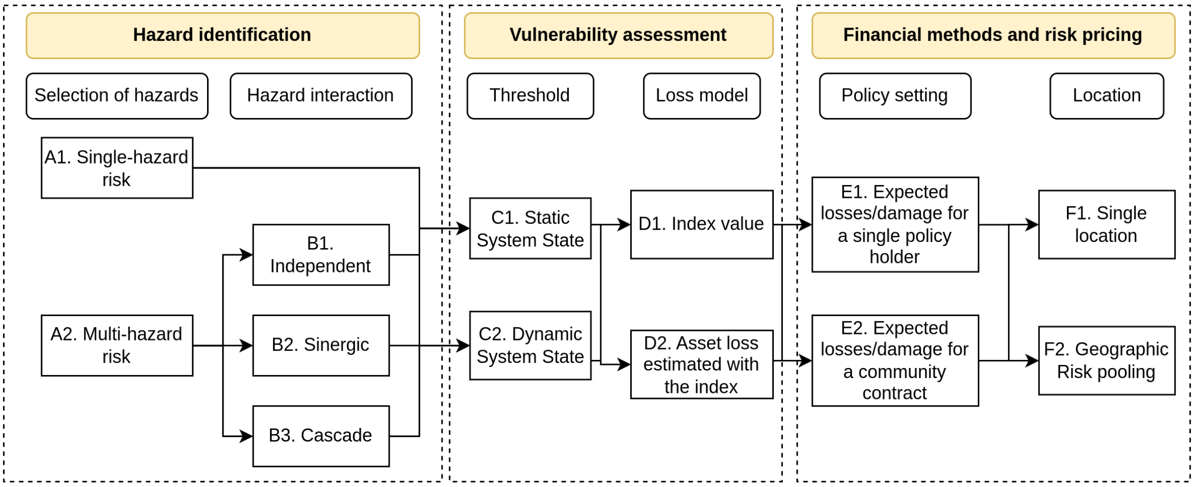


**Figure 2.3.** The framework illustrates the process of selecting and prioritizing hazards, defining index thresholds, modeling losses, and optimizing insurance risk premiums. The vulnerability assessment presents two types of systems: (a) Stationary System State, where both thresholds and hazards are stationary and represent an analysis based on observed historical information, and (b) Non-stationary System State, where both indices and hazards are non-stationary and reflect a combination of observed historical and projected data. The non-stationary system anticipates a potential increase in risk and optimizes risk premiums.

**Hazard identification** The previous discussion demonstrated that significant decisions in this step are whether to use a single (A1) or multi-hazard (A2). Single hazard risk is straightforward and should be reserved for situations where one hazard is dominant in a region. However, research has shown that single-hazard hypothesis and multi-hazard risks might include independent, synergistic, and cascade events.

According to (Gill e Malamud, 2014), independent hazards (B1) are events that can occur simultaneously in a region without any causal dependence. Synergistic hazards (B2) refer to a situation in which the occurrence of a particular hazard increases the probability of the occurrence of another. Cascade hazards (B3) represent a situation when an event triggers the occurrence of another event (that is, excessive rainfall that causes landslides in a particular region).

Alternative multi-hazard interaction hypotheses must be tested to depict weather and climate-related losses (Tilloy et al., 2019) and can influence the correct interpretation of loss modeling. The papers that addressed multi-hazard risks assumed independence between the events investigated; however, the combination of drivers and impacts of hazards contributes to risk analysis and is responsible for the occurrence of the most severe weather and climate-related impacts(Zscheischler et al., 2020). To overcome this challenge, Tilloy et al. (2019) presented methods to test multiple risk hypotheses, including copulas, classification algorithms, linear regression and physical models. The importance of these models will be explored in the illustrative example that can be visualized in Appendix .

**Vulnerability analysis** The vulnerability analysis for insurance design is translated into threshold definition and loss model selection. Thresholds, strikes, and triggers are all terms for the pre-agreed-upon index value that triggers claim payments when reached or exceeded. In the literature, we identified loss models represented in terms of index value (D1) or losses estimated with index values (D2).

The increasing frequency of extreme climate events has forced insurers to increase premium rates and threatens coverage availability. Losses and damages associated with extreme events have multiple drivers (Zscheischler et al., 2020), implying that losses have multiple thresholds and are associated with multiple variables. These thresholds vary with time and space (Hoek van Dijke et al., 2022).

As discussed in the vulnerability section, the selection of thresholds and consequential loss modeling consists of evaluating historical events. This creates a system that we call the stationary system state (C1), characterized by fixed thresholds even when multiple hazards are considered. The second case is the Nonstationary System State (C2). The frequency and severity of hazards change over time and the thresholds are dynamic, indicating either improvement or deterioration in resilience.

We proposed a conceptual framework for considering multi-hazard risk analysis that allowed us to analyze the interaction between hazards and two types of vulnerability: static and dynamic resilience (Figure 2.4). Static resilience refers to a stationary state system. Most of the papers represent this case. Dynamic resilience refers to a Non-Stationary State System and considers changing hazard patterns and vulnerability thresholds. Considering a Non-Stationary State System helps to anticipate increasing patterns of losses, therefore optimizing risk premiums to accelerate the adaptation and resilience of farmers against climate change.

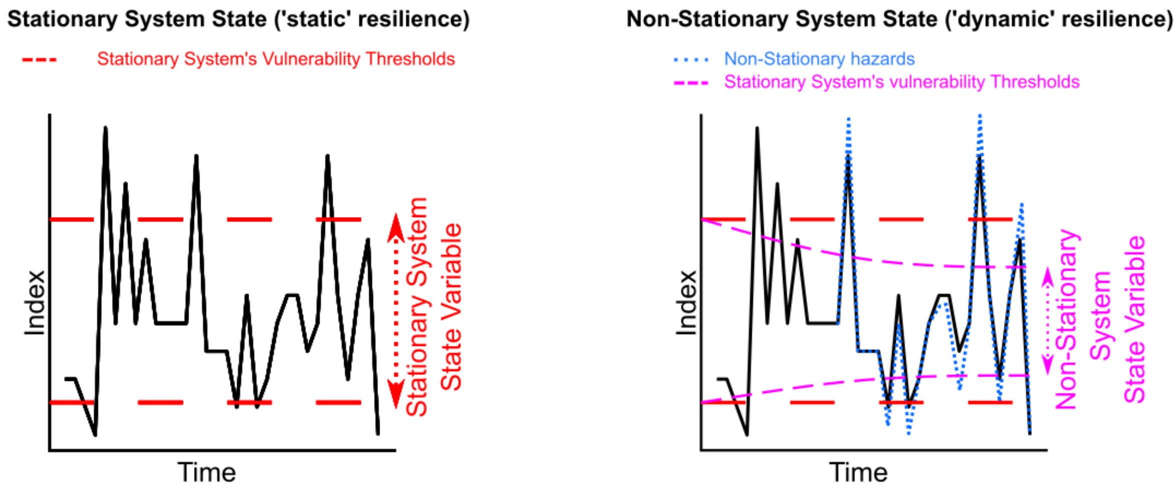


**Figure 2.4.** Graphical representation of static and dynamic state resilience systems

The dynamic threshold considers future scenarios that might assist in avoiding risk re-assessment by anticipating and diluting potential severe climate shocks. Shifts in frequency and severity of extreme events are evaluated using the Representative Concentration Pathways (RCP) (Van Vuuren et al., 2011) indicating possible changes in risk exposure. Shared Socioeconomic Pathways (SSP) (Riahi et al., 2017) will help us to understand risk in different vulnerability trajectories, that is, increasing, stationary, and decreasing resilience.

**Financial methods and risk pricing** Financial methods require defining parameters relevant to policy implementation, which depend on regional socioeconomic factors. Single contracts (E1) focus on individual policy and were the general option described in the literature review (Table 2.2). Community contracts (E2) are widely used for microinsurance contracts when small farmers are associated with government institutions, associations or companies to access affordable insurance policies and long-term financial support against weather and climate-related threats (Platteau et al., 2017).

Individual and community contracts can be tailored to a single region (F1), or different locations can be pooled (F2). In the derivative market, several locations can be insured in the same contract using the same weather index. This type of contract has been applied by retail (Štulec et al., 2019a) and is suitable for farmers and companies with operations in more than one location.

## 2.4 Conclusions

In conclusion, this study has outlined the relevance of index insurance in mitigating the impacts of climate extremes, with a particular focus on improving food security. This study reviewed the development and design of index insurance focusing on multi-hazard risk analysis and food security. We summarize the main methods of hazard analysis and index calculation, loss modeling, and risk pricing. We observed that the lack of studies on multi-hazard risks is the central gap in the literature.

Drought was the most researched hazard and the Cumulative Precipitation Index (CPI) was the index most used in the development of drought index insurance. The literature review also examined other hazards, such as excessive rainfall, temperature fluctuations, wind, and radiation. As food security is a complex concept, agricultural, hydrological, and sustainable energy insurance was also evaluated. In terms of hazard assessment, we conclude that future studies can incorporate multi-hazard risk analysis in order to improve the climate risk protection. The frequency and magnitude of climate extremes is expected to increase due to climate change; therefore, there is a growing demand for methods that help to select the most relevant hazards and their impact on food production. This is a learning process that is required to further advance impact studies in terms of providing methods that are generalized methods that can be applied globally and that can generate knowledge about multi-hazard impact and interactions that is relevant regionally.

The vulnerability analysis for insurance design was composed of the selection of a loss model and the definition of threshold values. Multi-hazard loss models were composed of generalized or separate additive models to calculate losses caused by different hazards, considering that the hazard occurrence was independent. Composite indices such as DOWKI or ambivalent indices such as CPI and SPI could capture excessive rainfall and droughts and are suitable for analyzing extreme conditions. Nevertheless, the reviewed papers did not fully explore other hypotheses of multi-hazard interaction, such as synergistic and cascade events.

Trigger values - the index value that triggers payouts - were attributed to both index values or losses estimated by the index. In the first case, thresholds were defined by a quantile or a range of quantiles, e.g., 70 to 80th quantiles. The threshold value depends on how risk-averse the policyholder is. The case of losses estimated by the index was typical in crop insurance, where the threshold is a percentage of the expected crop yield for a given year.

The determination of risk premiums followed methods based on historical data evaluation. These methods are based on the assumption that historical data provide enough information to characterize regional risk. However, recent findings (Cremades et al., 2018) demonstrate that this approach could lead to an underestimation of future risk. In the in-depth analysis of the most relevant articles, we found that the burning rate, probabilistic fit, and index modeling are the most prevalent risk pricing models.

We proposed a conceptual framework that incorporates hazard identification, vulnerability assessment, and financial methods and risk pricing. This paper outlines a series of steps that can be taken to guide the design of index insurance. We have connected these steps to the methods discussed in the literature. This framework is recommended for the field of food security, however, the design paths described in the framework can be applied to other fields.

Our paper demonstrates that, despite index insurance for food security has gained attention in the past years, there are still weaknesses and limitations that must be addressed in future work, e.g. a clear definition and analysis of multiple hazards instead of assuming single hazard risk; testing different hypotheses of the interaction between hazards, especially for coupled moisture-thermal events; evaluating how the multi-hazard risk selection affects basis risk; and analyzing the trade-offs between loss model accuracy and the policyholders willingness to pay.

## References

Abdi, M. J., Raffar, N., Zulkafli, Z., Nurulhuda, K., Rehan, B. M., Muharam, F. M., Khosim, N. A., e Tangang, F. (2022). Index-based insurance and hydroclimatic risk management in agriculture: A systematic review of index selection and yield-index modelling methods. *International Journal of Disaster Risk Reduction*, 67:102653.

Akter, S. (2012). The Role of Microinsurance as a Safety Net Against Environmental Risks in Bangladesh. *The Journal of Environment & Development*, 21(2):263–280.

Allen, R. G., Pereira, L. S., Raes, D., Smith, M., et al. (1998). Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56. *Fao, Rome*, 300(9):D05109.

Aria, M. e Cuccurullo, C. (2017). bibliometrix: An r-tool for comprehensive science mapping analysis. *Journal of informetrics*, 11(4):959–975.

Awondo, S. N. (2019). Efficiency of region-wide catastrophic weather risk pools: Implications for African Risk Capacity insurance program. *Journal of Development Economics*, 136:111–118.

Baldos, U., Haqiqi, I., Hertel, T., Horridge, M., e Liu, J. (2020). SIMPLE-G: A multiscale framework for integration of economic and biophysical determinants of sustainability. *Environmental Modelling & Software*, 133:104805.

Barnett, B. J., Barrett, C. B., e Skees, J. R. (2008). Poverty Traps and Index-Based Risk Transfer Products. *World Development*, 36(10):1766–1785.

Barrett, C. B. (2010). Measuring food insecurity. *Science*, 327(5967):825–828.

Blier-Wong, C., Cossette, H., Lamontagne, L., e Marceau, E. (2020). Machine learning in p&c insurance: A review for pricing and reserving. *Risks*, 9(1):4.

Bokusheva, R. (2018). Using copulas for rating weather index insurance contracts. *Journal of Applied Statistics*, 45(13):2328–2356.

Boyle, C. F., Haas, J., e Kern, J. D. (2021). Development of an irradiance-based weather derivative to hedge cloud risk for solar energy systems. *Renewable Energy*, 164:1230–1243.

Bucheli, J., Dalhaus, T., e Finger, R. (2021). The optimal drought index for designing weather index insurance. *European Review of Agricultural Economics*, 48(3):573–597.

Cobo, M., López-Herrera, A., Herrera-Viedma, E., e Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics*, 5(1):146–166.

Cremades, R., Surminski, S., Máñez Costa, M., Hudson, P., Shrivastava, P., e Gascoigne, J. (2018). Using the adaptive cycle in climate-risk insurance to design resilient futures. *Nature Climate Change*, 8(1):4–7.

Denaro, S., Castelletti, A., Giuliani, M., e Characklis, G. (2018). Fostering cooperation in power asymmetrical water systems by the use of direct release rules and index-based insurance schemes. *Advances in Water Resources*, 115:301–314.

Denaro, S., Castelletti, A., Giuliani, M., e Characklis, G. (2020). Insurance Portfolio Diversification Through Bundling for Competing Agents Exposed to Uncorrelated Drought and Flood Risks. *Water Resources Research*, 56(5).

Droogers, P. e Allen, R. G. (2002). Estimating reference evapotranspiration under inaccurate data conditions. *Irrigation and drainage systems*, 16:33–45.

Ejiyi, C. J., Qin, Z., Salako, A. A., Happy, M. N., Nneji, G. U., Ukwuoma, C. C., Chikwendu, I. A., e Gen, J. (2022). Comparative analysis of building insurance prediction using some machine learning algorithms. *International Journal of Interactive Multimedia and Artificial Intelligence ...*.

Eze, E., Girma, A., Zenebe, A., e Zenebe, G. (2020). Feasible crop insurance indexes for drought risk management in Northern Ethiopia. *International Journal of Disaster Risk Reduction*, 47.

FAO (2003). Chapter 2. food security: concepts and measurement. *Trade Reforms and Food Security. Conceptualizing the Linkages.*

FAO (2014). The water-energy-food nexus a new approach in support of food security and sustainable agriculture.

Figueiredo, R., Martina, M. L., Stephenson, D. B., e Youngman, B. D. (2018). A probabilistic paradigm for the parametric insurance of natural hazards. *Risk Analysis*, 38(11):2400–2414.

Foster, B., Kern, J., e Characklis, G. (2015). Mitigating hydrologic financial risk in hydropower generation using index-based financial instruments. *Water Resources and Economics*, 10:45–67.

Frees, E. W., Derrig, R. A., e Meyers, G. (2014). *Predictive modeling applications in actuarial science*, volume 1. Cambridge University Press.

Furuya, J., Mar, S. S., Hirano, A., e Sakurai, T. (2021). Optimum insurance contract of flood damage index insurance for rice farmers in Myanmar. *Paddy and Water Environment*, 19(2):319–330.

Ghosh, R. K., Gupta, S., Singh, V., e Ward, P. S. (2021). Demand for Crop Insurance in Developing Countries: New Evidence from India. *Journal of Agricultural Economics*, 72(1):293–320.

Gill, J. C. e Malamud, B. D. (2014). Reviewing and visualizing the interactions of natural hazards. *Reviews of Geophysics*, 52(4):680–722.

Gómez-Limón, J. A. (2020). Hydrological drought insurance for irrigated agriculture in southern Spain. *Agricultural Water Management*, 240:106271.

Guerrero-Baena, M. e Gómez-Limón, J. (2019). Insuring Water Supply in Irrigated Agriculture: A Proposal for Hydrological Drought Index-Based Insurance in Spain. *Water*, 11(4):686.

Guo, J., Mao, K., Zhao, Y., Lu, Z., e Xiaoping, L. (2019). Impact of Climate on Food Security in Mainland China: A New Perspective Based on Characteristics of Major Agricultural Natural Disasters and Grain Loss. *Sustainability*, 11(3):869.

Guzmán, D. A., Mohor, G. S., e Mendiondo, E. M. (2020). Multi-Year Index-Based Insurance for Adapting Water Utility Companies to Hydrological Drought: Case Study of a Water Supply System of the Sao Paulo Metropolitan Region, Brazil. *Water*, 12(11):2954.

Halcrow, H. G. (1949). Actuarial Structures for Crop Insurance. *Journal of Farm Economics*, 31(3):418.

Hillier, J. K., Matthews, T., Wilby, R. L., e Murphy, C. (2020). Multi-hazard dependencies can increase or decrease risk. *Nature Climate Change*, 10(7):595–598.

Hoek van Dijke, A. J., Herold, M., Mallick, K., Benedict, I., Machwitz, M., Schlerf, M., Pranindita, A., Theeuwen, J. J., Bastin, J.-F., e Teuling, A. J. (2022). Shifts in regional water availability due to global tree restoration. *Nature Geoscience*, 15(5):363–368.

Hohl, R., Jiang, Z., Vu, M. T., Vijayaraghavan, S., e Liong, S.-Y. (2020). Using a regional climate model to develop index-based drought insurance for sovereign disaster risk transfer. *Agricultural Finance Review*.

Hudson, P., Botzen, W. W., Feyen, L., e Aerts, J. C. (2016). Incentivising flood risk adaptation through risk based insurance premiums: Trade-offs between affordability and risk reduction. *Ecological Economics*, 125:1–13.

IPCC (2022). *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge.

Kahneman, D. e Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263.

Kapsambelis, D., Moncoulon, D., e Cordier, J. (2019). An Innovative Damage Model for Crop Insurance, Combining Two Hazards into a Single Climatic Index. *Climate*, 7(11):125.

Kath, J., Mushtaq, S., Henry, R., Adeyinka, A., e Stone, R. (2018). Index insurance benefits agricultural producers exposed to excessive rainfall risk. *Weather and Climate Extremes*, 22:1–9.

Kath, J., Mushtaq, S., Henry, R., Adeyinka, A. A., Stone, R., Marcussen, T., e Kouadio, L. (2019). Spatial variability in regional scale drought index insurance viability across Australia's wheat growing regions. *Climate Risk Management*, 24:13–29.

Keskitalo, E. C. H., Vulturius, G., e Scholten, P. (2014). Adaptation to climate change in the insurance sector: examples from the uk, germany and the netherlands. *Natural Hazards*, 71(1):315–334.

Kim, W., Iizumi, T., e Nishimori, M. (2019). Global patterns of crop production losses associated with droughts from 1983 to 2009. *Journal of Applied Meteorology and Climatology*, 58(6):1233–1244.

Komendantova, N., Mrzyglocki, R., Mignan, A., Khazai, B., Wenzel, F., Patt, A., e Fleming, K. (2014). Multi-hazard and multi-risk decision-support tools as a part of participatory risk governance: Feedback from civil protection stakeholders. *International Journal of Disaster Risk Reduction*, 8:50–67.

Kraehnert, K., Osberghaus, D., Hott, C., Habtemariam, L. T., Wätzold, F., Hecker, L. P., e Fluhrer, S. (2021). Insurance Against Extreme Weather Events: An Overview. *Review of Economics*, 72(2):71–95.

Leblois, A., Quirion, P., e Sultan, B. (2014). Price vs. weather shock hedging for cash crops: Ex ante evaluation for cotton producers in Cameroon. *Ecological Economics*, 101:67–80.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., Clarke, M., Devereaux, P. J., Kleijnen, J., e Moher, D. (2009). The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of clinical epidemiology*, 62(10):e1–e34.

Lyubchich, V., Newlands, N. K., Ghahari, A., Mahdi, T., e Gel, Y. R. (2019). Insurance risk assessment in the face of climate change: Integrating data science and statistics. *WIREs Computational Statistics*, 11(4).

Machado, M. L., Nascimento, N., Baptista, M., Gonçalves, M., Silva, A., Lima, J. d., Dias, R., Silva, A., Machado, E., e Fernandes, W. (2005). Curvas de danos de inundação versus profundidade de submersão: desenvolvimento de metodologia. *Revista de Gestão de Água da América Latina*, 2(1):35–52.

Martínez Salgueiro, A. (2019). Weather index-based insurance as a meteorological risk management alternative in viticulture. *Wine Economics and Policy*, 8(2):114–126.

Miranda, M. J. (1991). Area-Yield Crop Insurance Reconsidered. *American Journal of Agricultural Economics*, 73(2):233–242.

Mohor, G. S. e Mendiondo, E. M. (2017). Economic indicators of hydrologic drought insurance under water demand and climate change scenarios in a Brazilian context. *Ecological Economics*, 140:66–78.

Monteleone, B., Borzí, I., Bonaccorso, B., e Martina, M. (2022). Quantifying crop vulnerability to weather-related extreme events and climate change through vulnerability curves. *Natural Hazards*, pages 1–36.

Mortensen, E. e Block, P. (2018). ENSO Index-Based Insurance for Agricultural Protection in Southern Peru. *Geosciences*, 8(2):64.

Müller, A. e Grandi, M. (2000). Weather Derivatives: A Risk Management Tool for Weather-sensitive Industries. *The Geneva Papers on Risk and Insurance - Issues and Practice*, 25(2):273–287.

Mußhoff, O., Hirschauer, N., Grüner, S., e Pielsticker, S. (2018). Bounded rationality and the adoption of weather index insurance. *Agricultural Finance Review*, 78(1):116–134.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer Series in Statistics. Springer New York, New York, NY.

Norton, M. T., Turvey, C., e Osgood, D. (2013). Quantifying spatial basis risk for weather index insurance. *The Journal of Risk Finance*.

Odening, M. e Shen, Z. (2014). Challenges of insuring weather risk in agriculture. *Agricultural Finance Review*.

Paudel, Y., Botzen, W. J., e Aerts, J. C. (2015). Influence of climate change and socio-economic development on catastrophe insurance: a case study of flood risk scenarios in the netherlands. *Regional Environmental Change*, 15(8):1717–1729.

Pislyakov, V. e Shukshina, E. (2014). Measuring excellence in Russia: Highly cited papers, leading institutions, patterns of national and international collaboration. *Journal of the Association for Information Science and Technology*, 65(11):2321–2330.

Platteau, J.-P., De Bock, O., e Gelade, W. (2017). The demand for microinsurance: A literature review. *World Development*, 94:139–156.

Porth, L., Boyd, M., e Pai, J. (2016). Reducing risk through pooling and selective reinsurance using simulated annealing: An example from crop insurance. *The Geneva Risk and Insurance Review*, 41(2):163–191.

Raucci, G. L., Lanna, R., Silveira, D., F., H., e D., C. (2019b). Development of weather derivatives: evidence from the Brazilian soybean market. *Italian Review of Agricultural Economics*, 74(2):17–28.

Riahi, K., Van Vuuren, D. P., Kriegler, E., Edmonds, J., O'neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., et al. (2017). The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: an overview. *Global environmental change*, 42:153–168.

Ricome, A., Affholder, F., Gérard, F., Muller, B., Poeydebat, C., Quirion, P., e Sall, M. (2017). Are subsidies to weather-index insurance the best use of public funds? A bio-economic farm model applied to the Senegalese groundnut basin. *Agricultural Systems*, 156:149–176.

Righetto, J., Mendiondo, E., e Righetto, A. (2007). Modelo de Seguro para Riscos Hidrológicos. *Revista Brasileira de Recursos Hídricos*, 12(2):107–113.

Rodríguez, Y. E., Pérez-Uribe, M. A., e Contreras, J. (2021). Wind put barrier options pricing based on the nordix index. *Energies*, 14(4):1177.

Roznik, M., Brock Porth, C., Porth, L., Boyd, M., e Roznik, K. (2019). Improving agricultural microinsurance by applying universal kriging and generalised additive models for interpolation of mean daily temperature. *The Geneva Papers on Risk and Insurance - Issues and Practice*, 44(3):446–480.

Sacchelli, S., Cipollaro, M., e Fabbrizzi, S. (2018). A GIS-based model for multiscale forest insurance analysis: The Italian case study. *Forest Policy and Economics*, 92:106–118.

Sant, D. T. (1980). Estimating Expected Losses in Auto Insurance. *The Journal of Risk and Insurance*, 47(1):133.

Sarris, A. (2013). Weather index insurance for agricultural development: Introduction and overview. *Agricultural Economics (United Kingdom)*, 44(4-5):381–384.

Scopus (2022). What is scopus about?

Sekhri, S., Kumar, P., Fürst, C., e Pandey, R. (2020). Mountain specific multi-hazard risk management framework (MSMRMF): Assessment and mitigation of multi-hazard and climate change risk in the Indian Himalayan Region. *Ecological Indicators*, 118:106700.

Shirsath, P., Vyas, S., Aggarwal, P., e Rao, K. N. (2019). Designing weather index insurance of crops for the increased satisfaction of farmers, industry and the government. *Climate Risk Management*, 25:100189.

Skees, J. R. (2008a). Challenges for use of index-based weather insurance in lower income countries. *Agricultural Finance Review*, 68(1):197–217.

Smith, A. B. e Matthews, J. L. (2015). Quantifying uncertainty and variable sensitivity within the US billion-dollar weather and climate disaster cost estimates. *Natural Hazards*, 77(3):1829–1851.

Štulec, I., Petljak, K., e Naletina, D. (2019a). Weather impact on retail sales: How can weather derivatives help with adverse weather deviations? *Journal of Retailing and Consumer Services*, 49:1–10.

Štulec, I., Petljak, K., e Naletina, D. (2019b). Weather impact on retail sales: How can weather derivatives help with adverse weather deviations? *Journal of Retailing and Consumer Services*, 49:1–10.

Swiss Re, S. (2021a). *World insurance: the recovery gains pace.* Swiss Re.

Swiss Re, S. (2021b). World insurance: the recovery gains pace. *Sigma*, 3.

Tilloy, A., Malamud, B. D., Winter, H., e Joly-Laugel, A. (2019). A review of quantification methodologies for multi-hazard interrelationships. *Earth-Science Reviews*, 196:102881.

Turvey, C., Shee, A., e Marr, A. (2019). Addressing fractional dimensionality in the application of weather index insurance and climate risk financing in agricultural development: A dynamic triggering approach. *Weather, Climate, and Society*, 11(4):901–915.

UNDRR (2009). 2009 unisdr terminology on disaster risk reduction.

UNDRR (2014). Sendai framework for disaster risk reduction 2015-2030.

UNEP, F. (2012). *Principles for sustainable insurance*. Geneva: UNEP FI.

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., et al. (2011). The representative concentration pathways: an overview. *Climatic change*, 109(1):5–31.

Vroege, W., Bucheli, J., Dalhaus, T., Hirschi, M., e Finger, R. (2021). Insuring crops from space: the potential of satellite-retrieved soil moisture to reduce farmers' drought risk exposure. *European Review of Agricultural Economics*, 48(2):266–314.

Wang, S. S. (2002). A Universal Framework for Pricing Financial and Insurance Risks. *ASTIN Bulletin*, 32(2):213–234.

Yoshida, K., Srisutham, M., Sritumboon, S., Suanburi, D., Janjirauttikul, N., e Suanpaga, W. (2019). Evaluation of economic damages on rice production under extreme climate and agricultural insurance for adaptation measures in Northeast Thailand. *Engineering Journal*, 23(6):451–460.

Zara, C. (2010). Weather derivatives in the wine industry. *International Journal of Wine Business Research*, 22(3):222–237.

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., van den Hurk, B., AghaKouchak, A., Jézéquel, A., Mahecha, M. D., et al. (2020). A typology of compound weather and climate events. *Nature reviews earth & environment*, 1(7):333–347.

# 3 A DATA-DRIVEN FRAMEWORK FOR ASSESSING CLIMATIC IMPACT-DRIVERS IN THE CONTEXT OF FOOD PRODUCTION

A modified version of this chapter was submitted as follows: BENSO, Marcos Roberto; SILVA, Roberto Fray; GESUALDO, Gabriela Chiquito; MARQUES, Patricia Angélica; DELBEM, Alexandre; SARAIVA, Antônio Mauro; MENDIONDO, Eduardo Mario. A data-driven framework for assessing climatic impact-drivers in the context of food security. **Natural Hazards and Earth System Sciences**

## Abstract

Understanding how physical climate conditions affect crop production requires the effort to transform climate data into relevant information for regional impact studies. Data-driven methods can bridge this gap; however, more development must be done to create interpretable models, with special emphasis in regions with a lack of data availability. In this study, we adopt the climatic impact-driver (CID) approach proposed by Working Group I (WGI) in the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC). The aim of using the CID framework is selecting the most relevant indices that drive crop yield losses and identifying important thresholds for the indices, in which if they are exceeded, the probability of impacts increases. We examine the impact of two CID types (heat and cold, and wet and dry) represented by indices of climate extremes considering the impact on different crop yield datasets, with a specific focus on maize and soybeans in the main agro-producing municipalities in Brazil. We used a random forest model in a bootstrapping experiment to select the most relevant climate indices, and then applied Shapley Additive Explanations (SHAP) with XGBoost explanatory analysis to identify the thresholds of the indices that cause impact. We found that mean precipitation is a highly relevant CID; however, there is a window in which crops are more vulnerable to precipitation deficit. For soybeans, in many regions of Brazil, precipitation below 80 mm/month in December, January and February represents an increasing risk of crop yield losses, which is the end of the growing season, and for maize a similar pattern with precipitation bellow 100 mm/month in April and May. Indices of extremes are relevant to represent crop yield variability; however, the inclusion of climate means remains highly relevant and recommended for the study of the impact of climate risk to agriculture. Our findings contribute to a growing body of knowledge that is critical for informed decision-making, policy development, and adaptive strategies in response to climate change and its impact on agriculture.

## 3.1 Introduction

Climate extremes such as heat waves, droughts, floods, and excessive precipitation play a crucial role in determining crop yield shortfalls (Vogel et al., 2019). Based on empirical evidence, models that use data from multiple weather variables are more accurate in describing the variability of crop production than models that only use precipitation (Proctor et al., 2022; Ray et al., 2015). Consequently, it is critical to assess the risk of agricultural production through extreme climate indices, which are a fundamental part of monitoring hazards that can cause impacts on food production (Das et al., 2022; Schyns et al., 2015).

Several natural hazards create impacts for society; therefore, they can be named as ”impact-drivers”. According to Ruane et al. (2022), it is impossible to understand the impact-drivers without knowing of the vulnerability and exposure of the sector. Sectoral information can determine the magnitude of the driver’s effect, which can be beneficial or detrimental. This requires a co-criation process that aims at contextualizing climate information for decision-making. This concept was implemented as the climatic impact-drivers (CID) framework and was introduced by Working Group 1 of the Intergovernmental Panel on Climate Change (IPCC) in the Sixth Assessment Report (AR6) (Ranasinghe et al., 2021; Ruane et al., 2022).

The formal definition of the CID is the physical conditions of the climate that include means, extremes, and events, and the CID is framed in terms of two important aspects of risk assessment, defining *”Indices for climatic impact-drivers”* and identifying *”Thresholds for climatic impact-drivers”* (Ruane et al., 2022). These aspects reinforce the learning that is necessary for determining numerically computable indices that utilize one or a combination of climate variables to quantify the intensity and frequency of a CID, and the values of the indices when surpassed increase the danger of losses and damage.

Although there are many indices that have been used in the literature and summarized by Ranasinghe et al. (2021) in terms of their relevance to the agricultural sector, it is still necessary to tailor the approach based on regional characteristics that bridge global information with local solutions. The CID framework is still in its early stages of development, as it is in line with the United Nations Sendai Framework for Disaster Risk Reduction (UNISDR) 2015–2030 (UNDRR, nd) and is in accordance with the definitions of the UNISDR hazard list. However, the CID framework also takes into account climate change as a significant hazard, which is not included in the UNSIDR hazard list.

However, the lack of data and the interpretability of the models are two challenges that have been shown in the literature. One of the main challenges in evaluating the impact of weather extremes on food production is the limited availability of high-quality yield and weather data at suitable temporal and spatial scales (Vogel et al., 2019).

Reanalysis datasets offer extended time series with high spatial and temporal resolutions. However, reanalysis models can produce high uncertainty when estimating poorly monitored regions (Condom et al., 2020). The weather dataset from the European Centre for Medium-Range Weather Forecasts ERA5 has been recommended for agricultural applications (Almendra-Martín et al., 2021), however, more research is required to investigate the applicability of ERA5 in studies of crop yield responses to climate variability.

The paradigm of interpretability of machine learning models is a broad topic of discussion in supervised learning (Lipton, 2018). (i) To better interpret the causal association between data, it could be achieved using the correlation method. (ii) The training data can be imperfect to represent an environment that changes over time.

One way to understand the impact of climate change on agricultural production can be through Machine Learning Algorithms (ML). Machine learning algorithms can improve our understanding of the impact of climate on crop yields (Sidhu et al., 2023). Drawing on statistical learning theory (Vapnik, 1999b), these algorithms can generalize patterns and make predictions from available data. Several authors have applied machine learning algorithms to predict crop yields (Van Klompenburg et al., 2020; Vogel et al., 2019; Sidhu et al., 2023; Han et al., 2019; Pantazi et al., 2016; Schierhorn et al., 2021; Silva Fuzzo et al., 2020). In the study of Silva Fuzzo et al. (2020), a prediction model was employed to evaluate the variability of soybean crop yields in Paraná. They used satellite-derived surface radiant temperature (Ts) and the Normalized Difference Vegetation Index (NDVI), coupled with water balance performed using satellite evapotranspiration and precipitation data, along with reanalysis. The study findings demonstrated a strong alignment between the model predictions and actual crop yields. Moreover, the potential machine learning has been demonstrated by Pantazi et al. (2016) and Han et al. (2019).

Decision tree-based algorithms such as random forest (RF) models have been used to improve understanding of the impact of weather extremes on crop yield variability (Vogel et al., 2019; Jeong et al., 2016; Schierhorn et al., 2021). RF models are a combination of tree predictors that are split recursively and used for improved predictions (Breiman, 2001) and, by their nature, can provide information on the importance of each feature in the overall performance of the model. This was used by (Vogel et al., 2019) and (Schierhorn et al., 2021) to unravel the influence of extreme temperature and precipitation from mean climate variables on predictive RF models of soybean and maize. The two articles agree that mean climate variables over growing seasons are the most relevant features for predicting crop yields, however, extreme weather indices, especially for droughts and temperature extremes, can explain from 18-43% of crop yield variability (Vogel et al., 2019).

Despite the effort to increase the performance and interpretability of the model of machine learning models, the studies cited previously relied on a rather limited selection of indices. This means that other factors that influence crop yield variability can remain hidden and the underlying mechanisms of crop yield losses due to weather extremes can be neglected.

A generalized framework for variable and feature selection has been designed to improve the performance of ML models, provide faster models, and improve understanding of the underlying processes that generated the data (Guyon e Elisseeff, 2003). The backwards recursive feature elimination (RFE) presented by Svetnik et al. (2004) uses the ability of RF algorithms to generate variable importance as a variable reduction wrapper algorithm. Several articles have shown its applicability for suitable crop selection (Wang e Li, 2023) and the use of hyperspectral imaging to monitor pasture quality (Pullanagari et al., 2018).

Although eliminating redundant features and variables can improve our understanding of the structure of the data, the problem of interpretability cannot be fully addressed. We propose the use of the model-agnostic explanation method introduced by Lundberg e Lee (2017a) called SHAP (SHapley Additive exPlanations). There are some promising studies applying SHAP to environmental data (Wikle et al., 2023; Viana et al., 2021) with implications for soil moisture and determination of evapotranspiration. The use of *post hoc* explanation algorithms for crop yields was used by Mariadass et al. (2022), however, to our knowledge, this method has not been specifically applied to predict the impacts of extreme weather on food production.

To enhance the interpretability of tree-based models that utilize climate data to predict crop yield losses, we introduce a comprehensive modeling framework. This framework incorporates filtering methods for feature selection and determination of variable importance, along with *post hoc* analysis aimed at inferring the impact of each selected variable on the prediction.

In this research, our main objective is to assess the impact of climate extremes on food production. To achieve this, we have used the climatic impact-driver (CID) framework developed by Working Group 1 (IPCC WG1) of the Intergovernmental Panel on Climate Change (IPCC). This framework allows us to characterize climate extremes by creating numerically computable indices and determining relevant thresholds. The significance of this framework lies in its ability to provide a basis for incorporating climate information into studies, decision-making processes, and policy development. By applying this framework to our research, we aim to provide valuable insights that can inform critical decisions, policies, and strategies related to food production in the face of climate extremes.

## 3.2 Methodology

In this section, we present a framework for investigating the impacts of climate extremes on crop yields. The framework aims to investigate the impacts of climate extremes on crop yields using machine learning focusing on building a reproducible workflow, selecting features, and producing explainable model outputs. A good feature of a machine learning algorithm is **relevant** to explain the target variable (in this study, crop yields); however, it should not be **redundant** with any other relevant predictor (Yu e Liu, 2003). In addition to these concepts that are widely applied in machine learning methods, we add the concepts of **explainable** and **operational** features. In machine learning, a feature is any variable that is used as an input variable for prediction; therefore, the term feature and variable will be used interchangeably in this work. Calculations were carried out in the R environment using version 4.1.2 of the software (2021-11-01) (R Core Team, 2021).

### 3.2.1 Modeling Framework

This chapter introduces a modeling framework that seeks to understand the main climatic influences on soybean and maize crop yield, as well as to determine the thresholds of these climatic influences using a fully data-driven approach (Fig. 3.1). This framework consists of three steps. The first is data filtering; in this step, we removed highly correlated features (Pearson correlation greater than 0.9). Feature selection is considered to be a pre-processing step in machine learning models. The filtering process removes redundant features, and the other aspects, that is, relevancy, explainability, and operationability, will be explained in the next steps. The second step aims at selecting the most important variables. We use the abilities of the random forest to generate the importance of variables to rank the most important variables. The third step is to define variable thresholds. To do that, we have to apply another machine model in order to explain the first one. The Shapley Additive Explanations (SHAP) explanatory analysis is an explanation algorithm proposed by Štrumbelj e Kononenko (2014) and uses game theory (Strumbelj e Kononenko, 2010) to provide an efficient explanation of the predictions made by a machine learning algorithm. The SHAP method entails the use of a second algorithm model, most commonly the eXtreme Gradient Boosting (XGBoost) models (Ribeiro et al., 2016), to explain how each variable was used to make each individual prediction of the model.

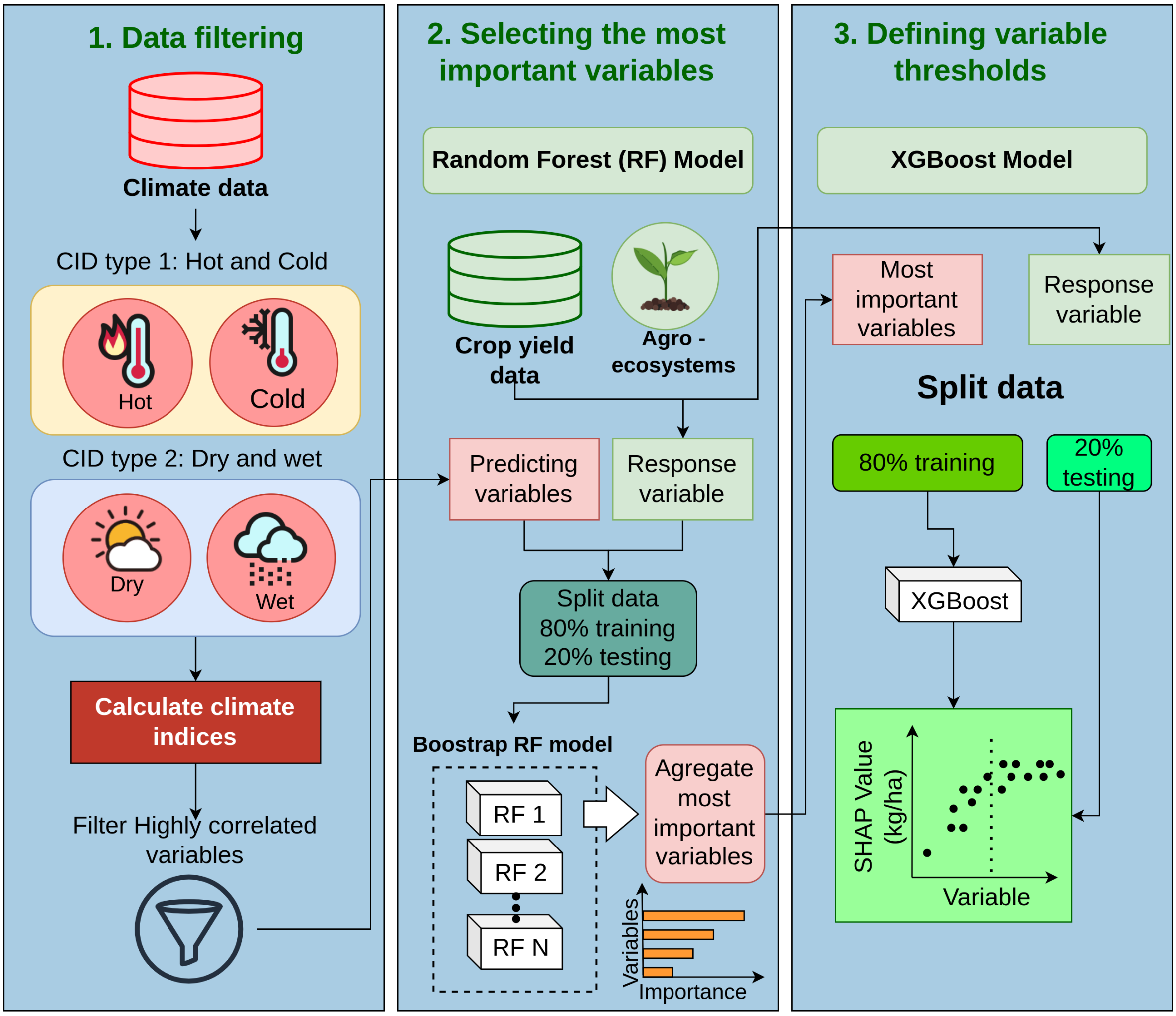


**Figure 3.1.** Flowchart illustrating the systematic methodology for analyzing the impact of climate extreme drivers (CID) on crop yield.

In the second step, we focus on identifying the most important climatic impact-drivers (CID) to assess their impact on food production. To achieve this, we used the random forest model. Different models were trained considering different combinations of input data, including precipitation means, temperature means, and combinations of means and extreme climate indices. The goal of this experiment was to identify the most important climate indices.

Random Forest (RF) algorithms are a combination of multiple classification and regression trees (CART) predictors that are sampled independently and with the same distribution (Breiman, 2001). This model has a wide range of applications and has gained a lot of attention for predicting the impacts of climate on regional crop yields (Jeong et al., 2016). Regarding the first step, redundancy is not a problem for tree-based algorithms, because algorithms are not built under the assumption of a fixed functional relationship between the predictor variables (Chan et al., 2022). For the sake of model performance, it might not be the problem; however, it can be a problem for the interpretation of machine learning models (Aydin e Iban, 2023).

RF algorithms are based on a model that creates a large number of trees through a bootstrapping process. The algorithm creates bootstraps from the original training datasets and then resamples different subsets of randomly selected features to make predictions using a wide range of decision trees or splits. Each prediction is then tested with a remaining part of the data that was not selected for the bootstrap dataset, which is also called Out-of-Bag (OOB) dataset. Decision trees that are more accurate receive a higher weight in the model and then are aggregated. However, in order to avoid highly correlated trees, a split-variable randomization is implemented, limiting the number of split variables to a value. For training and testing random forest models, we used the R-Package ranger (Wright e Ziegler, 2017). The performance of the models was measured using the coefficient of determination. To summarize the results, we calculated the performance for each state considering the median and the upper and lower 95 % confidence intervals. Medians and confidence intervals were calculated using R-package QuantileNPCI (Hutson et al., 2019), which implements Hutson's algorithm to calculate nonparametric confidence interval for quantiles (Hutson, 1999).

Building upon the results of the second step, the third step used the most relevant climate indices employing an XGBoost explanability with Shapley Additive Explanations (SHAP) explanatory analysis. This approach aimed to provide a detailed understanding of how these crucial indices were used by the model and, furthermore, attempted to identify significant thresholds for these influential climate variables. The XBoost models were implemented using the R-package xgboost (Chen et al., 2023) and the SHAP explanations were implemented using the R-package shapviz (Mayer, 2023).

The eXtreme Gradient Boosting, also called XGBoost, was created by Chen e Guestrin (2016) and is a variant of the boosting tree algorithm family. Boosting tree and random forest algorithms have the same principle of combining the outputs of individual decision trees; however, the way each individual tree is built is different between the two, and this also affects the way the outputs of each tree are combined (Sutton, 2005). Boosting algorithms build trees sequentially, rather than randomly; this means that the algorithm trains a single tree and then evaluates the quality of this tree using a penalty function, so each new tree tries to correct the mistakes of the previous one in a form of optimization problem (Breiman, 1997). This optimization problem gives a weight to each created tree and the prediction is the sum of all trees (Sutton, 2005). XGBoost is considered to provide state-of-the-art results for agricultural applications (Mariadass et al., 2022), and its complexity and set of equations can be further explored in Chen e Guestrin (2016).

The SHAP approach is based on explaining how each individual feature of the model was used to make a single prediction. The first step is to define a base prediction, which in this case is the expected value of the set $X$, which is the values of the adjusted and detrended crop yields $E[f(X)]$. Then, the XGBoost algorithm is used to make a prediction $f(x)$ for a single value of crop yield in a certain municipality and in a certain year. The difference between the expected value and the prediction is called the SHAP value and preserves the unit of crop yields; therefore, here it will be in tons per hectare. In order to identify how each feature was used to generate the SHAP value, the algorithm recursively adds each feature and tests the importance of the feature for that prediction. However, the order in which the feature is added to the model is important; that is why the principles of game theory introduced by Lloyd Stowell-Shapley were used to solve the problem of allocating the order of each feature and extracting the importance of it for the prediction. More details of the game theory used in SHAP explanations can be learned in Strumbelj e Kononenko (2010).

In Fig. 3.2, we provide a graphical illustration of the black box approach in which the XGBoost algorithm uses three features, Temperature (Temp), Precipitation (Prec), and Standardised Precipitation Evapotranspiration Index (SPEI) to make a prediction. With SHAP explanations, the difference between base prediction and the model output, also called SHAP Value, is explained. As we perform the SHAP explanations for each individual prediction, we can generate a partial dependence plot, which is the relationship between the feature value and the contribution to the SHAP value. From this approach we can unveil what values of the feature are critical for crop yield losses and then establish thresholds and contributing for the CID framework. The SHAP value is obtained through iteration of all possible subsets of features and provides the marginal contribution of each feature to the prediction. This implies that the evaluation of the interaction between features is also possible. Since the interaction among features is also evaluated, a second analysis is performed, which uses Gaussian copulas to evaluate the combined effect of the variables.

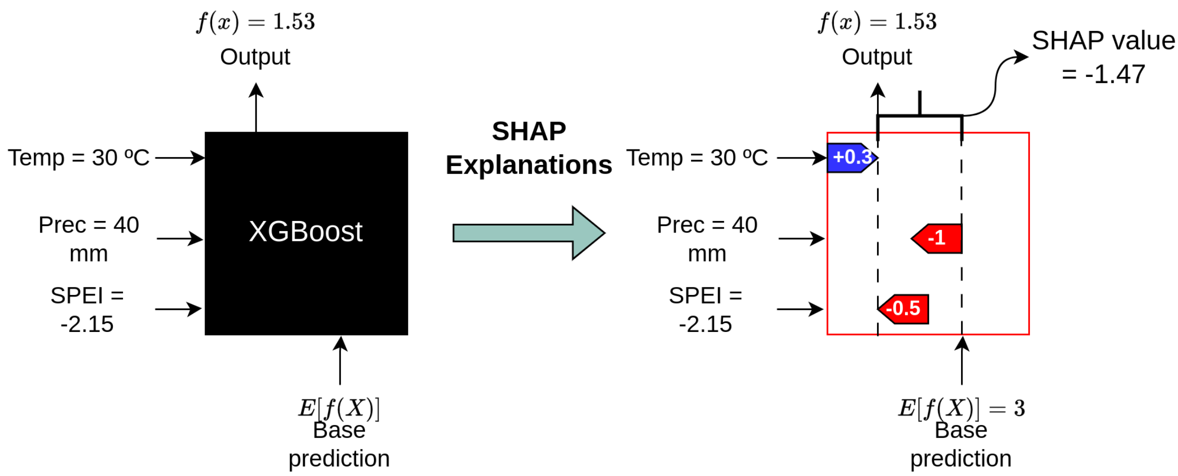


**Figure 3.2.** Demonstration of SHAP Explanations. Adapted from Lundberg e et al. (2023)

### 3.2.2 Study Area

The richness of plant biodiversity offers a wide range of possibilities for human nutrition. Currently, the existence of around 374,000 known and described plant species is accepted (Christenhusz e Byng, 2016). According to Rapoport e Drausal (2013), there are more than 16,000 species that are considered edible. However, the diets of most of the human population are based on one or more of the main staple crops, such as soybeans (Hartman et al., 2011) and maize (Shiferaw et al., 2011). Although we recognize that food production and food security require the study of a broader range of plants and animal products, in this work, we focus on the production of soybean and maize second cycle. This is because these two products represent the main staple food products in terms of area and quantity. The second reason for this choice is because the cycle of these two cultures is well defined across the main producing regions in Brazil.

Brazil is a major producer of agricultural goods, as reported by the Food and Agriculture Organization (FAO) (FAO, nd). This country is responsible for more than 10% of the world's maize and more than 30% of the global soybean production. Brazil is one of the four leading agricultural producers in the world, along with China, India and the United States, with a cultivated area of soybean and maize of 58 million hectares. In Fig. 3.3 we show the delimitation of the study area. The map shows 452 selected municipalities that encompass the states of Rio Grande do Sul (RS), Santa Catarina (SC), Paraná (PR), São Paulo (SP), Mato Grosso do Sul (MS), Minas Gerais (MG). The selection criteria will be explained in the sub-section Crop yield data. The map shows the percentage of cropland derived from the work of Potapov et al. (2022).

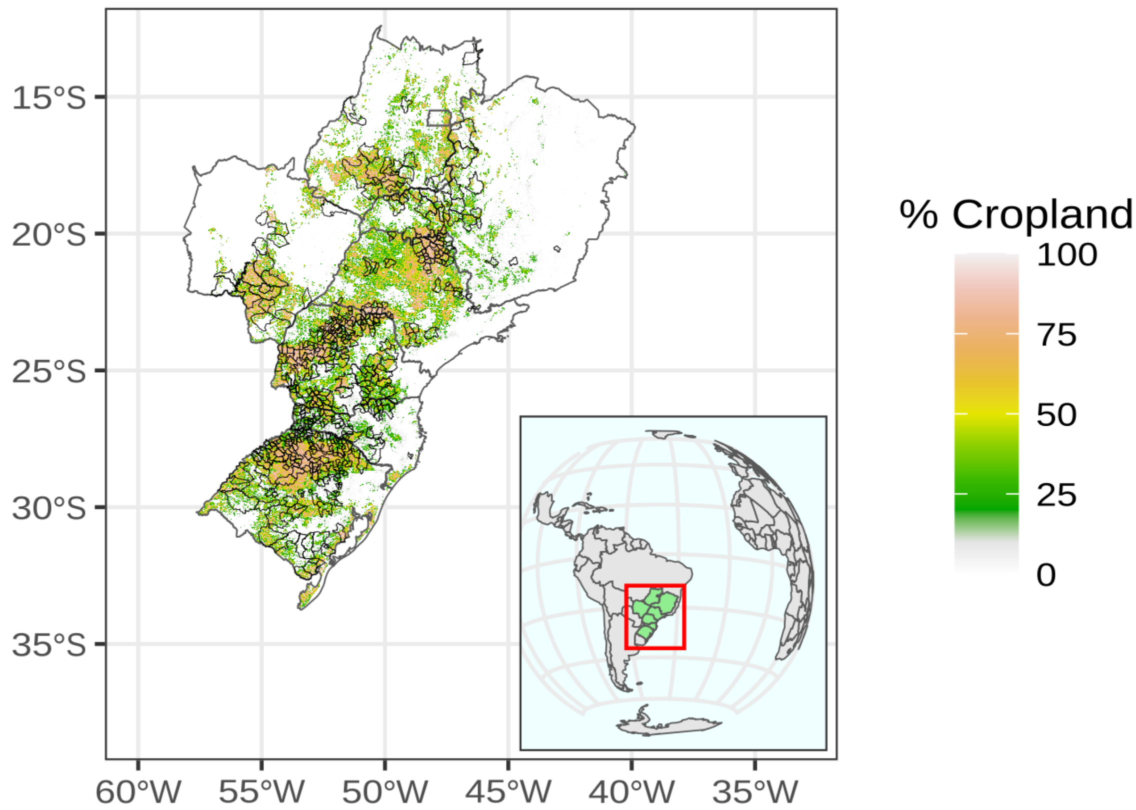


**Figure 3.3.** Location of selected study municipalities with respect with observed cropland extent between from 2003 to 2019. The map show the percentage of cropland within the pixel

The growing season in the study area was defined using the global crop calendar for soybeans and maize using data from Sacks et al. (2010). We consider that the planting dates follow a normal distribution, with the mean date being the most probable date for farmers, and the maximum and minimum dates being considered twice the standard deviation. For soybeans, sowing dates start in the middle of Austral spring in October, peak in November, and end in December. The harvest begins in the late summers and extends to fall, ranging from February to March. Since the second cycle of maize peaks in February, we consider that the end of soybean is in February. For the second cycle of maize, planting begins at the end of January, after soybean harvest, peaks in February, and ends in the beginning of April. The harvest starts in June, peaks in August, and ends in October.

### 3.2.3 Data collection and Processing

In this section, we present the description of datasets used to analyze the impact of climate variables on soybean and maize crops in Brazil. We used two criteria to select a dataset, the first one is that the data must comply with FAIR principles, i.e., data must have findability, accessibility, interoperability, and reusability; and the second is that climate data must be updated frequently, ideally, with minimum update frequency daily. We used three different datasets, Paraná state statistical yearbooks (Parana, 2021), Municipal Agricultural Production Survey by the Brazilian Institute of Geography and Statistics (IBGE) (de Geografia e Estatística, 2022) and Global dataset of historical yields (GDHY) (Iizumi e Sakai, 2020). For climate analysis, we used data from the fifth generation European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 Land reanalysis dataset (Muñoz-Sabater et al., 2021). We summarized the main characteristics of each dataset used in Table 3.1.

**Table 3.1.** Description of datasets used in the methods

| Dataset | Variable | Spatial resolution | Temporal resolution | Time frame |
|---|---|---|---|---|
| Crop yields IBGE | Soybean | Municipality level | Annual | 1974-2022 |
| Crop yields IBGE | Maize second cycle | Municipality level | Crop cycle | 2003-2022 |
| Crop yields Paraná | Soybean and Maize | Municipality level | Crop cycle | 1997-2021 |
| Crop yields GDHY | Soybean and Maize | 0.5 degree cell | Crop cycle | 1981-2016 |
| ERA5 Land | Precipiation, temperature | 0.1 degree cell | Daily | 1950-Current |
| SoilGrids 2.0 | Clay, Silt and Sand | 250 m cell | - | - |

#### 3.2.3.1 Crop yield data

**Data from Brazilian Institute of Geography and Statistics**

Crop yield data is a key component in understanding the impacts of climate on food production. Crop yields are generally made available at the municipality level. In Brazil, the Brazilian Institute of Geography and Statistics (IBGE) and state agencies collect agricultural data using surveys, interviews, and expert elicitation to create an annual database for more than 40 crops at the municipal level (de Geografia e Estatística, 2022). The raw data collected from IBGE can be assessed in the Multidimensional Statistics Database (BME). One of the main challenges of this dataset is that the data represent an annual statistics and do not represent different cycles. In the study area, the double cropping system is widely adopted, therefore representing a potential bottleneck. However, IBGE has started to collect maize in the first and second cycle since 2003. This matches the period of time where maize second cycle is intensified in Brazil. The selection resulted in 452 municipalities for soybeans that comprise the states of Rio Grande do Sul (RS), Santa Catarina (SC), Paraná (PR), São Paulo (SP), Mato Grosso do Sul (MS), Minas Gerais (MG), and Goiás (GO) and 216 municipalities for maize second cycle for Paraná (PR), São Paulo (SP), Mato Grosso do Sul (MS), Minas Gerais (MG), and Goiás (GO).

**Data from Parana state statistical yearbooks**

The Departament of Rural Economy (Deral) of the Paraná state, Brazil, is also responsible for collecting crop data at the municipal level. The method of collecting and processing of data is similar to what is done by IBGE, therefore, a high level of redundancy is expected from these two datasets. This redundancy is important to validate data and remove outliers that might reduce the quality of a model. The same number of municipalities selected using IBGE data was used in Deral data. This dataset is derived from the gross value of production divided by the harvested area, which is derived from the price and quantity of production of 30 crops. Therefore, we obtained the crop yields in tons per hectare.

**Global dataset of historical yields (GDHY)**

The Global Dataset of Historical Yields is a global annual time series of 0.5 º grid-cell estimates for maize, rice, wheat, and soybean from 1981 to 2016. For each grid cell, crop yields are estimated in ton/ha based on Food and Agriculture Organization (FAO) country-level yield statistics and then corrected using the remote-sensed leaf area index (LAI), the fraction of photosynthetically active radiation (FPAR), and crop-specific radiation use efficiency derived from reanalysis. Crop areas and crop calendars were derived from Monfreda et al. (2008) and Sacks et al. (2010). More details on the dataset are described in Iizumi e Sakai (2020) and Iizumi et al. (2014). The dataset was aggregated to the municipal level using zonal statistics in the terra package (Hijmans, 2023a) in R Studio.

In the literature, three major problems have been reported in regard to the quality of crop yield data for risk analysis, namely the presence of outliers, technological trends, and heterokedasticity. Removal of outliers is a complex problem since we are working with extreme events. The definition of an outlier must be carefully taken to eliminate useful data. According to Ozaki et al. (2008), soybean and maize crop yield data tend to correlate, considering drought years, present correlation within approximately 150 km, and, within this range, the assumption of normality can be supported. Therefore, for the sake of simplification, we assumed that crop yields within the immediate IBGE region are highly correlated, that is, $R^2 > 0.6$ and p-value $<$ 0.05.

Changes in technology in seed production, fertilizers and land management impact on crop yields (Liu e Ker, 2020). This effect is well documented in the agronomic literature and not only increases averages, but also leads to changes in non-constant variance, i.e., heteroskedasticity (Tolhurst e Ker, 2015; Harri et al., 2011). The effect of timely technological adoption and improvements on crop yields was treated in a two-step process as proposed by Zhu et al. (2011), first by removing trends and then testing and adjusting the heteroskedasticity of the residuals.

In the first step, we tested the presence of monotonic trends using the Mann-Kendall test (Mann, 1945) with the Kendall R Package (McLeod, 2022). If the Mann-Kendall indicate a p value lower than 0.05, the null hypothesis is accepted and the crop yield series is considered to have a monotonic trend that must be corrected. We choose the Local Polynomial Regression Fitting (LOESS) (Cleveland et al., 2017) to model crop yields $y_t$ for the year $t$. The residuals $\epsilon_t$ are considered to be the detreded crop yields.

We tested heteroskedasticity in the data using the Pagan-Breusch test (Breusch e Pagan, 1979). If the p-value of the test is lower than 0.05, then the null hypothesis is rejected, and the yield series is considered to be heteroskedastic. In this case, we compute the normalized residuals $y_t^n$ for two cases. The first case is when the errors are proportional to the yield level, then we obtain the $y_t^n$ considering the proportional error $\varepsilon_t$, which is calculated by dividing the error term from the LOESS prediction function $\epsilon_t$ by the yield predicted by the function. Then the proportional errors are multiplied by the yields observed in 2021. With this procedure, the past yields are expressed in terms of the 2010 technology. Otherwise, when the residuals are not proportional to the yield levels, $y_t^n$ is calculated by adding 2010 yields to the residuals of the LOESS prediction function.

After obtaining a consistent time series that is corrected for outliers, trend, and heteroskedasticity, the next step is to adjust for a distribution function. Statistical modeling of crop yields is relevant for risk management, because we do not have enough time series to empirically evaluate risk (Liu e Ker, 2020). Several parametric and nonparametric distribution functions have been proposed to model crop yields (Ozaki et al., 2008), we selected the gamma distribution because it allows one to simulate positive and negative skewness, flexibility to assume different shapes according to the parameters c$(\alpha, \beta)$, the shape and rate parameters, respectively. For the gamma distribution, when $\alpha$ is greater than 1, the distribution is skewed to the right, that is, skewed towards higher crop yields, otherwise, the distribution is skewed towards lower yields.

#### 3.2.3.2 ERA5 Land reanalysis dataset

Weather data was sourced from the fifth generation European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 Land Reanalysis dataset (Muñoz-Sabater et al., 2021). The data set has a 0.1 x 0.1 lon lat-lon grid and was aggregated to the municipal area to match the spatial discretization of SBY. The data collection spanned 1980 to 2023, with daily observations. The climate variables included precipitation, maximum temperature, and minimum temperature. We used different climate indices to evaluate multi-hazard risks. Since mean climate conditions of precipitation and temperature have been shown to be the most relevant Moriondo et al. (2011), we considered monthly precipitation, maximum and minimum temperatures, as well as total precipitation and mean temperature over growing seasons.

#### 3.2.3.3 Indices for Climatic Impact-Drivers

The WGI of the IPCC has presented the climatic impact-drivers (CID) as a new approach to assessing climate data to analyze their effects on society. The CIDs must be represented by numerically computable indices and are categorized into several types. In this thesis, we considered wet and dry and hot and cold CIDs. For computing indices, first we considered the indices indicated by the Expert Team on Climate Change Detection and Indices (ETCCDI), which is supported by the World Meteorological Organization (WMO) Commission for Climatology, the Joint Commission for Oceanography and Marine Meterology (JCOMM), and the Research Program on Climate Variability and Predictability (CLIVAR).

We also considered two drought-related indices, the Standardized Precipitation Index (SPI) (McKee et al., 1993; McKee, 1995) and the Standardized Precipitation and Evapotranspiration Index (SPEI) (Vicente-Serrano et al., 2010). The SPI is based on the probability of monthly precipitation on different time scales, and it is recommended to be calculated with a time series of at least 30 years. The monthly time series must be fitted to a cumulative distribution function (CDF); here we adopted the gamma distribution, and then the data is transformed to the standard normal distribution to calculate the SPI, which is a value standardized in terms subtracting the transformed precipitation by mean value and dividing by the standard deviation. The SPI can be calculated using different time scales that represent previous meteorological conditions typically ranging from 1 to 48 months. For agricultural applications, 3-month SPI has been shown to be the most suitable Kim et al. (2019), however, Lam et al. (2022) has shown a counter example in Kenya showing that longer time scales could also be applied for agriculture.

The SPEI incorporates temperature in the calculation of SPI. A new step is added to the procedure consisting of calculating the monthly potential evapotranspiration (PET) and calculating the SPI using the same procedure previously described with the value of monthly precipitation minus monthly PET. PET was calculated using the Hargreaves method, which is calculated using maximum and minimum temperature and extraterrestrial radiation (RA) (Droogers e Allen, 2002). Evapotranspiration was estimated using the R-package Evapotranspiration (Guo et al., 2022) and the SPI and SPEI indices were calculated using the R-package SPEI (Beguería e Vicente-Serrano, 2023).

The primary motivation for using distinct indices derived from the same fundamental data is to identify which features of the extremes are the most significant. All indices were calculated in a monthly time step. Is it the magnitude of an extreme, the length of time it lasts, or values that are either above or below a certain threshold? Researchers such as Vogel et al. (2019) have demonstrated the importance of extreme events in understanding the variability of crop yields. We summarize all the indices according to CID type and category in Table 3.2.

**Table 3.2.** Description of the Climatic Impact-Drivers (CID) considered in this study and their respective Indices

| CID Type | CID Category | CID Index Abbreviation | CID Index Description |
|---|---|---|---|
| Heat and Cold | Mean air temperature | Temp | Monthly temperature mean |
| | | DTR | Daily temperature range: Monthly mean difference between maximum and minimum daily temperature |
| | Extreme heat | TX90p | Monthly percentage of days when maximum daily temperature is higher than the 90th percentile |
| | | TN90p | Monthly percentage of days when minimum daily temperature is higher than the 90th percentile |
| | | SU | Number of summer days: monthly number of days when maximum daily temperature is higher than 25 º C |
| | | TR | Number of tropical nights: monthly number of days when minimum daily temperature is higher than 20 º C |
| | | TXX | Monthly maximum value of daily maximum temperature |
| | | TXN | Monthly maximum value of daily minimum temperature |
| | Cold spell | TX10p | Monthly percentage of days when maximum daily temperature is lower than the 10th percentile |
| | | TN10p | Monthly percentage of days when minimum daily temperature is lower than the 10th percentile |
| | | TNN | Monthly minimum value of daily minimum temperature |
| | | TXN | Monthly minimum value of daily maximum temperature |
| Wet and dry | Mean precipitation | Prctot | Monthly precipitation sum |
| | Heavy precipitation | R10mm, R20mm | Monthly count of days when daily precipitation is higher than 10 and 20 mm. |
| | | Rx1day, Rx5day | Monthly maximum 1-day and 5-day precipitation |
| | Agricultural and ecological drought | SPEI 3, 6 | Standardised Precipitation and Evapotranspiration Index for 3 and 6 moths accumulations |
| | | SPI 3, 6 | Standardised Precipitation Index for 3 and 6 moths accumulations |

#### 3.2.3.4 Soil data

The soil property data used in this study were sourced from SoilGrids 2.0, a globally recognized resource that offers standardized soil profile data Poggio et al. (2021). SoilGrids offers soil data, including the percentage of clay, silt, and sand, at a resolution of 250 meters, which was used for our analysis at a depth of 30 centimeters. The content of each soil component was combined at the municipal level to correspond with the crop yield data.

## 3.3 Results and Discussion

### 3.3.1 Spatiotemporal analysis of crop yield data

Our findings show that the hypothesis of high correlation of crop yields within the studied regions was confirmed for the data from soybean (Fig. 3.4) and maize (Fig. 3.5) indicating that this strategy is appropriate. Since soybeans have a longer record, the correlations were more stable in the immediate regions and tended to have higher values.

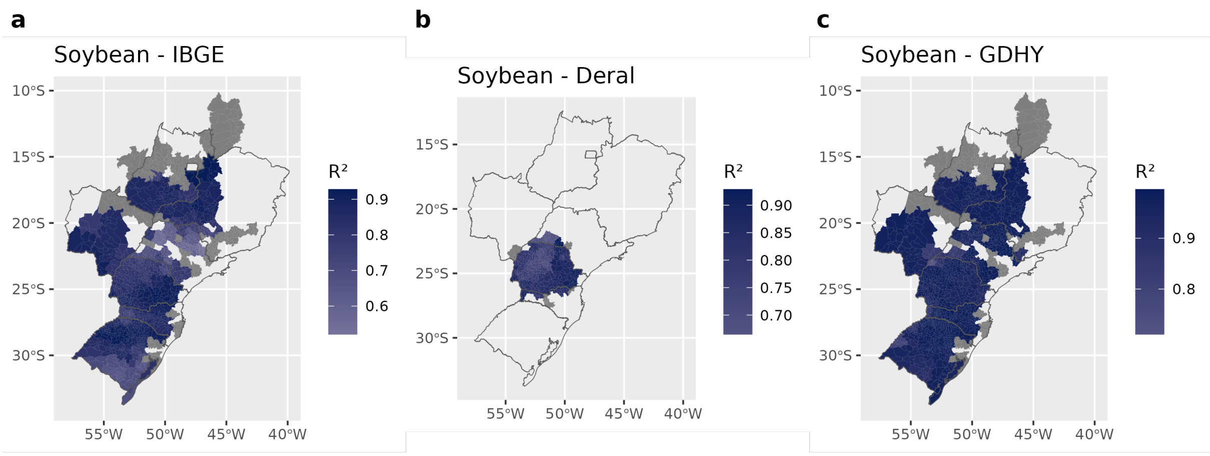


**Figure 3.4.** Spatial correlation of municipal soybean crop yields for each dataset, IBGE, Deral and GDHY

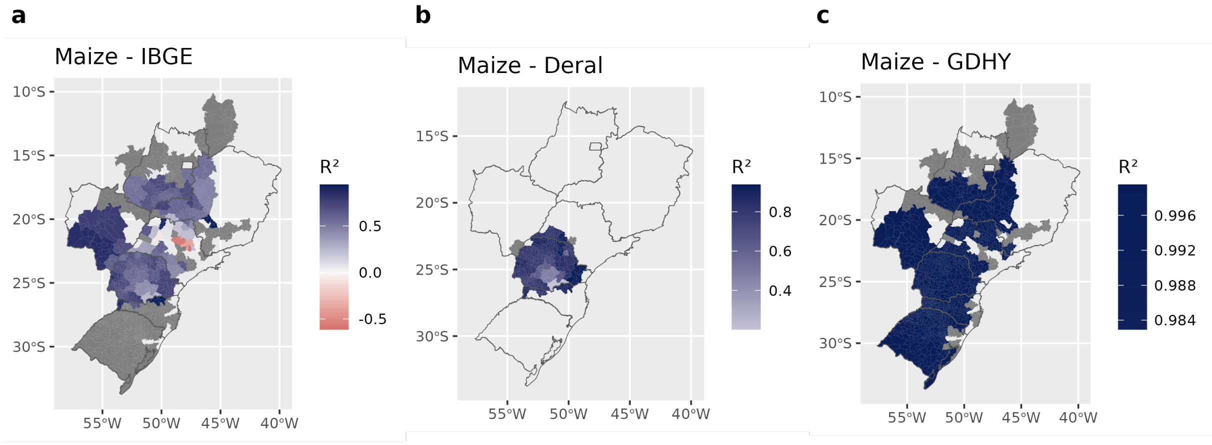


**Figure 3.5.** Spatial correlation of municipal maize crop yields for earth dataset, IBGE, Deral and GDHY. They grey areas represent no data.

The comparison of the datasets used in this study is important to evaluate the reliability of the data. High-quality crop yield data improves calibration of crop growth models (Rosenzweig et al., 2014), however, they have a broader application in geosciences. Crop yield data is used to parameterize watershed hydrological models, especially in agricultural catchments, and improve soil moisture simulation (Sinnathamby et al., 2017). For the management of water resources, the use of higher quality crop yield data has improved global knowledge on the water-food-energy nexus (Ai e Hanasaki, 2023; Wang et al., 2023).

We compared crop yields at the municipal level in Brazil. As observed in Figures 3.6 and 3.7, the IBGE and Paraná Deral data for soybeans and maize are highly correlated; however, outliers were detected in both datasets. The outlier removal process improved the agreement between the two datasets, suggesting that the elimination of data improved the quality of the dataset. Since Deral is only available in Paraná, for the other states of Brazil, only GDHY and IBGE were compared. The global dataset of historical yields aggregated at the municipal level has a weak association with the other datasets. This result confirms what was reported by Iizumi et al. (2014). The GDHY is based on satellite data collected from a fixed cropland map. In many regions of Brazil, as shown in Fig. 3.3, there is a noticeable increase in croplands, and this can influence the estimation of GHDY. In addition, the cropland areas within each municipality can vary from year to year.

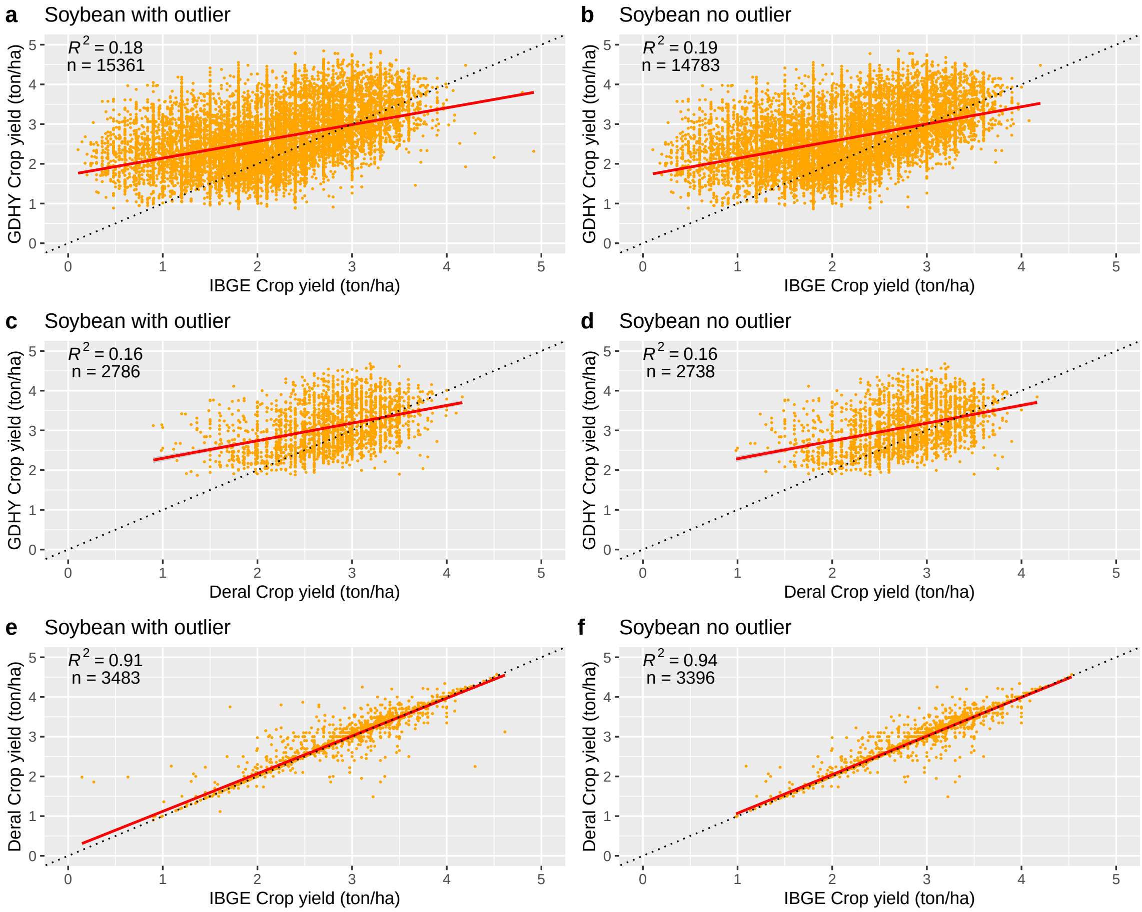


**Figure 3.6.** Correlation analysis of three different datasets for soybean crop yields: (Global dataset of historical yields ($yield_{gdhy}$), Brazilian Institute for Geography and Statistics ($yield_{IBGE}$), and Paraná Department of Rural Economy ($yield_{PR}$)

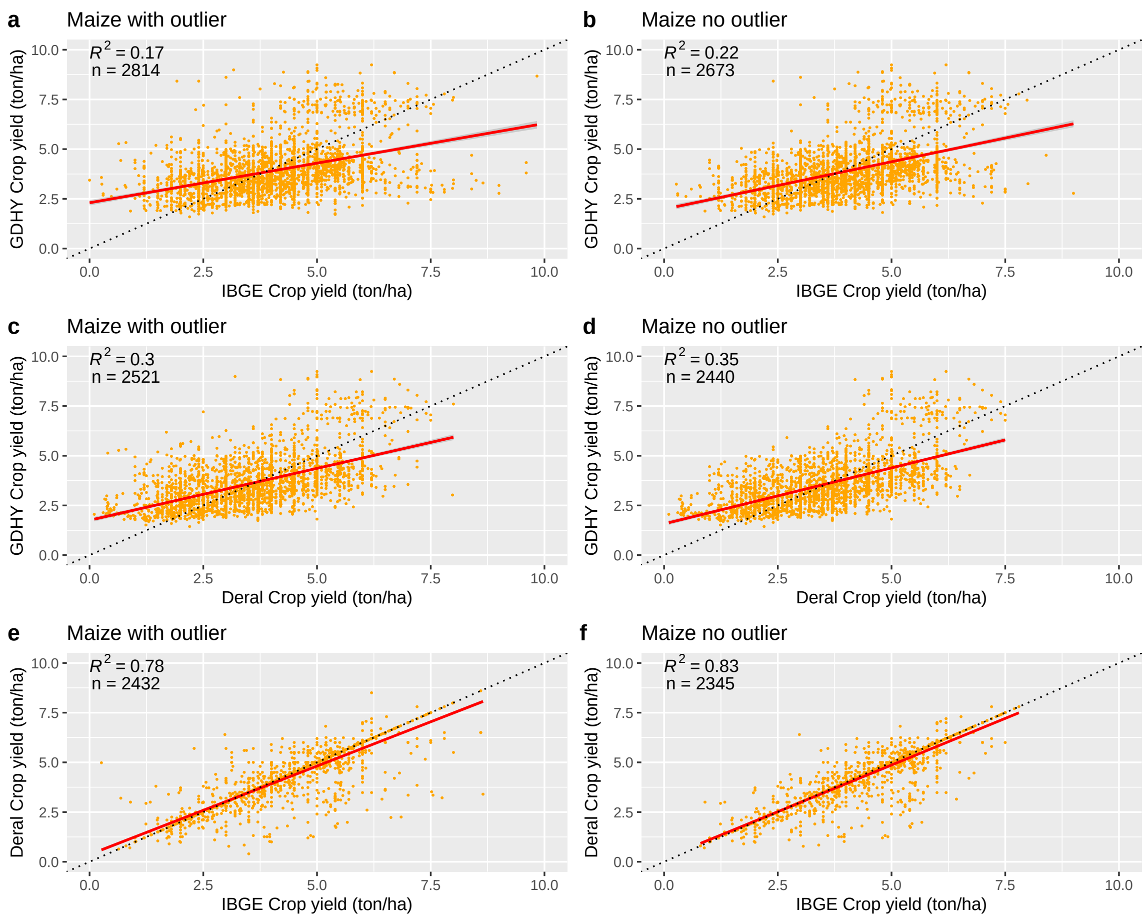


**Figure 3.7.** Correlation analysis of three different datasets for maize second cycle crop yields: (Global dataset of historical yields ($yield_{gdhy}$), Brazilian Institute for Geography and Statistics ($yield_{IBGE}$), and Paraná Department of Rural Economy ($yield_{PR}$)

In order to evaluate risk, testing and removing trend is a fundamental step to remove the effects of technology advances on the data (Harri et al., 2011). Mann-Kendall trend analysis of soybean and maize yields for Brazilian municipalities considering three datasets unveiled a consistent pattern of trends. Data for all municipalities included in our study presented significant positive monotic trends considering p values less than 0.05. For maize, on the other hand, significant positive monotonic trends were observed in most municipalities, but were not universally present in the states of MS, GO, SP, and MG in the IBGE dataset. The lack of trends can be attributed to a limitation of the dataset, particularly related to the lack of long-term data for the second maize cycle.

Since most of the data presented positive trends, we applied a LOESS model for all municipalities. The residuals of the LOESS models were then tested for heteroskedasticity. Other studies evaluated the presence of heteroskedasticity in crop yield data, Vicente (2004) was tested using the Brazilian agricultural census 1995/1996, Ozaki et al. (2008) for soybean, maize, and wheat in Paraná, and RODRIGUES et al. (2013) in farm-level studies in São Paulo. The presence of heteroskedasticity represents systematic changes in crop yield data Yang et al. (1992).

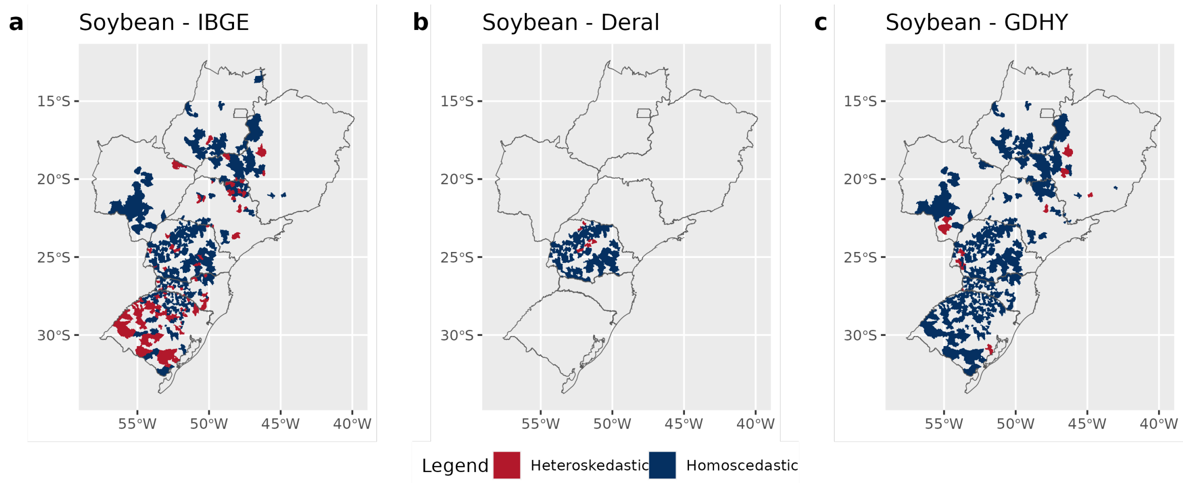


**Figure 3.8.** Heteroskedasticity Test Results for soybean crop yields in Brazilian Municipalities

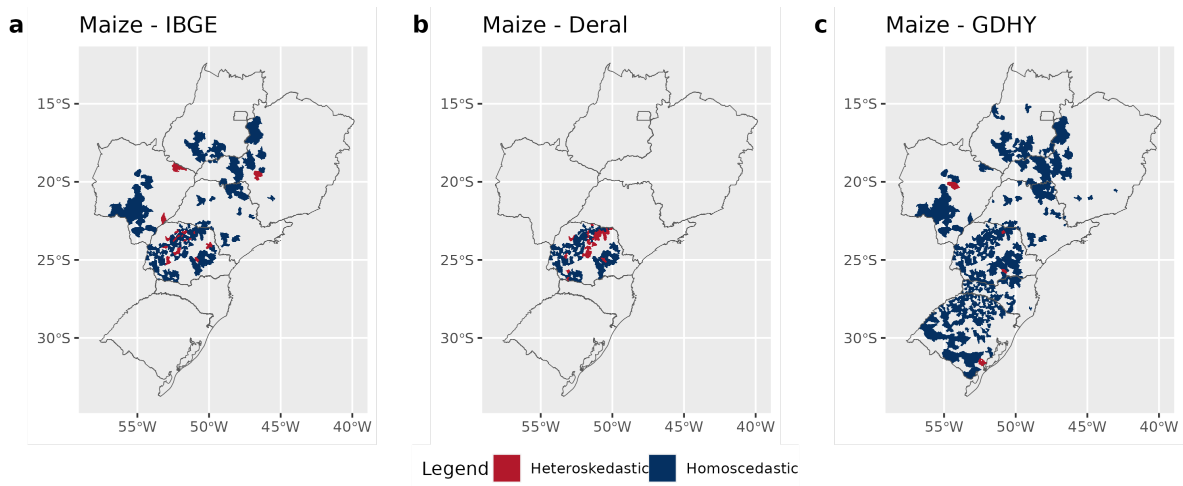


**Figure 3.9.** Heteroskedasticity Test Results for maize crop yields in Brazilian Municipalities

### 3.3.2 Identifying Key Climate Impact-Drivers

In this study, we examine different indices that assess mean precipitation, mean temperature, and extremes. The main purpose of using separate indices based on the same underlying data was to determine the most important characteristics of these extreme events. All indices were calculated monthly. We initially tested the machine learning technique using various inputs to demonstrate its ability to illustrate the variability of crop yield in the states examined. In Fig. 3.10, we demonstrate the performance of the random forest model considering different input variables and datasets. Taking the coefficient of determination, the climate variables explained the variability of soybean crop yields, on average, from 30 to 40% of IBGE, from 25 to 45% for GDHY and from 30 to 35% for Deral. For maize, the climate was explained from 30 to 40% for IBGE and Deral and from 20 to 45% for GDHY.

The coefficient of determination quantified the proportion of the variance in the crop yield data (dependent or target variable) that can be explained by the random forest model. The results are consistent with the values found in other similar studies, Ray et al. (2015) used municipal level data to quantify the impact of climate variability on yields using regression models and determined that in Brazil, climate variability explains 26–34% of soybean yields and 41% of maize yields. Vogel et al. (2019) used the same dataset as Ray et al. (2015) considering South America and applied a random forest model defining the values of 28 % for soybeans and 25% for maize. It is important to note that none of these studies separated maize first and second season.

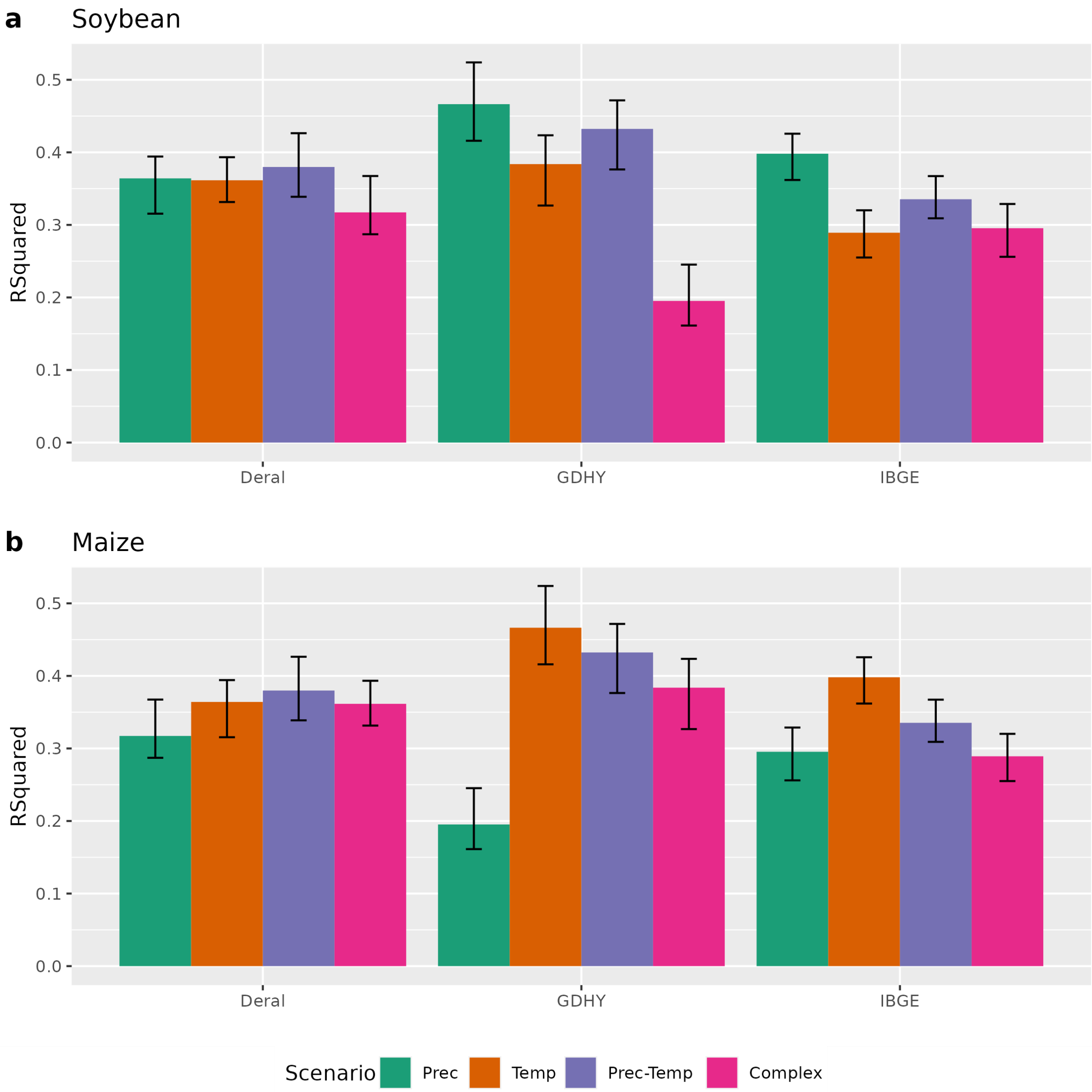


**Figure 3.10.** Performance evaluation of regression model for (a) Soybean and (b) Maize. Considering data from Department of Rural Economics (Deral) with data only for Paraná, Global Dataset on Historical Yields (GDHY) and Brazilian Institute of Geography and Statistics (IBGE) for the major historical agricultural producing municipalities

In general, the model performance was higher for GDHY than for the other datasets; for soybean in the southern states (RS, SC, and PR). The models based on precipitation means and the combination with temperature and extremes explain more the variability of crop yield than only the temperature means and this was observed in all three datasets. For maize second season, for MS, MG and GO, the models that combine mean temperature to mean precipitation and extremes explained more the variability of crop yields than only temperature, and this was observed in the GDHY and IBGE dataset. The IBGE database showed that the maize model was much more effective in São Paulo than in any other state.

The random forest models used in this study were helpful in obtaining the most relevant variables and these variables were classified into CID types, i.e., wet and dry, and hot and cold and subcategories, as described in Table 3.2. The model demonstrates the importance of climate variables in explaining the variability in crop yields and allowed us to determine the key climatic impact-drivers considering each state and dataset. In the following sections, we summarize the key CID for soybean and maize. Figures demonstrating the most important variables are in the Appendix for soybean Fig. A.4 and for maize Fig. A.5.

#### 3.3.2.1 Wet and dry

Changes in mean precipitation pose a threat to agricultural production. Precipitation deficit leads to reduced available soil moisture, affecting plant development and reducing crop yields, and it is considered the most important environmental factor that reduces crop yields (Bray, 2007). In our analysis, mean precipitation was one of the most important climatic impact-driver for soybean crop yields during the months of January and February for RS, SC, PR, MS and also in December in PR. The state of Rio Grande do Sul has historically been affected by El Niño Southern Oscillation (ENSO), with a stronger influence from November to May caused by the La Niña phenomenon (Gelcer et al., 2013), which is responsible for droughts and the impact on soybean crop yields. The model did not indicate that mean precipitation was the most important for the state of SP. For maize production, mean precipitations during April and May were considered important for PR, MS, MG, and GO, also with the exception of SP. Agricultural systems require minimum rainfall, or they rely on irrigation. In Brazil, the states of SP, MS, MG and GO have a well-defined difference between wet and dry seasons, usually the wet season starting in October and ending in May, and the Soybean-Maize double cropping system depends on the length of the wet season in the aforementioned states.

Agricultural and ecological drought indices are directly related to a precipitation deficit and excessive temperature (Lesk et al., 2022; Sarhadi et al., 2018; Lesk et al., 2021), which affects the ability of plants to grow and reduce plant transpiration. The duration and timing of droughts play a significant role, in general, we observed that droughts occurring in January and February were the most important for soybeans. The droughts that occurred in February and March also affected the second-cycle crop yields of corn. Droughts that occurred at the end of the maize growing season also affected crop yields.

In this study, we considered climate extreme indices on different time scales and, as we added the temporal dimension in the analysis, we revealed that a 3-month SPEI in October in the state of Rio Grande do Sul (RS) was selected on the list of most relevant variables. This indicates that pre-sowing meteorological factors that can reduce soil moisture conditions also influence crop yields. This result corroborates Santini et al. (2022), which revealed that drought analysis should not neglect antecedent conditions, since it influences factors such as soil workability and crop development.

#### 3.3.2.2 Hot and cold

The mean air temperature influences many aspects of crop cultivation. In RS and SC, soybean growing season starts when the mean temperatures are higher than the minimum temperature thresholds for soybeans (Battisti e Sentelhas, 2014). As the temperature increases, the development (phenology) of the plant is affected and increased thermal stress is expected Lesk et al. (2021). With the exception of RS, mean temperatures during all soybean growing seasons were considered important variables. The same behavior was observed for maize, during all growing seasons mean temperatures were considered important.

The exposure to temperatures above a certain limit or threshold can lead to lower yields. The value of these thresholds depends on the crop specie and farm management. For soybeans, extreme temperature indices affected crop yields throughout the growing season, especially in January and February.

### 3.3.3 Determining Thresholds and Significance

With the results of the selection of key climatic impact-divers, we improved the understanding of the impacts on climate variables that significantly influence crop yield losses considering different types of indices and critical periods. The insights obtained from the combination of random forest model applied to different datasets facilitate a robust understanding of climate-crop interactions and make it possible to compare what results the datasets have in common; therefore, increasing the reliability of the results. However, the random forest model did not provide information on the values of each variable that are important and can help us define the threshold values of these indicators that are associated with an increased risk of crop yield losses.

To improve our understanding of how each climate extreme indicator was used to predict, we used SHapley Additive Explanations (SHAP). This technique allowed us to extract insights by coupling the results of a Random Forest model with those of an XGBoost model, thus providing a comprehensive perspective on the drivers of crop yield fluctuations. The results of SHAP-derived explanations revealed a clear pattern with respect to the most influential variable that affects soybean yields. We highlight important loss events by evaluating the prediction of the model for a particular city in a given year.

In 2019, an important widespread drought event was observed in Brazil and was considered a mega-drought that affected many regions of Brazil, especially the Paraná river basin Marengo et al. (2021). We highlight the model explanation for the state of Paraná considering two important agro-producing municipalities, namely Borrazópolis and Uniflor, as shown in Fig. 3.11. The two datasets presented an agreement in terms of crop yields below the expected value $E[X]$, however, they varied in terms of the magnitude of these losses. For the municipality of Borrazópolis, the Deral yields were lower than IBGE, 1.53 and 3.27, respectively. The predictions made by the two models were similar, and the main variables also performed similarly. High temperatures represented by the maximum value of the daily minimum temperature in the month of February were the main driver of losses combined with the 3-month SPEI in february. For Deral, accumulated precipitation in December was also a major driver of losses and for IBGE 3-month SPEI in January. The other variables represented a negligible influence on crop yield losses. For the city of Uniflor, actual yields of Deral and IBGE were similar, 2.59 and 2.85, respectively. The main influences on crop yield losses were precipitation in December (prctot_Dec) and high temperatures (tnx_Feb). The 3-month SPEI values in January and February can be considered redundant. Standardized indices are referent previous conditions, therefore, the values overlap in two months (December and January).

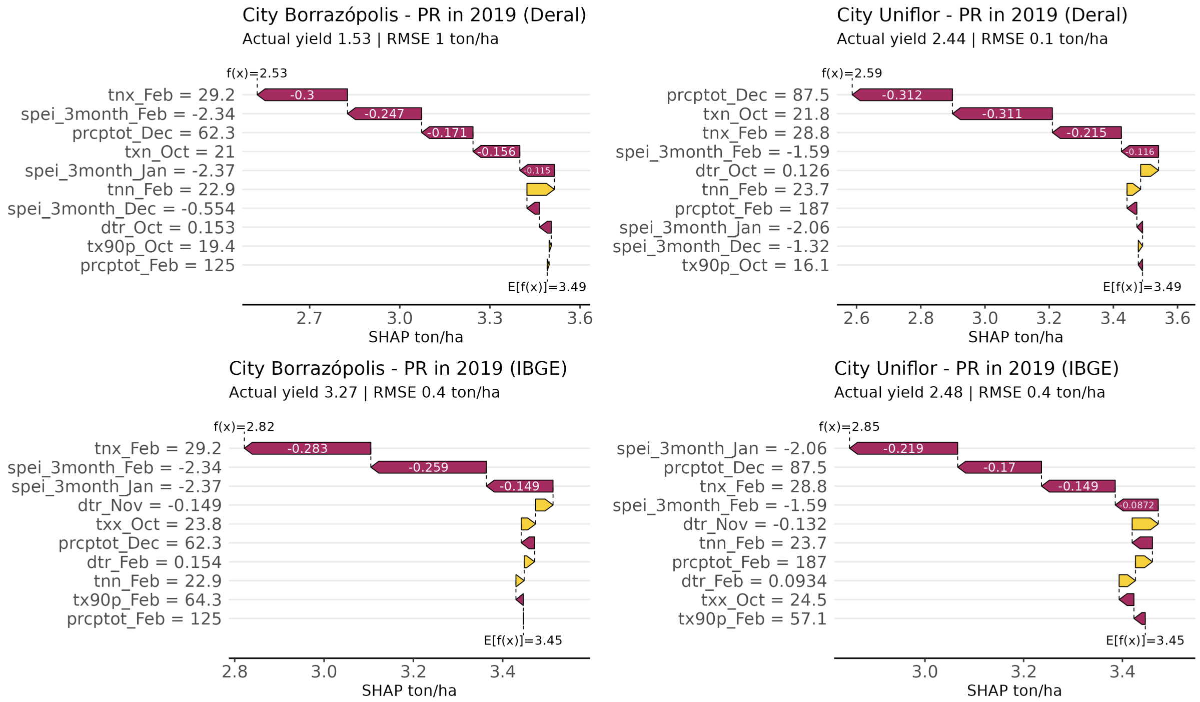


**Figure 3.11.** SHAP waterfall plot visualizing the key climatic impact-drivers contributions to crop yield losses for the state of Paraná (PR) in 2019, a drought year.Monthly Maximum Value of Daily Minimum Temperature (tnx), Minimum Value of Daily Maximum Temperature (txn), Minimum Value of Daily Minimum (tnn), Maximum Value of Daily Maximum (txx), Percentage of days when TX > 90th percentile, Standardized Precipitation Evapotranspiration Index (SPEI), Total Precipitation (prctot), Daily temperature range (DTR)

In 2014/2015, a severe drought occurred in southeastern Brazil, causing an unprecedented water supply shortage in the Cantareira Water Supply System and affecting many cities in São Paulo (Deusdará-Leal et al., 2019). The drought had repercussions in many regions of the state of São Paulo, therefore, we compared the results of IBGE and GDHY for two municipalities of the state of So Paulo, Orlândia and Pontal. The expected values of GDHY were lower than those of IBGE, and the two datasets diverged in terms of yields below the expected value, i.e., IBGE indicated losses and GDHY did not indicate losses. According to a report by the Brazilian Ministry of Agriculture, Livestock and Food Supply (MAPA), the state of São Paulo was one of the most affected by losses in the agricultural year 2014/2015 (MAPA, 2022). This result suggests that, although it has been suggested that GDHY is recommended in data-scarce regions (Iizumi e Sakai, 2020), the use of this dataset requires caution.

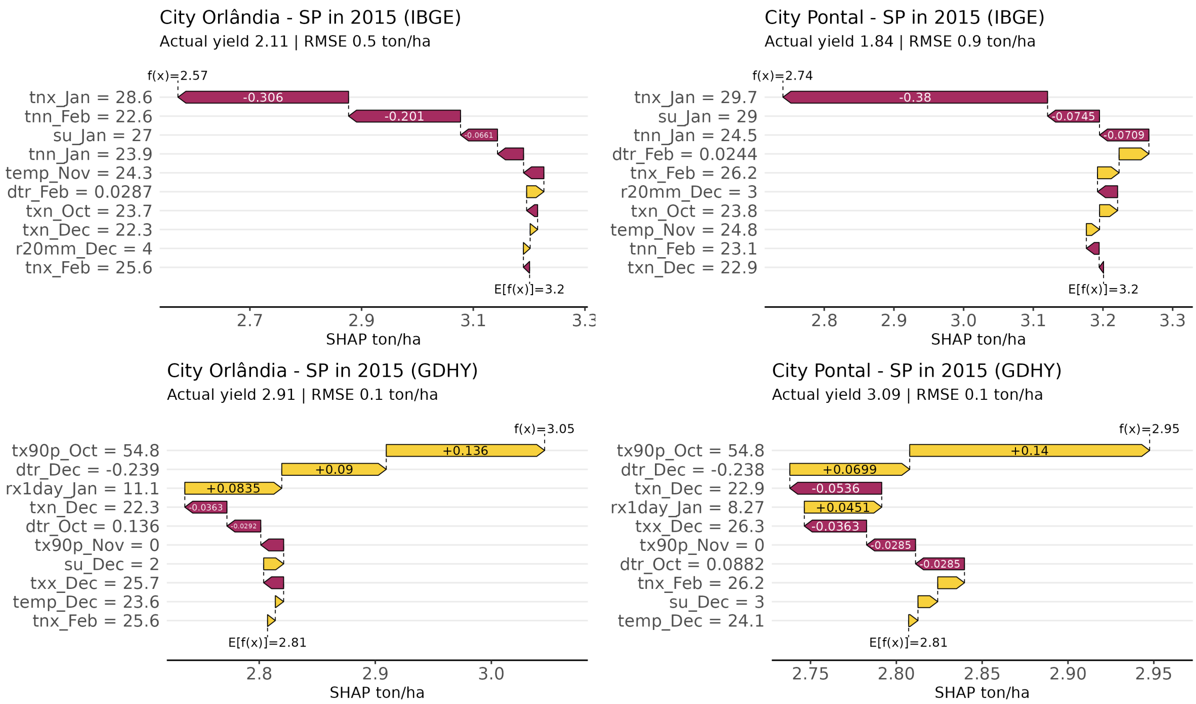


**Figure 3.12.** SHAP waterfall plot visualizing the key climatic impact-drivers contributions to crop yield losses for the state of São Paulo (SP) in 2015, a drought year. Monthly Maximum Value of Daily Minimum Temperature (tnx), Minimum Value of Daily Maximum Temperature (txn), Minimum Value of Daily Minimum (tnn), Maximum Value of Daily Maximum (txx), Percentage of days when TX > 90th percentile, Standardized Precipitation Evapotranspiration Index (SPEI), Total Precipitation (prctot), Daily temperature range (DTR)

The SHAP methodology analyzes each individual prediction and shows how each variable was used in the model. This helps us to create a partial dependence plot, which relates the value of the variable with the impact in terms of crop yield losses represented by the SHAP value. This analysis is illustrated in Fig. 3.13, which shows partial dependence plots for the state of Rio Grande do Sul. We compared two datasets. The IBGE data set shows that the 3-month SPEI in February can influence crop yield losses and has an upper and lower threshold. The lower threshold is -1.5, values below this value can represent losses of up to 0.8 ton/ha. In general, 3-month SPEI values below -1.5 are considered critical and are used as a reference for the severity of the drought (Chiang et al., 2021). The upper limit indicates that extreme wet conditions also affect crop yields in the RS state. We find threshold 1. Excess rainfall can have the same impact as droughts (Li et al., 2019), however, little attention has been paid to this analysis. Our results suggest that excessive precipitation can be responsible for up to 0.4 ton/ha of losses. More studies are suggested on this type of hazard. The cumulative precipitation in February presented a threshold of around 100 mm and the potential to cause losses of approximately 1 ton/ha. The cumulative precipitation in December was the only indicator that had a similar result in both the IBGE and GDHY datasets. In terms of potential losses, both agree in a value of up to 0.8 ton/ha, however, the for IBGE is 150 mm and for GDHY it is 100 mm.

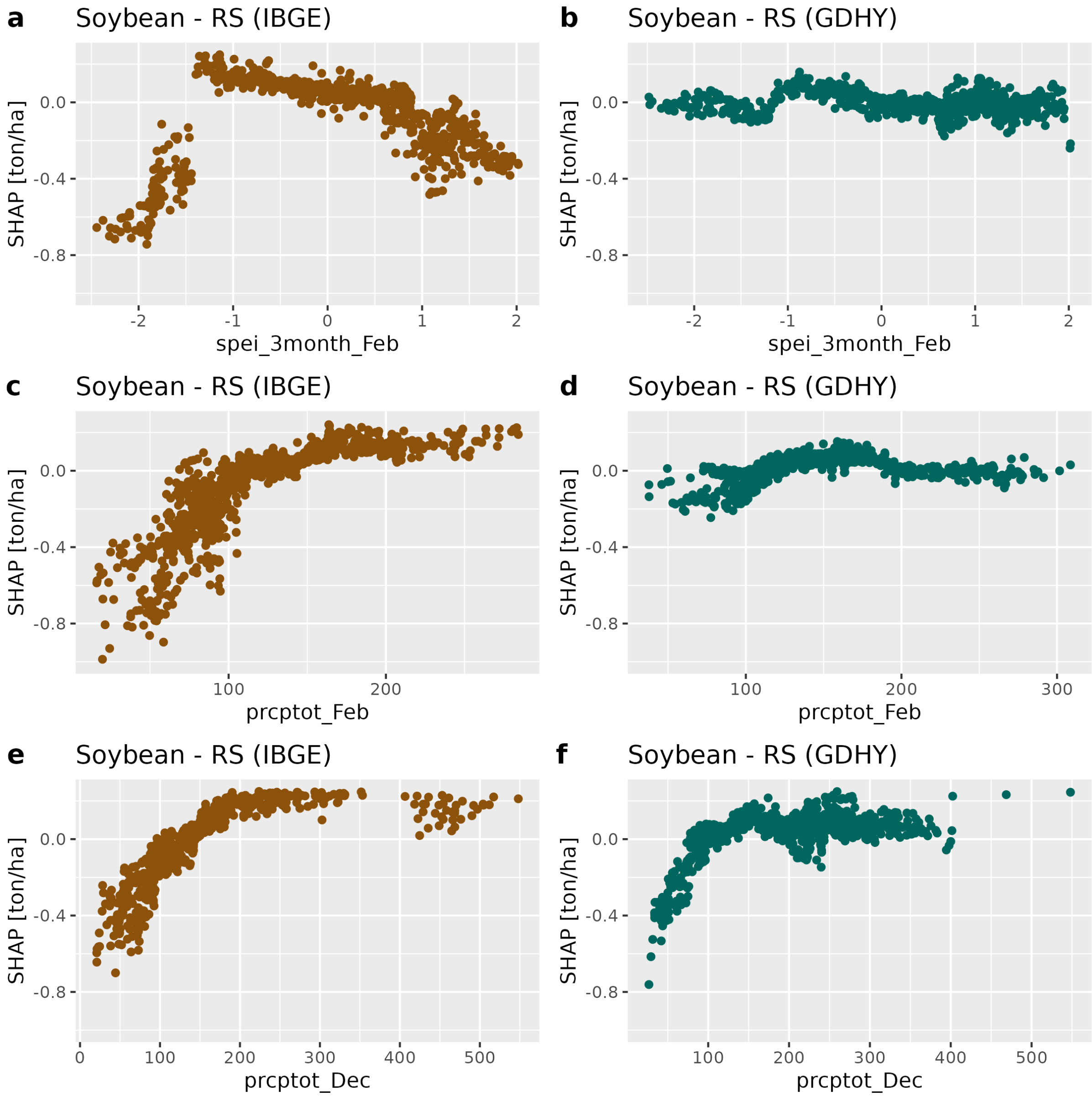


**Figure 3.13.** A comparison of the key climatic impact-drivers derived from the Brazilian Institute of Geography and Statistics (IBGE) from 2013 to 2021 and the Global Dataset on Historical Yields (GDHY) from 2009 to 2016 annual data aggregated at the municipal level was used to create a dependence plot for the soybean explanation model for validation data in RS. The data spanned from 2013 to 2021 for IBGE and from 2008 to 2016 for GDHY.. Considering 3-month SPEI in February (spei_3month_Feb) for (a) IBGE and (b) GDHY; precipitation accumulated in February (prcptot_Feb) for (c) IBGE and (d) GDHY, and precipitation accumulated in December (prcptot_Dec) for (e) IBGE and (f) GDHY

The patterns of crop yield losses observed in the region raise two main concerns. The first is that the severe crop yield losses presented in the previous examples have happened only once in the entire time series, representing an imbalance in the values of the data set. One implication of this situation is that models may not have sufficient cases of severe failure to be adequately trained and may underestimate losses. The second concern is related to the decision to do with these anomalous events. Possible solutions include using it for training, using it for testing, or removing it from the dataset. We opted to keep these events in the analysis with the warning that this might interfere with model performance. However, we wanted to evaluate the ability of the model to predict unprecedented loss events. Future studies can address these challenges in a combination of data-driven techniques, modeling, and expert evaluation. Data-Driven Techniques such as data augmentation have been proven effective in increasing the accuracy of crop yield predictions. The trigonometric Euclidean-smoother interpolation (TESI) method for continuous time-series and non-time-series data augmentation, which demonstrated promising results in preventing over-fitting and under-fitting in deep neural networks (DNNs). Derraz et al. (2023). Building ensemble models that incorporate the predictions from many models. Even with imbalanced datasets, ensemble approaches can increase the model's robustness and generalizability (Zhang et al., 2022; Prodhan et al., 2022). In the context of modeling crop yield losses, experts can provide domain-specific knowledge and expertise that contribute to understanding crop yield losses Monteleone et al. (2023)

### 3.3.4 Evaluating Combined Hazards

The SHAP algorithm also allowed us to investigate the compound effect of climate indicators. In Fig. 3.14 we exemplify the detection of compound event effects, considering the most important variable in the state of PR (prcptot_Dec) with four other variables that were considered important. We observed the presence of hot and dry compound events, that is, a combination of high temperatures and a precipitation deficit. Compound hot and dry events has been cited as an increasing threat to food production (Lesk et al., 2022; Zscheischler et al., 2018; Hamed et al., 2021). From this analysis, we observed that precipitation in December had a similar effect on crop yield losses as 3-month SPEI in January. As discussed previously, this is expected, since SPEI takes into account previous conditions. It is important to note that December is a critical month for droughts in PR and other states, such as Rio Grande do Sul (RS), Santa Catarina (SC), and Mato Grosso do Sul (MS) (see Section 3.3.2.1). Interestingly, the indices based on the minimum daily temperature were those that best reflected the impact of hot days. The maximum value of the daily minimum temperature in the month of February (tnx_Feb) presented critical values of 27.7 °C and the minimum value of the daily minimum temperature in February (tnn_Feb) presented a critical value of 22 °C.

Our use of XGBoost with SHAP explanation provides an advance by enabling quantification of the combined effect of multi-hazards on food production. In the realm of risk for food production, this method could be applied to explain the seasonal impact forecast made with composite indicators such as the Integrated Drought Index (IIS) (Cunha et al., 2018; Marengo et al., 2017). This method can also be readily applied to other natural hazards, such as landslides, floods, and wildfires.

## 3.4 Conclusions

This study aimed to assess the impacts of climate extremes on food production, using explainable machine learning algorithms. To achieve this goal, we conducted an extensive examination of various datasets, focusing on soybean and second season maize in Brazil. Our data sources included the Department of Rural Economy, the Brazilian Institute of Geography

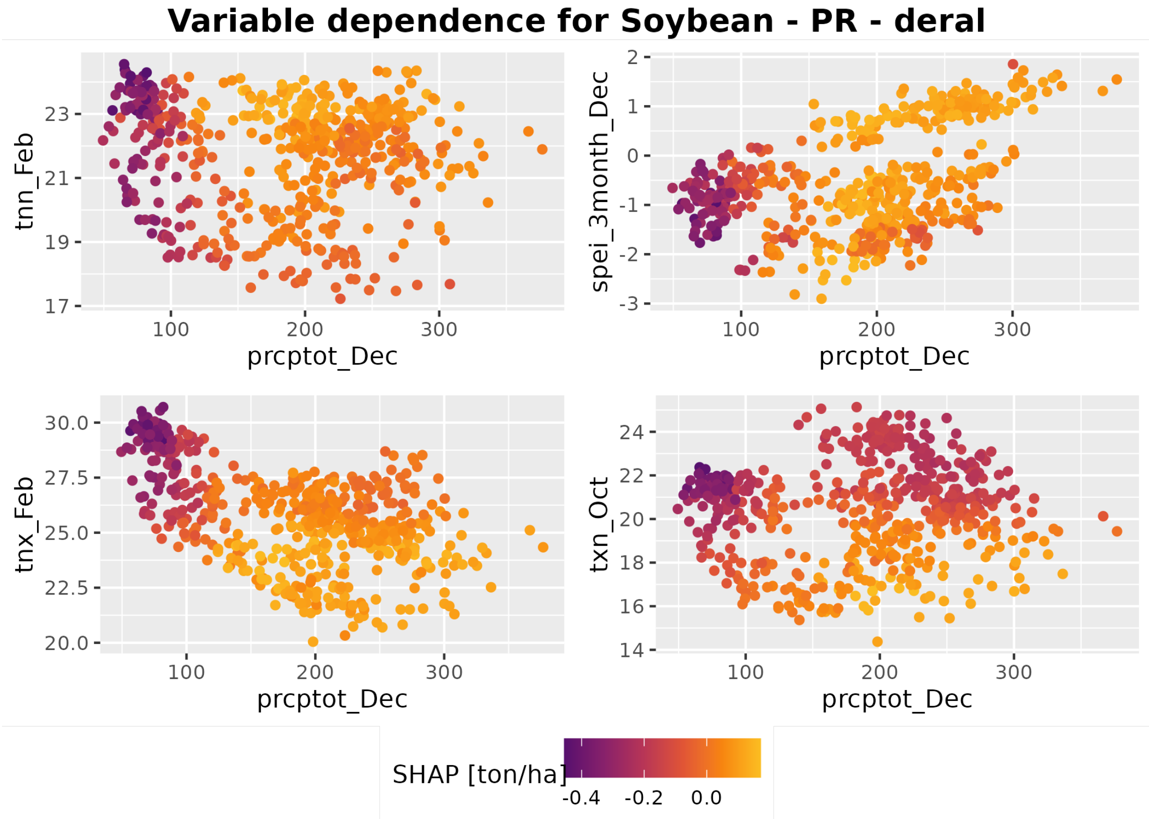


**Figure 3.14.** 2D partial dependence derived from from the Department of Rural Economics (Deral) of the state of Parana from 2016 to 2021 of the key climatic impact-driver total precipitation in December (prctot_Dec) with the most correlated CID and their combined impact on SHAP values (ton/ha): daily temperature range in October (dtr_Oct); Considering 3-month SPEI in Jan (spei_3month_Jan); Daily Minimum Temperature in February (tnx_Feb), Minimum Value of Daily Minimum (tnn_Feb)

and Statistics (IBGE), and the Global Dataset for Historical Yields (GDHY). Through a machine learning analysis, we examined the effects of climate extremes on crop yield production, ultimately providing critical insights for the agricultural sector. Our analysis incorporated data from several Brazilian states, including RS, SC, PR, SP, MS, MG and GO for soybeans, and PR, SP, MS, MG and GO for the second cycle of maize.

We conducted two machine models to achieve our research objectives. In the first model, we explored different combinations of input data, encompassing precipitation means, temperature means, as well as more complex combinations, including precipitation, temperature means, and extremes. This approach allowed us to determine the most relevant climate indices for the regions under investigation. In particular, this experiment validated the robustness of our methodology, as it successfully identified climate indices of particular significance for regional studies.

In the second model, we took the most relevant indices derived from the first experiment and used them in an XGBoost model. We then applied Shapley Additive Explanations (SHAP) explanatory analysis to explore how the important indices were utilized by the random forest model to predict the impact of climate extremes on food production. This analysis not only revealed the impact of these indices, but also provided insights that may be crucial in establishing significant thresholds and guidelines for effective climate-driven decision making.

In conclusion, our research exemplifies the potential of machine learning to understand and harness the influence of climate variables on food production. By determining the most pertinent CIDs and exploring their significance in a regional context, our findings contribute to a growing body of knowledge that is critical for informed decision-making, policy development, and adaptive strategies in the face of climate change and its impact on agriculture. The combination of data-driven insights and advanced modeling techniques, as demonstrated in our study, offers a valuable pathway toward ensuring food security under climate change.

## References


Ai, Z. e Hanasaki, N. (2023). Simulation of crop yield using the global hydrological model h08 (crp. v1). *Geoscientific Model Development*, 16(11):3275–3290.

Almendra-Martín, L., Martínez-Fernández, J., González-Zamora, Á., Benito-Verdugo, P., e Herrero-Jiménez, C. M. (2021). Agricultural drought trends on the iberian peninsula: An analysis using modeled and reanalysis soil moisture products. *Atmosphere*, 12(2):236.

Aydin, H. E. e Iban, M. C. (2023). Predicting and analyzing flood susceptibility using boosting-based ensemble machine learning algorithms with shapley additive explanations. *Natural Hazards*, 116(3):2957–2991.

Battisti, R. e Sentelhas, P. C. (2014). New agroclimatic approach for soybean sowing dates recommendation: A case study. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 18:1149–1156.

Beguería, S. e Vicente-Serrano, S. M. (2023). *SPEI: Calculation of the Standardized Precipitation-Evapotranspiration Index*. R package version 1.8.1.

Bray, E. A. (2007). Plant response to water-deficit stress. *Encyclopedia of Life Sciences*.

Breiman, L. (1997). Arcing the edge. Technical report, Citeseer.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Breusch, T. S. e Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the econometric society*, pages 1287–1294.

Chan, J. Y.-L., Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z.-W., e Chen, Y.-L. (2022). Mitigating the multicollinearity problem and its machine learning approach: a review. *Mathematics*, 10(8):1283.

Chen, T. e Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., e Yuan, J. (2023). *xgboost: Extreme Gradient Boosting*. R package version 1.7.5.1.

Chiang, F., Mazdiyasni, O., e AghaKouchak, A. (2021). Evidence of anthropogenic impacts on global drought frequency, duration, and intensity. *Nature communications*, 12(1):2754.

Christenhusz, M. J. e Byng, J. W. (2016). The number of known plants species in the world and its annual increase. *Phytotaxa*, 261(3):201–217.

Cleveland, W. S., Grosse, E., e Shyu, W. M. (2017). Local regression models. In *Statistical models in S*, pages 309–376. Routledge.

Condom, T., Martínez, R., Pabón, J. D., Costa, F., Pineda, L., Nieto, J. J., López, F., e Villacis, M. (2020). Climatological and hydrological observations for the south american andes: in situ stations, satellite, and reanalysis data sets. *Frontiers in Earth Science*, 8:92.

Cunha, A. P. M. d. A., Marengo, J. A., Cuartas, L. A., Tomasella, J., e Leal, K. R. D. (2018). Drought monitoring and impacts assessment in brazil: The cemaden experience. *ICHARM*.

Das, S., Das, J., e Umamahesh, N. (2022). Copula-based drought risk analysis on rainfed agriculture under stationary and non-stationary settings. *Hydrological Sciences Journal*, 67(11):1683–1701.

de Geografia e Estatística, I. B. (2022). Produção agrícola municipal 2022. http://www.sidra.ibge.gov.br/bda/pesquisas/pam.

Derraz, R., Muharam, F. M., Jaafar, N. A., e Yap, N. K. (2023). Trigonometric-euclidean-smoother interpolator (tesi) for continuous time-series and non-time-series data augmentation for deep neural network applications in agriculture. *Computers and Electronics in Agriculture*, 206:107646.

Deusdará-Leal, K., Cuartas, L., Zhang, R., Mohor, G., Carvalho, L., Nobre, C., Mendiondo, E., Broedel, E., Seluchi, M., e Alvalá, R. (2019). Implication of the new operation rules for cantareira system: Re-reading of the 2014/2015 water crisis. *Water Resour. Res.*

Droogers, P. e Allen, R. G. (2002). Estimating reference evapotranspiration under inaccurate data conditions. *Irrigation and drainage systems*, 16:33–45.

FAO (n.d.). Agricultural production statistics 2000–2021. `https://www.fao.org/3/cc3751en/cc3751en.pdf`. (Accessed on 10/11/2023).

Gelcer, E., Fraisse, C., Dzotsi, K., Hu, Z., Mendes, R., e Zotarelli, L. (2013). Effects of el niño southern oscillation on the space–time variability of agricultural reference index for drought in midlatitudes. *Agricultural and Forest Meteorology*, 174:110–128.

Guo, D., Westra, S., e Peterson, T. (2022). *Evapotranspiration: Modelling Actual, Potential and Reference Crop Evapotranspiration*. R package version 1.16.

Guyon, I. e Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182.

Hamed, R., Van Loon, A. F., Aerts, J., e Coumou, D. (2021). Impacts of compound hot–dry extremes on us soybean yields. *Earth System Dynamics*, 12(4):1371–1391.

Han, L., Yang, G., Dai, H., Xu, B., Yang, H., Feng, H., Li, Z., e Yang, X. (2019). Modeling maize above-ground biomass based on machine learning approaches using uav remote-sensing data. *Plant methods*, 15(1):1–19.

Harri, A., Coble, K. H., Ker, A. P., e Goodwin, B. J. (2011). Relaxing heteroscedasticity assumptions in area-yield crop insurance rating. *American Journal of Agricultural Economics*, 93(3):707–717.

Hartman, G. L., West, E. D., e Herman, T. K. (2011). Crops that feed the world 2. soybean—worldwide production, use, and constraints caused by pathogens and pests. *Food Security*, 3:5–17.

Hijmans, R. J. (2023a). *terra: Spatial Data Analysis*. R package version 1.7-29.

Hutson, A. D. (1999). Calculating nonparametric confidence intervals for quantiles using fractional order statistics. *Journal of Applied Statistics*, 26(3):343–353.

Hutson, N., Hutson, A., e Yan, L. (2019). *QuantileNPCI: Nonparametric Confidence Intervals for Quantiles*. R package version 0.9.0.

Iizumi, T. e Sakai, T. (2020). The global dataset of historical yields for major crops 1981–2016. *Scientific Data*, 7(1):97.

Iizumi, T., Yokozawa, M., Sakurai, G., Travasso, M. I., Romanenkov, V., Oettli, P., Newby, T., Ishigooka, Y., e Furuya, J. (2014). Historical changes in global yields: major cereal and legume crops from 1982 to 2006. *Global ecology and biogeography*, 23(3):346–357.

Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim, K.-M., Gerber, J. S., Reddy, V. R., et al. (2016). Random forests for global and regional crop yield predictions. *PloS one*, 11(6):e0156571.

Kim, W., Iizumi, T., e Nishimori, M. (2019). Global patterns of crop production losses associated with droughts from 1983 to 2009. *Journal of Applied Meteorology and Climatology*, 58(6):1233–1244.

Lam, M. R., Matanó, A., Van Loon, A. F., Odongo, R., Teklesadik, A. D., Wamucii, C. N., van den Homberg, M. J., Waruru, S., e Teuling, A. J. (2022). Linking reported drought impacts with drought indices, water scarcity, and aridity: the case of kenya. *EGUsphere*, pages 1–25.

Lesk, C., Anderson, W., Rigden, A., Coast, O., Jägermeyr, J., McDermid, S., Davis, K. F., e Konar, M. (2022). Compound heat and moisture extreme impacts on global crop yields under climate change. *Nature Reviews Earth & Environment*, 3(12):872–889.

Lesk, C., Coffel, E., Winter, J., Ray, D., Zscheischler, J., Seneviratne, S. I., e Horton, R. (2021). Stronger temperature–moisture couplings exacerbate the impact of climate warming on global crop yields. *Nature Food*, 2(9):683–691.

Li, Y., Guan, K., Schnitkey, G. D., DeLucia, E., e Peng, B. (2019). Excessive rainfall leads to maize yield loss of a comparable magnitude to extreme drought in the united states. *Global change biology*, 25(7):2325–2337.

Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3):31–57.

Liu, Y. e Ker, A. P. (2020). When less is more: on the use of historical yield data with application to rating area crop insurance contracts. *Journal of Agricultural and Applied Economics*, 52(2):194–203.

Lundberg, S. e et al. (2023). A game theoretic approach to explain the output of any machine learning model. [Accessed 30-10-2023].

Lundberg, S. M. e Lee, S.-I. (2017a). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica: Journal of the econometric society*, pages 245–259.

MAPA, M. d. A. P. e. A. (2022). Histórico de perdas na agricultura brasileira : 2000-2021. `https://www.gov.br/agricultura/pt-br/assuntos/riscos-seguro/seguro-rural/publicacoes-seguro-rural/historico-de-perdas-na-agricultura-brasileira-2000-2021.pdf`. [Accessed 28-10-2023].

Marengo, J. A., Alves, L. M., Alvala, R., Cunha, A. P., Brito, S., e Moraes, O. L. (2017). Climatic characteristics of the 2010-2016 drought in the semiarid northeast brazil region. *Anais da Academia Brasileira de Ciências*, 90:1973–1985.

Marengo, J. A., Cunha, A. P., Cuartas, L. A., Deusdará Leal, K. R., Broedel, E., Seluchi, M. E., Michelin, C. M., De Praga Baião, C. F., Chuchón Angulo, E., Almeida, E. K., et al. (2021). Extreme drought in the brazilian pantanal in 2019–2020: characterization, causes, and impacts. *Frontiers in Water*, 3:639204.

Mariadass, D. A., Moung, E. G., Sufian, M. M., e Farzamnia, A. (2022). Extreme gradient boosting (xgboost) regressor and shapley additive explanation for crop yield prediction in agriculture. In *2022 12th International Conference on Computer and Knowledge Engineering (ICCKE)*, pages 219–224. IEEE.

Mayer, M. (2023). *shapviz: SHAP Visualizations*. R package version 0.9.1.

McKee, T. B. (1995). Drought monitoring with multiple time scales. In *Proceedings of 9th Conference on Applied Climatology, Boston, 1995*.

McKee, T. B., Doesken, N. J., Kleist, J., et al. (1993). The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology, California, 1993*.

McLeod, A. (2022). *Kendall: Kendall Rank Correlation and Mann-Kendall Trend Test.* R package version 2.2.1.

Monfreda, C., Ramankutty, N., e Foley, J. A. (2008). Farming the planet: 2. geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global biogeochemical cycles*, 22(1).

Monteleone, B., Giusti, R., Magnini, A., Arosio, M., Domeneghetti, A., Borzì, I., Petruccelli, N., Castellarin, A., Bonaccorso, B., e Martina, M. L. (2023). Estimations of crop losses due to flood using multiple sources of information and models: The case study of the panaro river. *Water*, 15(11):1980.

Moriondo, M., Giannakopoulos, C., e Bindi, M. (2011). Climate change impact assessment: the role of climate extremes in crop yield simulation. *Climatic change*, 104(3-4):679–701.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., et al. (2021). Era5-land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13(9):4349–4383.

Ozaki, V. A., Goodwin, B. K., e Shirota, R. (2008). Parametric and nonparametric statistical modelling of crop yield: implications for pricing crop insurance contracts. *Applied Economics*, 40(9):1151–1164.

Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R. L., e Mouazen, A. M. (2016). Wheat yield prediction using machine learning and advanced sensing techniques. *Computers and electronics in agriculture*, 121:57–65.

Parana (2021). Levantamento da produção agropecuária. `https://www.agricultura.pr.gov.br/deral/ProducaoAnual`.

Poggio, L., De Sousa, L., Batjes, N., Heuvelink, G., Kempen, B., Ribeiro, E., e Rossiter, D. (2021). Soilgrids 2.0: producing soil information for the globe with quantified spatial uncertainty, soil, 7, 217–240.

Potapov, P., Turubanova, S., Hansen, M. C., Tyukavina, A., Zalles, V., Khan, A., Song, X.-P., Pickens, A., Shen, Q., e Cortez, J. (2022). Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century. *Nature Food*, 3(1):19–28.

Proctor, J., Rigden, A., Chan, D., e Huybers, P. (2022). More accurate specification of water supply shows its importance for global crop production. *Nature Food*, 3(9):753–763.

Prodhan, F. A., Zhang, J., Sharma, T. P. P., Nanzad, L., Zhang, D., Seka, A. M., Ahmed, N., Hasan, S. S., Hoque, M. Z., e Mohana, H. P. (2022). Projection of future drought and its impact on simulated crop yield over south asia using ensemble machine learning approach. *Science of The Total Environment*, 807:151029.

Pullanagari, R. R., Kereszturi, G., e Yule, I. (2018). Integrating airborne hyperspectral, topographic, and soil data for estimating pasture quality using recursive feature elimination with random forest regression. *Remote Sensing*, 10(7):1117.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ranasinghe, R., Ruane, A. C., Vautard, R., Arnell, N., Coppola, E., Cruz, F. A., Dessai, S., Saiful Islam, A., Rahimi, M., Carrascal, D. R., et al. (2021). Chapter 12: Climate Change Information for Regional Impact and for Risk Assessment — ipcc.ch. `https://www.ipcc.ch/report/ar6/wg1/chapter/chapter-12/`. [Accessed 08-10-2023].

Rapoport, E. H. e Drausal, B. S. (2013). Edible plants. *Encyclopedia of Biodiversity (Second Edition)*.

Ray, D. K., Gerber, J. S., MacDonald, G. K., e West, P. C. (2015). Climate variation explains a third of global crop yield variability. *Nature communications*, 6(1):5989.

Ribeiro, M. T., Singh, S., e Guestrin, C. (2016). " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144.

RODRIGUES, M., Corá, J. E., Castrignanò, A., Mueller, T. G., e Rienzi, E. (2013). A spatial and temporal prediction model of corn grain yield as a function of soil attributes. *Agronomy Journal*, pages 1878–1887.

Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A., Boote, K. J., Folberth, C., Glotter, M., Khabarov, N., et al. (2014). Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proceedings of the national academy of sciences*, 111(9):3268–3273.

Ruane, A. C., Vautard, R., Ranasinghe, R., Sillmann, J., Coppola, E., Arnell, N., Cruz, F. A., Dessai, S., Iles, C. E., Islam, A. S., et al. (2022). The climatic impact-driver framework for assessment of risk-relevant climate information. *Earth's Future*, 10(11):e2022EF002803.

Sacks, W. J., Deryng, D., Foley, J. A., e Ramankutty, N. (2010). Crop planting dates: an analysis of global patterns. *Global ecology and biogeography*, 19(5):607–620.

Santini, M., Noce, S., Antonelli, M., e Caporaso, L. (2022). Complex drought patterns robustly explain global yield loss for major crops. *Scientific reports*, 12(1):5792.

Sarhadi, A., Ausín, M. C., Wiper, M. P., Touma, D., e Diffenbaugh, N. S. (2018). Multidimensional risk in a nonstationary climate: Joint probability of increasingly severe warm and dry conditions. *Science Advances*, 4(11):eaau3487.

Schierhorn, F., Hofmann, M., Gagalyuk, T., Ostapchuk, I., e Müller, D. (2021). Machine learning reveals complex effects of climatic means and weather extremes on wheat yields during different plant developmental stages. *Climatic Change*, 169(3-4):39.

Schyns, J. F., Hoekstra, A. Y., e Booij, M. J. (2015). Review and classification of indicators of green water availability and scarcity. *Hydrology and Earth System Sciences*, 19(11):4581–4608.

Shiferaw, B., Prasanna, B. M., Hellin, J., e Bänziger, M. (2011). Crops that feed the world 6. past successes and future challenges to the role played by maize in global food security. *Food security*, 3:307–327.

Sidhu, B. S., Mehrabi, Z., Ramankutty, N., e Kandlikar, M. (2023). How can machine learning help in understanding the impact of climate change on crop yields? *Environmental Research Letters*.

Silva Fuzzo, D. F., Carlson, T. N., Kourgialas, N. N., e Petropoulos, G. P. (2020). Coupling remote sensing with a water balance model for soybean yield predictions over large areas. *Earth Science Informatics*, 13:345–359.

Sinnathamby, S., Douglas-Mankin, K. R., e Craige, C. (2017). Field-scale calibration of crop-yield parameters in the soil and water assessment tool (swat). *Agricultural water management*, 180:61–69.

Strumbelj, E. e Kononenko, I. (2010). An efficient explanation of individual classifications using game theory. *The Journal of Machine Learning Research*, 11:1–18.

Štrumbelj, E. e Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and information systems*, 41:647–665.

Sutton, C. D. (2005). 11 - classification and regression trees, bagging, and boosting. In Rao, C., Wegman, E., e Solka, J., editors, *Data Mining and Data Visualization*, volume 24 of *Handbook of Statistics*, pages 303–329. Elsevier.

Svetnik, V., Liaw, A., Tong, C., e Wang, T. (2004). Application of breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules. In *Multiple Classifier Systems: 5th International Workshop, MCS 2004, Cagliari, Italy, June 9-11, 2004. Proceedings 5*, pages 334–343. Springer.

Tolhurst, T. N. e Ker, A. P. (2015). On technological change in crop yields. *American Journal of Agricultural Economics*, 97(1):137–158.

UNDRR (n.d.). What is the Sendai Framework for Disaster Risk Reduction? — undrr.org. `https://www.undrr.org/implementing-sendai-framework/what-sendai-framework`. [Accessed 24-10-2023].

Van Klompenburg, T., Kassahun, A., e Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177:105709.

Vapnik, V. N. (1999b). An overview of statistical learning theory. *IEEE transactions on neural networks*, 10(5):988–999.

Viana, C. M., Santos, M., Freire, D., Abrantes, P., e Rocha, J. (2021). Evaluation of the factors explaining the use of agricultural land: A machine learning and model-agnostic approach. *Ecological Indicators*, 131:108200.

Vicente, J. R. (2004). Economic efficiency of agricultural production in brazil. *Revista de Economia e Sociologia Rural*, 42:201–222.

Vicente-Serrano, S. M., Beguería, S., e López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of climate*, 23(7):1696–1718.

Vogel, E., Donat, M. G., Alexander, L. V., Meinshausen, M., Ray, D. K., Karoly, D., Meinshausen, N., e Frieler, K. (2019). The effects of climate extremes on global agricultural yields. *Environmental Research Letters*, 14(5):054010.

Wang, J., Wei, J., Shan, W., e Zhao, J. (2023). Modeling the water-energy-food-environment nexus and transboundary cooperation opportunity in the brahmaputra river basin. *Journal of Hydrology: Regional Studies*, 49:101497.

Wang, Y. e Li, Y. (2023). Mapping the ratoon rice suitability region in china using random forest and recursive feature elimination modeling. *Field Crops Research*, 301:109016.

Wikle, C. K., Datta, A., Hari, B. V., Boone, E. L., Sahoo, I., Kavila, I., Castruccio, S., Simmons, S. J., Burr, W. S., e Chang, W. (2023). An illustration of model agnostic explainability methods applied to environmental data. *Environmetrics*, 34(1):e2772.

Wright, M. N. e Ziegler, A. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, 77(1):1–17.

Yang, S.-R., Koo, W. W., e Wilson, W. W. (1992). Heteroskedasticity in crop yield models. *Journal of Agricultural and Resource Economics*, pages 103–109.

Yu, L. e Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proceedings of the 20th international conference on machine learning (ICML-03)*, pages 856–863.

Zhang, Y., Wang, L., Chen, X., Liu, Y., Wang, S., e Wang, L. (2022). Prediction of winter wheat yield at county level in china using ensemble learning. *Progress in Physical Geography: Earth and Environment*, 46(5):676–696.

Zhu, Y., Goodwin, B. K., e Ghosh, S. K. (2011). Modeling yield risk under technological change: Dynamic yield distributions and the us crop insurance program. *Journal of Agricultural and Resource Economics*, pages 192–210.

Zscheischler, J., Westra, S., Van Den Hurk, B. J., Seneviratne, S. I., Ward, P. J., Pitman, A., AghaKouchak, A., Bresch, D. N., Leonard, M., Wahl, T., et al. (2018). Future climate risk from compound events. *Nature Climate Change*, 8(6):469–477.

# 4 A DYNAMIC RISK APPROACH USING MULTI-HAZARD INDEX INSURANCE DESIGN FOR CLIMATE EXTREMES

A modified version of this chapter will be submitted as follows: BENSO, Marcos Roberto; MENDIONDO, Eduardo Mario. A dynamic resilience approach using a multihazard index insurance design for extreme weather events. **International Journal of Disaster Risk Reduction**

### Abstract

Climatic impact-drivers are defined as climate means, events, and extremes that affect society and ecology. The threat that these drivers pose to food production is increasing, and further research is needed to understand their relationship and how climate change influences them in Brazil to design better risk mitigation strategies. Despite the multi-hazard nature of climate threats to food production, research in index insurance has mostly focused on single index insurance. This paper presents a multi-hazard risk approach for index insurance design for food production and climate change adaptation. We used municipal-level crop yield data for soybean (1980-2022) in 452 municipalities and for maize (2003-2022) in 216 municipalities and climate data (1980-2022). We employ a multi-hazard insurance model to evaluate the effectiveness of risk transfer strategies to improve financial resilience and increase food production under current climate conditions and also to simulate the impacts of projected climate change (1980-2100) under moderate (SSP 5 8.5) and high emission (SSP 5 8.5) scenarios. The results showed that the inclusion of climate extreme indices improves the risk reduction potential for soybeans and maize, and these improvements were statistically significant ($p \leq 0.01$) under current conditions for soybeans and maize. The projected climate changes in both scenarios suggest that monthly mean temperatures could rise to 5 °C at the end of the century, while there is no evidence of any trend in monthly precipitation averages. Droughts (SPEI less than -1.5) are projected to increase 2 to 5 times the probability and they are located in the northwest of Paraná and in the states of São Paulo, Mato Grosso do Sul, Minas Gerais and Goias. Climate change projections also indicate an increase in high temperature indices and an increase in heavy rainfall, especially in South Brazil. The dynamics of risk represents a potential increased pressure for insurance market, on average, projected climate extremes represent an increase in 26.5 % risk premiums for soybeans, and temperature means an increase in 24.6 % for maize. Our study highlights the importance of considering climate extremes to reduce the risk of financial losses for farmers and provides an overview of how the risk dynamic caused by climate change is projected to affect risk premium prices and insurance design.

## 4.1 Introduction

Climate extremes such as droughts, floods, heat waves, and heavy rains are a major concern for food security. When a natural hazard encounters conditions of exposure, vulnerability, and capacity of human systems, causing direct and indirect losses and damages that affect their welfare. The projections show that there is a potential of 20 26% more people to face hunger by 2050 due to climate extremes (Hasegawa et al., 2021), which places the link between food insecurity and resilience to climate risk management in a central position.

Resilience is a concept used in various applications of society and has different definitions. Economic resilience is the ability of society to minimize the welfare losses from an extreme event of a given magnitude (Hallegatte, 2014). Ecology-based resilience refers to the ability of a system to withstand stress, also called the 'environmental load' (Holling, 1973). Climate resilience is the capacity of a system to absorb a shock while maintaining function (Simonovic e Peck, 2013). Regarding food security, Bullock et al. (2017) theorizes that resilience has different purposes that are relevant to the present study, considering social structures and capacity building; analyzing sensitivity in models linking human population growth to food supply; maintaining agricultural production under climate change.

According to Surminski et al. (2016) payments in the aftermath of an extreme event are often the only option for people affected by climate extremes in many regions of the world. However, these payments are unpredictable, can be delayed, and are much more expensive than pre-event measures such as insurance. Food systems face various stressors, including pests, soil degradation, population growth, economic and political crises, and climate change, as noted by Godfray et al. (2010). Among these stressors, climate change is considered the most significant threat to food production.

There are numerous climate extremes that pose a threat to food production. Studies on the main cereals such as maize, wheat, rice and soybeans show that heat stress, soil moisture deficit, and excess are the main drivers of crop losses worldwide (Benso et al., 2023; Zampieri et al., 2017). Benso et al. (2023) reviewed the design of multi-hazard risk index insurance and demonstrated that although there is evidence in the literature that multi-hazard events and moisture-thermal compound events are critical

In recent years, Brazil has faced severe impacts from hydrological extremes on food production. These extremes cause severe losses and damages that affect food availability, access, utilization, and stability, which are the pillars of food security (Barrett, 2010). In the context of farming, climate extremes can induce shocks in food production, that is, sudden losses in crop yields (Cottrell et al., 2019).

As the climate system suffers the impact of human actions that leads to the global phenomenon of global warming, climate extremes become more frequent and severe (Matthews e Wynes, 2022; IPCC, 2022); therefore, understanding the role of risk mitigation strategies such as insurance is of paramount importance in protecting agricultural systems and food security. This matter paves the way for the analysis of how climate risk changes over time and the ability of farmers to cope and recover from the adverse effects of natural hazards.

There are numerous climate extremes that pose a threat to food production. Studies on cash crops show that heat stress, soil moisture deficit, and excess are the main drivers of crop losses worldwide (Benso et al., 2023; Zampieri et al., 2017). Benso et al. (2023) reviewed the design of multi-hazard risk index insurance and highlighted the importance of considering compound events involving both moisture and thermal factors, as evidenced in the literature.

In recent years, Brazil has experienced severe impacts of hydrological extremes on food production. These extremes cause severe losses and damages that affect food availability, access, utilization, and stability, which are the pillars of food security (Barrett, 2010). Temperature and moisture present a synergic relationship, also known as temperature-moisture couplings, that pose a great threat to food production (Heino et al., 2023). According to Lesk et al. (2021), one example of a temperature-moisture coupling is the atmospheric circulation coupling, that is, the combination of rising temperatures and decreasing precipitation. This phenomenon is detected considering the correlation between mean atmospheric temperature and total precipitation sum over a certain period, in this work represented as a specific season that represents growing conditions for cash crops such as soybeans and maize second cycle.

Climate extremes present challenges not only to food production but also across various sectors. This topic has been widely discussed in the literature, but there is still a lack of research linking climate indicators, their effects, and prevention methods. Consequently, the objective of this paper is to evaluate the effectiveness of multi-hazard index insurance for increasing the resilience and adaptive capacity of agricultural systems in coping with climate extremes under climate change scenarios. This research evaluates the impact of index insurance on agricultural systems' ability to cope with and effectively respond to the various difficulties posed by changing temperature and precipitation patterns. It does this by carefully analyzing the relationship between these two climate-related risks.

## 4.2 Material and methods

### 4.2.1 Index insurance model

In our study, we used a conceptual framework for multi-hazard risk index-based insurance design according to (Benso et al., 2023), outlining three key modules: hazard identification, vulnerability assessment, financial methods, and risk pricing Fig. 4.1. Hazard identification involves choosing between single-hazard and multi-hazard risk insurance and considers independent, synergistic, and cascade events. The vulnerability analysis adopted in this work involves evaluating how susceptible crop yields are to climate risks. Financial methods and risk pricing depend on regional socioeconomic factors and can involve individual or community contracts, with options for pooling multiple locations. The dynamic resilience framework anticipates increasing patterns of losses and aims to optimize risk premiums for climate change adaptation and resilience. These components interact to design insurance policies that address specific needs, with options supported by existing literature. Calculations were carried out in the R environment using version 4.1.2 of the software (2021-11-01) (R Core Team, 2021).

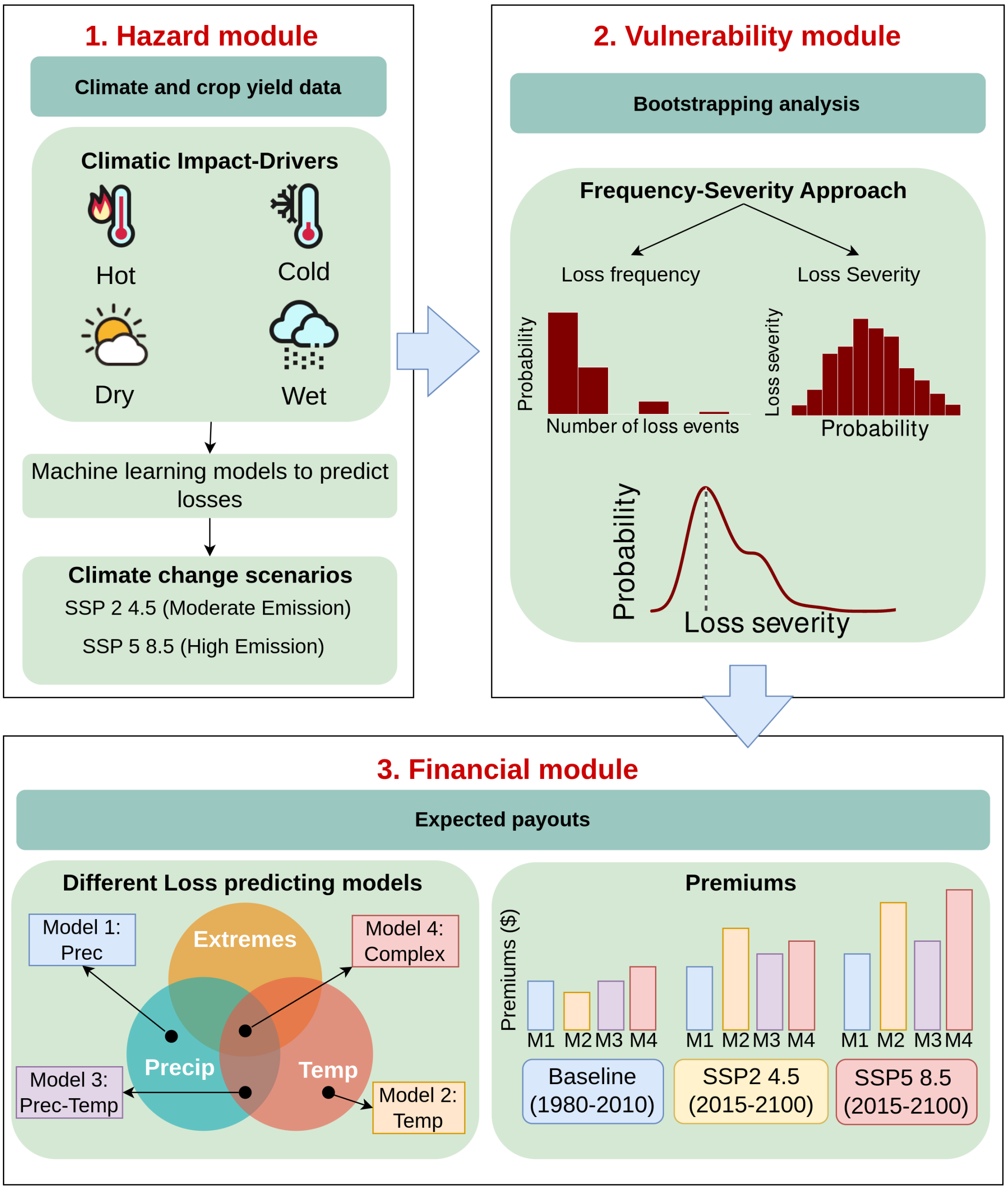


**Figure 4.1.** Description of the methodological steps presented in this study

Here, we present an insurance design and evaluation that simulates risk decision making considering a fully developed index insurance market of policyholders $I$ $i = 1, ..., I$. This methodology is used to evaluate the effectiveness of a multi-hazard risk index insurance approach.

The underlying structure of the index insurance design is the option pricing model proposed by Black e Scholes (1973). The B-S model is based on a series of assumptions that simplify the rating process; we used five that, to the best of our knowledge, fit into the main goal of this work, that is, the interest rate is risk-free, the price of commodities follows a geometric Brownian motion, the stock does not pay any dividend, the market is based on no arbitrage opportunity, the transactions, buying and selling policies, do not have fees or costs, and insurance contract operate as an European options. We assume that farmers will purchase the most effective insurance policy in terms of minimization of financial losses on the farm.

**Payout model**: The risk pricing process, as mentioned in Chapter 2, is analogous to the pricing of options or derivatives. Our contract is a modified version of a binary European option that has three tiers, instead of one only. We use the Skees et al. (1997) payoff equation $I_t$, which is given by Eq. 4.1, where $E[y]$ is the expected yield, which is the trigger value for claim payments, the yield is the yield estimated by the random forest model that associates climate extremes with crop yields and $\lambda$ is the degree of coverage.

$$I_t = \frac{max(0, E[y] \cdot \lambda - yield)}{E[y] \cdot \lambda} \tag{4.1}$$

**Trigger**: In this study, we assume that every farm within a municipality is equally covered by the insurance scheme and equally exposed to climate risks. We acknowledge that crop production within a municipality is idiosyncratic and farmers can have different levels of exposure and vulnerability to extremes, however, this effect is not evaluated in this study. If a loss event exceeds the trigger value, all farmers in the municipality will receive a proportional claim amount.

**Premium**: Risk premiums $pr$ were calculated using the expected value of the payouts $E(I)$. To calculate risk premiums, we used the probability $P(x)$ associated with each crop yield value $x$, therefore we used the adjusted crop yield data for trends and heteroskedasticity explained in Chapter 3. For each municipality, we fit the following candidate distributions: Normal, Gamma, Cauchy, Weibull and Log-normal using the maximum likelihood estimation (MLE) and using the R-package fitdistrplus (Delignette-Muller e Dutang, 2015). We then selected the most appropriate distribution according to Akaike's Information Criterium (AIC) (Akaike, 1974).

$$pr = E(I) = \sum x \cdot P(x) \tag{4.2}$$

**Claim Payment**: Claim payments in index insurance usually are a fixed small payment associated with the triggering event. Insurers will pay farmers indemnities whenever the trigger of each contract is reached or exceeded. This will happen in the month after the last month considered in the random forest model.

**Resilience**: The resilience of food production systems and insurance was evaluated using metrics that focus on farms and insurance companies. We used a resilience metric $R_{t,c}$ (Eq. 4.3) for year $t$ and municipality $c$ considering the balance between the sum of production costs $c_{t,c}$ and premiums $pr_{t,c}$ divided by the sum of indemnizations $I_{t,c}$ and the $Profit_{t,c}$. For simplicity, the cost of production was considered as 30 % of the expected yields in that municipality. The profit is the actual yield registered in a given year and municipality according to IBGE data.

$$R_{t,c} = \frac{(pr_{t,c} + c_{t,c})}{I_{t,c} + Profit_{t,c}} \tag{4.3}$$

To evaluate the effectiveness of different index insurance models, we applied the risk reduction potential (RRP) concept adapted from Mechler (2016), which consists in generating two curves: the first is the resilience with risk reduction, that is, with index insurance; and the second without risk reduction. In the second curve, we consider premiums and indemnizations in Eq. 4.3 are equal to zero. The curves are generated considering the resilience measure and the empirical cumulative distribution of each value, which represents the exceedance probability. To estimate the RRP, we calculate the difference between two areas, as demonstrated in Fig. 4.2.

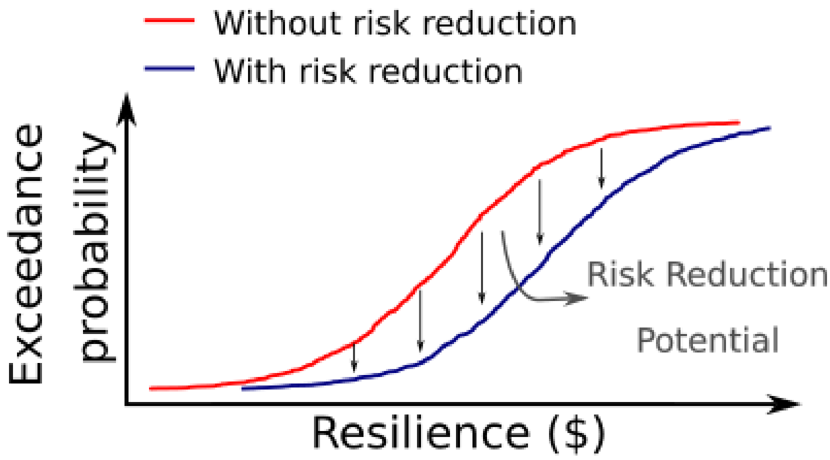


**Figure 4.2.** Theoretical definition of risk reduction potential based on the impact-probability curve that is formed by an indicator of financial impact on the sector (x-axis) and the exceedance probability (y-axis). The concept is adapted from Mechler (2016)

### 4.2.2 Data collection and processing

The fifth generation European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 Land Reanalysis dataset was used as the source of weather data (Muñoz-Sabater et al., 2021). This data set has a longitude and latitude grid of 0.1 x 0.1 °and was combined with the municipal area to match the spatial resolution of the crop yield data. The data was collected over a 41-year period, 1980 to 2021, with daily readings. The weather variables taken into account were precipitation, maximum temperature, and minimum temperature. We used various climate indices to assess multi-hazard risks. The mean climate conditions of precipitation and temperature have been shown to be the most pertinent Moriondo et al. (2011).

To assess future changes or risk dynamics, we used projected monthly mean maximum and minimum temperature and total precipitation from a group of 19 bias-corrected climate models from CMIP6 downscaled to 0.25 °for the Brazilian territory (Ballarin et al., 2023). Both historical and climate change scenario data are presented in different resolutions and projections. To make them comparable, we coarse the monthly ERA5-Land data to 0.25 °using the bilinear method using Climate Data Operators (CDO) (Schulzweida, 2023).

Crop yield data was obtained from the Municipal Agricultural Survey of the Brazilian Institute of Geography and Statistics (PAM / IBGE), at muncipal-level. We selected soybean crop data from 1974 to 2022 and maize crop yield data from 2003 to 2022 to match the period in which the double crop system was intensified in Brazil (Arvor et al., 2014; Pires et al., 2016). The precipitation and temperature of each season of the year represent crop-specific growing seasons. Crop calendars are based on a combination of favorable climatic conditions, crop phonological requirements, and crop management (Dias e Sentelhas, 2017). For this reason, we calculated climate extremes for each growing season, and for simplicity, we focused on the soybean and maize second cycle to determine the impacts of multi-hazard risks on food production.

### 4.2.3 Selection of climate models and crop calendars

Since climate is one of the most important determinants of the suitability of agricultural production (Van Wart et al., 2013), we argue that the reliability of a climate model for impact studies on agricultural areas should be measured by the ability of the model to reproduce historical agroclimatic zones. We use the three climate zones, also referred to in this work as seasonality type, according to the definition of Waha et al. (2012) based on the intra-annual variability, that is, the coefficient of variation (CV) of precipitation and temperature. The CV is characterized by the annual variation of precipitation ($CV_{PREC}$) and temperature ($CV_{PREC}$) that allows us to classify the types of seasonality:

- No seasonality ($CV_{TEMP} \leq 0.01$ and $CV_{PREC} \leq 0.4$),
- Precipitation Seasonality ($CV_{TEMP} \leq 0.01$ and $Cv_{PREC} \geq 0.4$),
- Temperature seasonality and
- Both types of seasonality ($CV_{TEMP} \geq 0.01$ and $CV_{PREC} \geq 0.4$).

The types of seasonality influence not only the suitability of a crop in a region, but also the crop calendars. When there is no seasonality, the sowing dates can be on any day of the year and, by default, they can be set to January first. In the case of precipitation seasonality, farmers choose to start the growing season when there is sufficient soil moisture for germination; this can be determined by the relationship between precipitation and potential evapotranspiration, usually, when this ratio is greater than 0.5, there is an indication of higher soil moisture available in the soil (Minoli et al., 2019). For the case of temperature seasonality, if the temperature of the coldest month is greater than 10 °C, the sowing dates are determined by the primary wet season. On the other hand, if the temperature of the coldest month is lower than 10 °C, the sowing dates are determined by crop-specific temperature thresholds (Waha et al., 2012). Seasonality types and crop calendars were calculated using the R-package cropCalendar (Minoli, 2023).

To select the climate models that best represent the climatic conditions during the soybean and maize growing seasons, we evaluated the ability of the models to estimate the types of seasonality by comparing them with historical simulations. This evaluation was carried out using precision and kappa metrics derived from a confusion matrix (Velez et al., 2007). For further details, see Velez et al. (2007).

### 4.2.4 Climate-based multi-hazard indices

For computing climate extreme indices, we used the indices indicated by the Expert Team on CLimate Change Detection and Indices (ETCCDI), which is supported by the World Meteorological Organization (WMO) Commission for Climatology, the Joint Commission for

Oceanography and Marine Meterology (JCOMM) and the Research Program on Climate Variability and Predictability (CLIVAR).

We also used two drought-related indices, the Standardized Precipitation Index (SPI) (McKee et al., 1993; McKee, 1995) and the Standardized Precipitation and Evapotranspiration Index (SPEI) (Vicente-Serrano et al., 2010). The SPI is based on the probability of monthly precipitation on different time scales, and it is recommended to be calculated with a time series of at least 30 years. The monthly time series must be fitted to a cumulative distribution function (CDF); here we adopted the gamma distribution, and then the data is transformed to the standard normal distribution to calculate the SPI, which is a value standardized in terms subtracting the transformed precipitation by mean value and dividing by the standard deviation. The SPI can be calculated using different time scales that represent previous meteorological conditions typically ranging from 1 to 48 months. For agricultural applications, 3-month SPI has been shown to be the most suitable Kim et al. (2019), however, Lam et al. (2022) has shown a counter example in Kenya showing that longer time scales could also be applied for agriculture.

The SPEI incorporates temperature in the calculation of SPI. A new step is added to the procedure consisting of calculating the monthly potential evapotranspiration (PET) and calculating the SPI using the same procedure previously described with the value of monthly precipitation minus monthly PET. PET was calculated using the Hargreaves method, which is calculated using maximum and minimum temperature and extraterrestrial radiation (RA) (Droogers e Allen, 2002). In summary, this chapter uses the climate indices presented in Section 3.2.3.3 Indices for Climatic Impact-Drivers.

## 4.3 Results and discussion

### 4.3.1 Selection of climate change models

We show the distribution of seasonality types across the Brazilian territory (Fig. 4.3). Based on the seasonality map, we defined that the climate models that best represented climate variability are UKESM1.0, K.ACE and HadGEM3.GC3.1 with Kappa of 0.62, 0.57, 0.53, respectively. However, MPI-ESM1.2-HR, KIOST-ESM models, which had Kappa values of 0.55 and 0.52, were discarded after visual inspection. A higher uncertainty was observed in the states of Paraná (PR), the southern region of Mato Grosso do Sul (MS) in the Midwest of Brazil and Amazonas (AM) and Acre (AC) in northern Brazil. The northern region of Paraná (PR) is a significant crop production area in Brazil, known particularly for its Soybean-Maize double-cropping system.

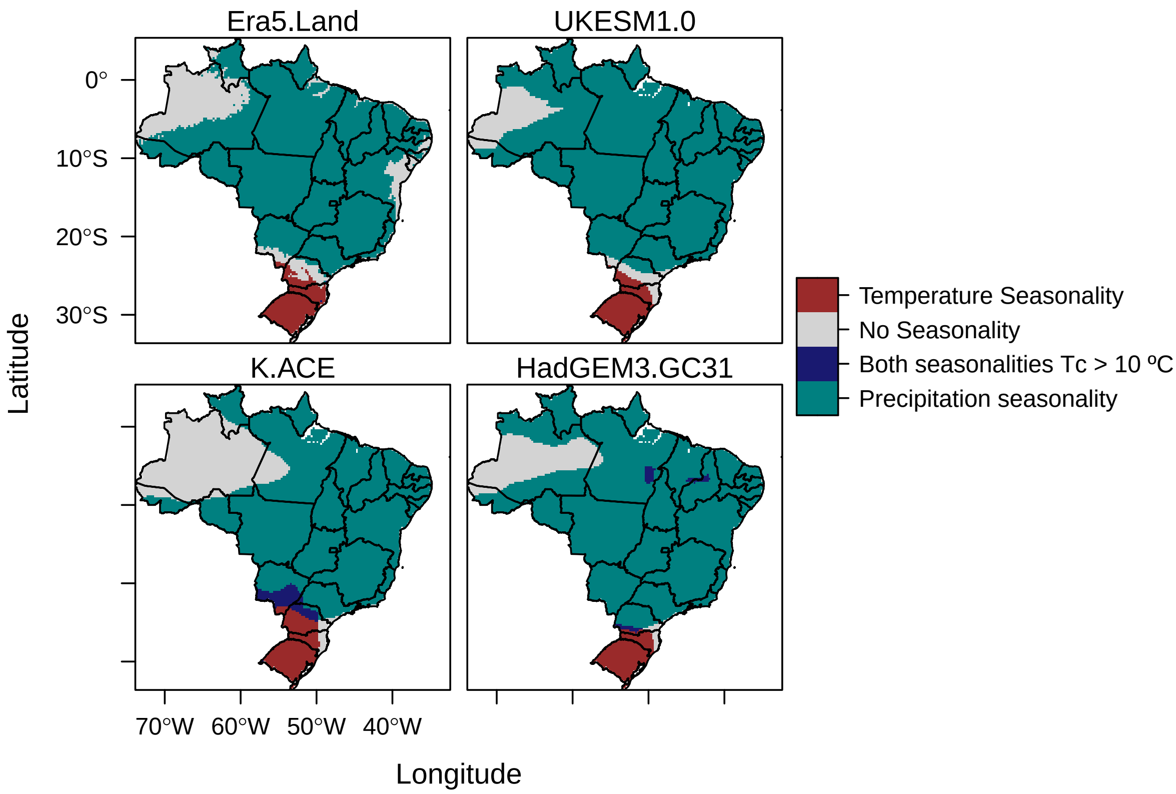


**Figure 4.3.** Comparison between seasonality maps for Brazil considering observed data (Era5 Land) and the three climate models considering historical simulations of the climate models K.ACE, UKESM1.0, and HadGEM3.GC3.1 considering the data from 1980 to 2010

We provide the calculated crop calendars for the study region, organized by agroclimatic zones in Fig. 4.3. Simulations present location-specific sowing dates and the duration of the growing season according to climate models. For the purpose of the present study, only soybeans were considered for the analysis. The crop yield cycle was fixed in 120 days.

In general, Brazil is considered highly favorable for the production of soybean-maize double cropping system (de Souza Nóia e Sentelhas, 2020). Our simulations indicated an average soybean sowing date in the late October and the beginning of November. Oligini et al. (2021) recommended sowing dates in the beginning of October or earlier to make the second season of maize feasible in Paraná. Other studies are consistent with these dates for the Midwest (Júnior e Sentelhas, 2019; Becker et al., 2020). In South Brazil, soybean sowing is recommended from

August 11 th to September 21 th and 21 th December 1 st to January 1 st (Battisti e Sentelhas, 2014), which also aligns with the simulation results.

Comparison with the literature indicates that the method proposed by Waha et al. (2012) and Minoli et al. (2019), although intended for global studies, was able to inform the sowing dates for site-specific applications only considering precipitation and temperature. Furthermore, this indicates that the selected models provided a reliable estimate of suitable planting dates based on climate variability in the regions studied.

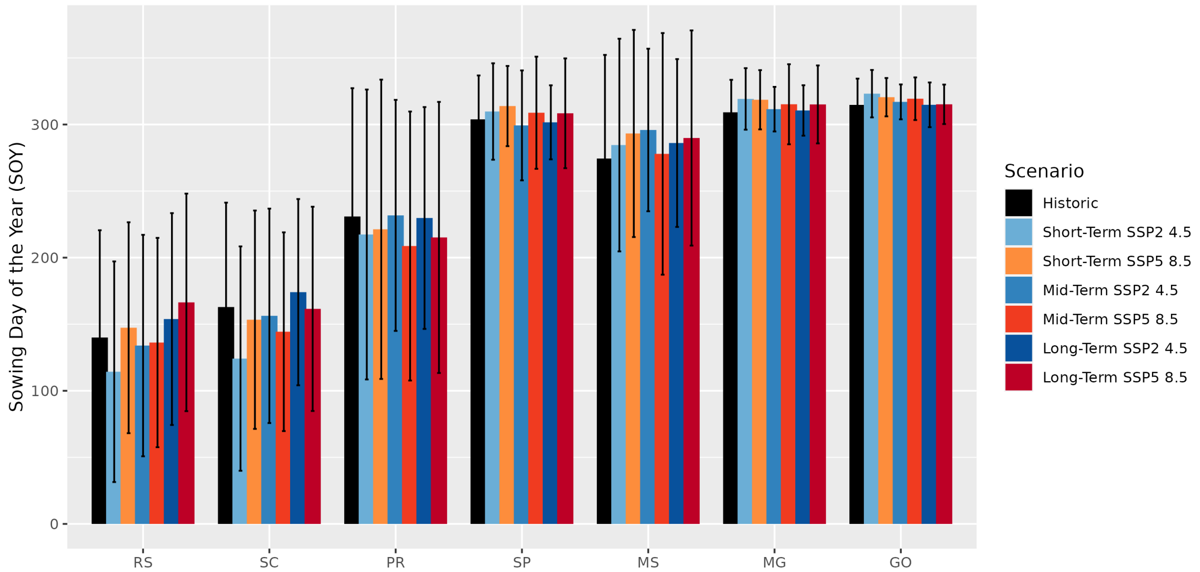


**Figure 4.4.** Definition of Sowing Day of the Year estimated using deterministic rules and K.ACE, UKESM1.0, and HadGEM3.GC3.1. climate change historic simulations and projection data for historic (1980-2010), short-term (2015-2044), mid-term (2045-2074) and log-term (2075-2100) considering moderate (SSP2 4.5) and high (SSP5 8.5) emission scenarios

Taking into account the median values, when comparing historic and future simulations, the change in the Sowing Day of the Year is less than a month for all states. This is consistent with the results presented by Minoli et al. (2019). The result indicates that the same growing season used for historic simulations also applies to future scenarios. However, it is important to state that this might not be the case for other crops such as wheat, which is much more sensitive to increasing temperatures (Franch et al., 2022).

### 4.3.2 Projected trends of Climate Means and Extremes from 1980 to 2100

The purpose of this subsection is to characterize the historical and projected temperature and precipitation for the study region, focusing on the baseline period (1980-2010) and future scenarios (2015-2100) under moderate (SSP2 4.5) and high (SSP5 8.5) emission derived from the climate models UKESM1.0, K.ACE and HadGEM3.GC3.1 of the CMIP6. The monthly averages of rainfall and temperature at the municipal level for the periods studied and the scenarios of climate change are presented as time series (Figure 4.5).

For SSP5-8.5, the comparison with historical simulations indicates a median shift to warmer conditions. The projected climate changes in both scenarios suggest that monthly mean temperatures could rise to 5 °C and 4.5 °C for high and moderate emission scenarios, respectively, at the end of the century.

In terms of the mean monthly accumulated precipitation, there is no evidence of any trend. However, it is important to evaluate recent observed trends that might be overlooked by global simulations. The simulations of the CMIP6 model from Ballarin et al. (2023) have a historic time frame from 1980 to 2013 and a future time frame from 2015 to 2100. As observed in Fig. 4.5, compared to the observed ERA5, the period 2010 and 2020 shows average precipitation means lower than the means of SSP2 4.5 and SSP5 8.5. As reported by Cunha et al. (2019), severe drought events have been identified in this decade. Other articles outline severe mega-droughts in this decade, such as the drought in 2014/2015 that affected the Cantareira Water Supply System that supplies water to the São Paulo Metropolitan Region (Deusdará-Leal et al., 2019), and the mega-drought that occurred in the Brazilian pantanal, the world's largest wetland in 2019/2020 (Marengo et al., 2021).

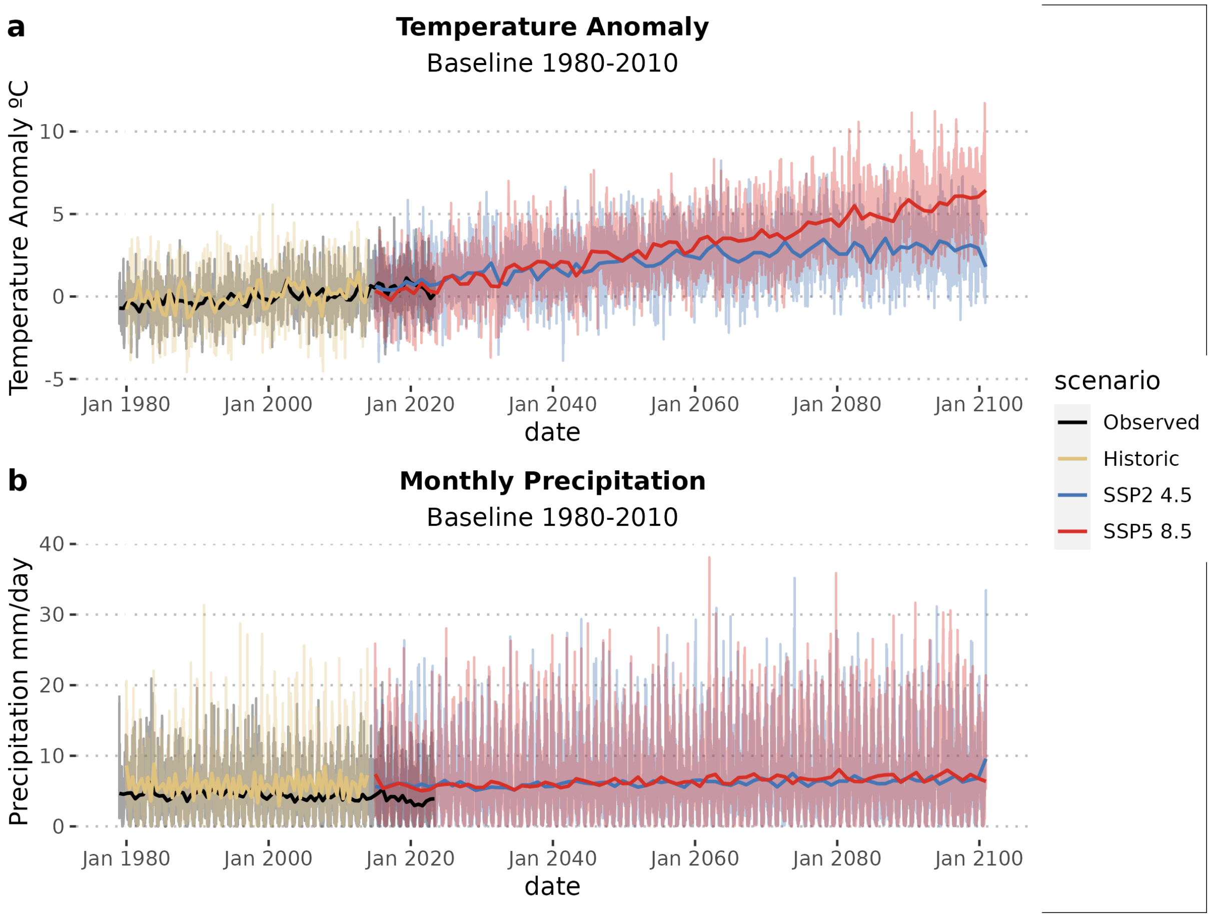


**Figure 4.5.** Observed climate from ERA5 Land for the period of 1980 to 2022 and historic simulated period from 1980 to 2013 and projected period from 2015 to 2100 from climate models K.ACE, UKESM1.0, and HadGEM3.GC3.1. Fig. (a) is the observed and simulated changes in the temperature anomaly, the anomaly, and (b) is the monthly accumulated precipitation

Using simulations of climate models, we provide a spatial meteorological perspective of the regions that are most likely to experience greater changes in drought frequency. We used SPEI values lower than −1.5 as the drought threshold as proposed by Chiang et al. (2021). From an agro-meteorological perspective, we observed 452 municipalities, including the soybean and soybean-maize double cropping system regions; We determined from our SPEI-based drought analysis that the frequency of drought will shift especially in the northwestern portion of Paraná (PR), Mato Grosso do Sul (MS), São Paulo (SP), Minas Gerais (MG), and Goiás (GO) for moderate and high emission scenarios. From a geographical perspective, we observed that lower latitudes present an increasing probability ratio (PR) of drought, indicating that there is a latitude gradient. An increased frequency of compound hot and dry events has been investigated by Geirinhas et al. (2021), in the states of São Paulo (SP), Rio de Janeiro (RJ) and Minas Gerais (MG), our results corroborate that this increased frequency can be attributed to climate change, however, this statement must be carefully examined in future studies.

The projections indicate an increasing risk of drought using a meteorological indicator, while the propagation of meteorological drought to other types of drought such as hydrological is highly correlated (Cuartas et al., 2022), factors such as increased water demand, deforestation, increasing the number of dams and dam capacity affect the impacts of meteorological droughts. We also acknowledge biases related to rainfall patterns in Brazil that are influenced by major climate drivers such as El Nino (Awange et al., 2016). In the literature, CMIP6 models have been documented to have a bias in the simulation of tropical sea surface temperature (SST), which can influence the accuracy of the simulation of indices such as El Niño-Southern Oscillation (ENSO) (Chiang et al., 2021; McKenna et al., 2020). Another source of precipitation error, which affects mainly the tropics, is the model representation of double-intertropical convergence zone (ITCZ), which has an influence in heavy rainfall (Tian e Dong, 2020).

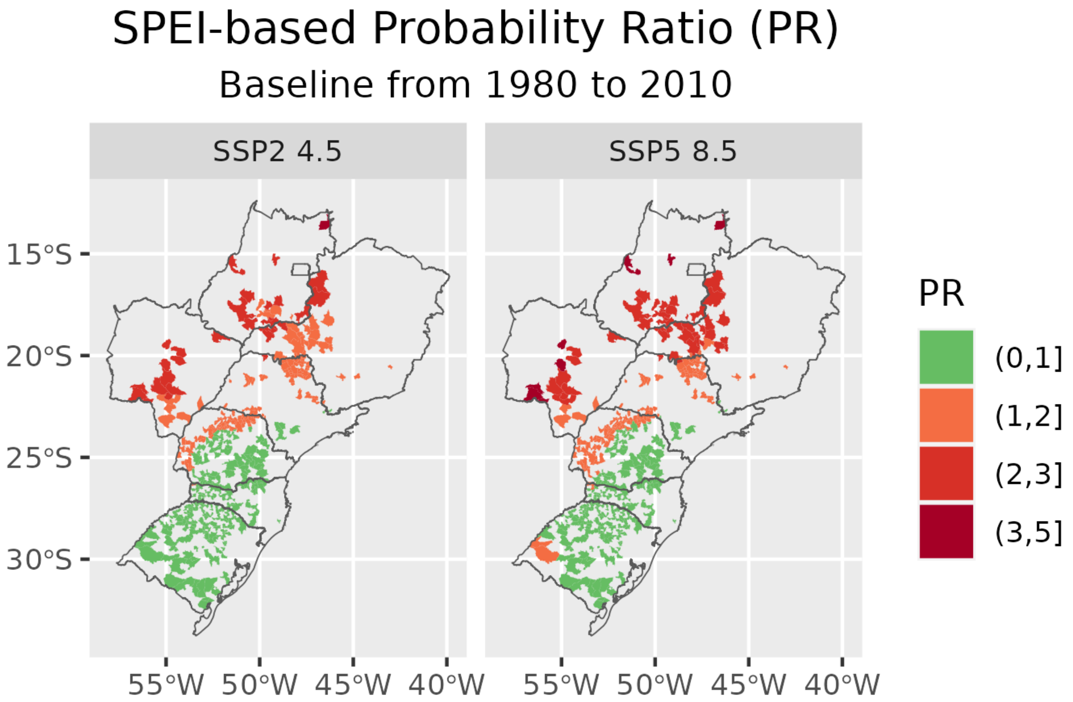


**Figure 4.6.** Map of the Probability Ratio (PR) $PR = \frac{P_{fut}}{P_0}$ of each municipality which is calculated by designating 3-month SPEI dips less than -1.5 as drought events, comparing historic SPEI $P_0$ to (a) moderate emission scenario (SSP2 4.5) and (b) high emission scenario (SSP5 8.5). Values above 1 indicate higher risks of drought events in forced conditions, while values below 1 indicate lower risks of drought as compared to observed values. The models used were K.ACE, UKESM1.0, and HadGEM3.GC3.1.

The results of the trend analysis of three extreme climate indicators for excessive precipitation (R20mm), high temperatures (TXX), and low temperatures (TNN) are shown in Fig. 4.7 for the period 2015–2100 relative to the baseline period (1980–2010). We highlight that the indicators showed statistically significant trends with $p < 0.05$. The R20mm, as defined in this study, is the monthly number of days when precipitation is greater than 20mm. The spatial patterns of R20mm indicate an increased impact of heavy rainfall affecting south Brazil, with a slope ranging from 0.1 to 0.2 days/month/decade. This is in agreement with the historical trends of extreme precipitation observed between 1960 and 2000, which was particularly observed in southern Brazil, Paraguay, Uruguay, and northeast Argentina (Haylock et al., 2006). Heavy rainfall is an important trigger for flash floods in various regions of Brazil (Marengo et al., 2023) and this represents a great threat to people living in vulnerable areas. Heavy rainfall can cause severe losses and damage to farmers and crop yields (Nóia Júnior et al., 2023), although in Chapter 3 we observed that the extreme wet growing season was relevant for soybean prediction losses, our crop loss prediction model was able to detect the impact of heavy rainfall on crop yields. This can represent an important underestimation in future scenarios, and more studies are needed on the impact of excessive wetness as climate change can accelerate these conditions.

Fig. 4.7 also illustrates the regional and spatial changes of the high and low temperature indices. Both low temperatures and high temperatures present an upward trajectory. Widespread upward trends in maximum and minimum temperatures were observed by Regoto et al. (2021) from 1961 to 2018 using ground-based temperature measurements. Increasing temperatures are attributed to climate change; these trends have been confirmed by Avila-Diaz et al. (2020) using both ground-based observations, reanalysis data, and global climate model simulations.

In general, the results of the analysis of extreme changes indicate a worsening of the climate conditions for agriculture with respect to (i) increasing severity of droughts in lower latitudes of Brazil, (ii) increasing in high temperatures, (iii) increasing in low temperatures, and (iv) increasing heavy rains especially in southern Brazil. This impact can be extended to other sectors of society such as increased respiratory and heart disease, wildfires, water shortages, floods, landslides, pests, and others (Masson-Delmotte et al., 2022). Although the results presented in this section demonstrate a worrying scenario, the values should not be regarded as exact indicators of the increase in the severity of extremes. However, these values are important for detecting areas that are projected to be more affected by extreme events and, therefore, support decisions such as how to support areas after disaster and how to price insurance to better protect insurance companies and policyholders (Stott, 2016). Global model simulations forced with climate change scenarios corroborate that these trends can be attributed to global warming. Although the attribution of increased climate extremes to climate change is a complicated topic and requires parsimony, denying it is no longer accepted in the impact studies community (of Sciences Engineering et al., 2016).

Although the models were evaluated for different regions, the spatial dependence of changes in climate extremes was not fully explored. The non-stationarity across space was proven to be relevant for extreme events statistics by Einmahl et al. (2022), and the impacts of spatially compounding extremes have been studies by Brunner et al. (2020).

### 4.3.3 Effectiveness of multi-hazard index insurance

In this section, we focus on evaluating how to reduce the risk of climate-related financial losses using index insurance. Fig. 4.2 summarizes the risk reduction potential (RRP) for each random forest model at the state level. Taking into account the average of all models, for soybeans, the RRP was on average 16 % in Rio Grande do Sul (RS), 4 % in Santa Catarina (SC) and Paraná (PR), 2.5 % in São Paulo (SP) and Mato Grosso do Sul (MS), and less than 2 % in Minas Gerais (MG) and Goiás (GO). For maize, the RRP was 5 % for PR, less than 1% for SP, 12.5 % for MS and 2.5 % for MG and GO. In general, our results for RS were compatible with the reduction of the rainfall-based weather derivative tested by Raucci et al. (2019a) in five municipalities.

For evaluating the effectiveness of multi-hazard index insurance, we performed a risk reduction potential analysis with an out-of-bag experiment. Our results showed that, in most cases, the risk reduction potential of the complex model improved significantly. This indicates that when climate extremes are taken into consideration using a complex model and a machine learning technique, farmers will be more protected against financial losses caused by extremes; therefore, the complex is the best suited model for risk protection. The results are robust and reliable, which is evidenced by the fact that the improvement is statistically significant $p < 0.01$.

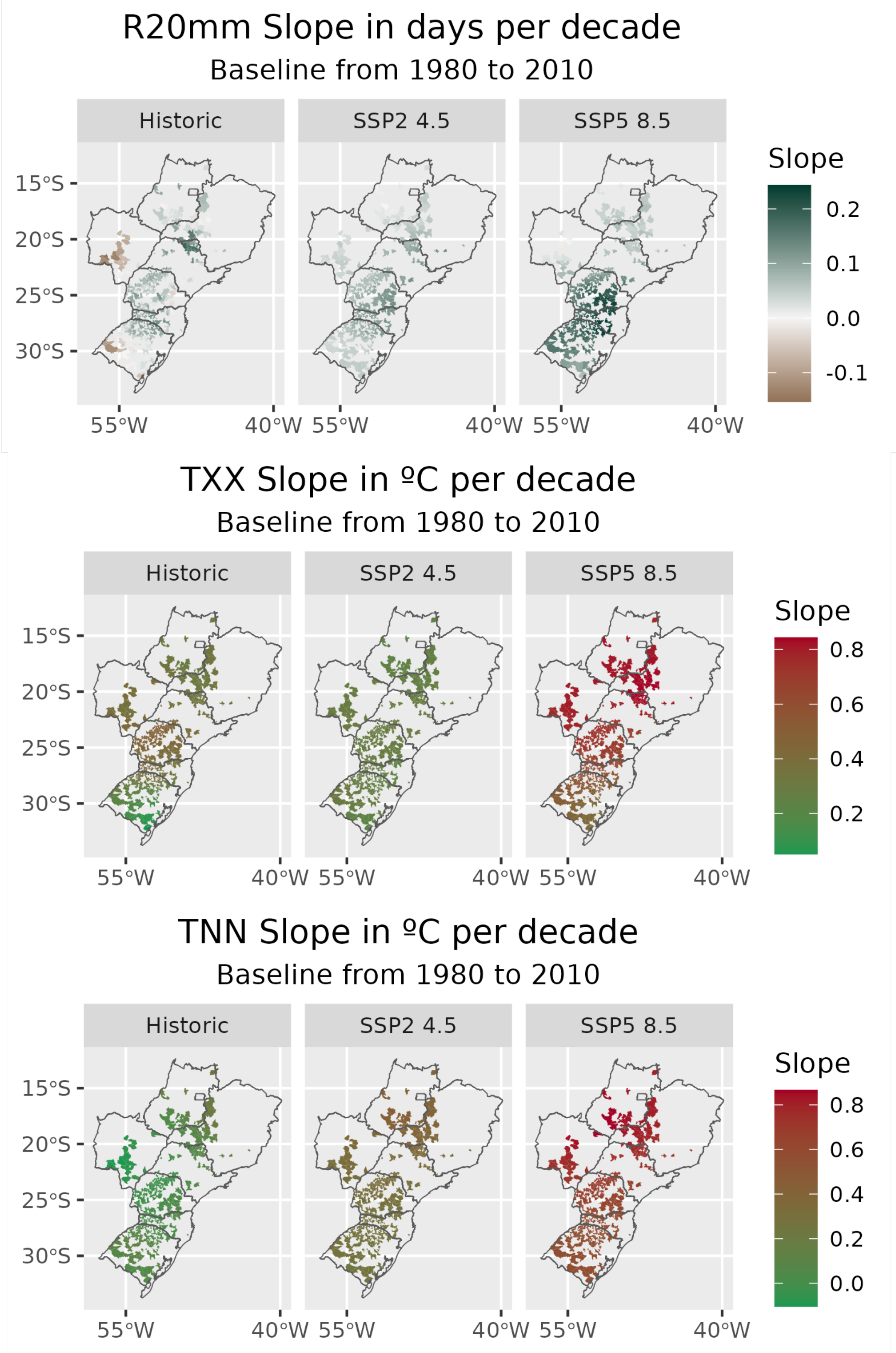


**Figure 4.7.** Spatial distribution of trends in climate extreme indices of at monthly time scaale for the historic period (1980 to 2010), and future period (2015 to 2100) for moderate (SSP2 4.5) and high emission (SSP5 8.5) scenarios: monthly count of days when PRCP 20mm (R20mm), monthly maximum value of daily maximum temperature (TXX), and Monthly minimum value of daily minimum temperature (TNN). The models used were K.ACE, UKESM1.0, and HadGEM3.GC3.1.

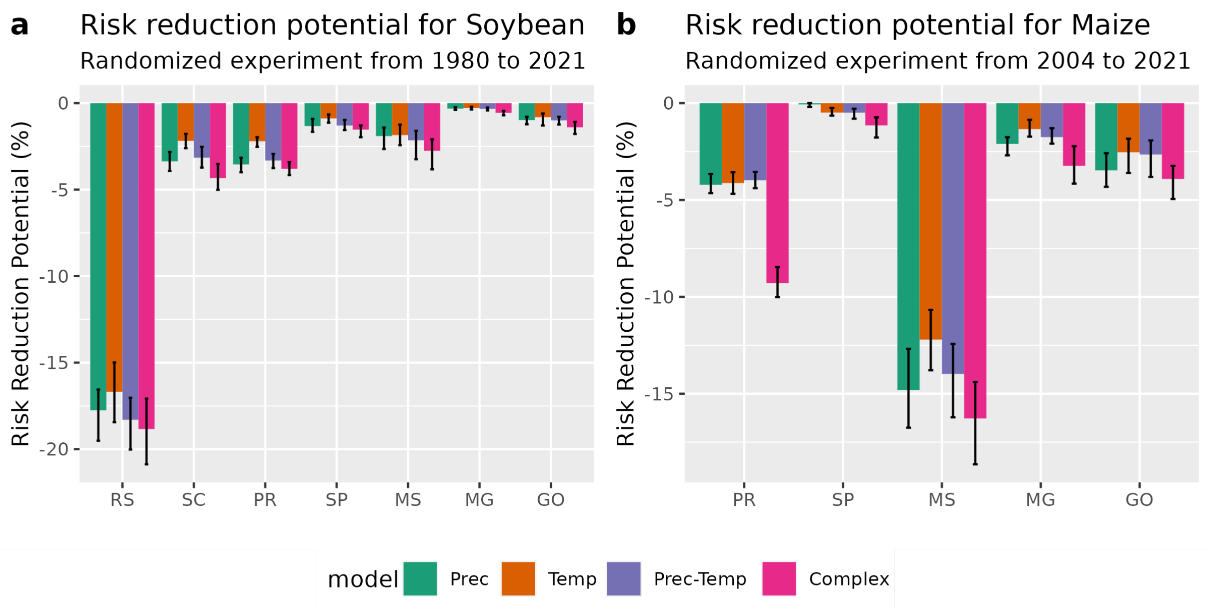


**Figure 4.8.** Analysis of Risk Reduction Potential for Soybean (a) and Maize (b) in Response to Four Policy Scenarios: precipitation means (Prec), temperature means (Temp), coupled precipitation and temperature means (Prec-Temp), and means and extreme events (Complex)

The findings of this study have significant implications for risk management, particularly when it comes to climate change and extreme events. Incorporating machine learning and climate-extreme indices into the design and implementation of index insurances appears to be a promising avenue for improving their effectiveness. Further research and validation of these findings would be valuable, as they have the potential to inform and potentially transform the insurance industry's approach to mitigating risks associated with climate-related events. In Chapter 3, we developed an analysis of climatic impact-drivers (CID) and defined the most important drivers of losses, their thresholds, and interactions. We acknowledge that these results can be used in the index insurance design process and should be considered in future work.

The main goal of this research was to understand the role of multi-hazard risk in insurance design and not to select the best machine learning model for index insurance or to compare it with traditional models. In fact, machine learning models have been proven to outperform traditional regression models for rainfall index insurance for floods (Cesarini et al., 2021; Hernández-Rojas et al., 2023). However, we acknowledge that one of the main underlying principles of index insurance is transparency (Skees et al., 1997), and machine learning is not transparent, although some explainability was provided in Chapter 3, the level of trust that policyholders could have on this type of contract is still a subject of study. It is important to state that the use of model interpretation proposed in Chapter 3 has been criticized by Rudin (2019), the author suggests that the solution for this problem is the use of models that are inherently interpretable. We argue that since we provided a method for selecting and evaluating many indices, the power of machine learning models could be used as a tool to improve traditional models. Another way forward is to use secure frameworks. An example is blockchain, which is a transaction database that allows one to make transactions encrypted (Nofer et al., 2017), therefore, the machine learning model that uses climate indices as inputs and predicts losses for index insurance could be underwritten as a so-called smart contract (Xiong et al., 2020; Kshetri, 2021).

### 4.3.4 Increased pressure in premiums

In Fig. 4.9, we illustrate the additional projected impact of climate change scenarios on soybeans and maize. Using historic simulations of global models (1980-2010) can be used to check the reliability of the results. For soybeans, the average yield considering all 452 municipalities is 3.32 ton/ha. When comparing with the average yields, the risk premiums estimated using the four models, we obtained the following results, the complex model was 2 % (0.07 ton/ha), prec 4 % (0.12 ton/ha), prec-temp 2 % (0.1 ton/ha), and temp 2 % (0.08 ton/ha). For maize, the average yield considering all 216 municipalities is 4.55 ton/ha. Compared to average yields, risk premiums are the following, for complex 10 % (0.5 ton/ha), for prec 16 % (0.75 ton/ha), prec-temp 13 % (0.6 ton/ha), and temp 16 % (0.75 ton/ha).

In terms of the impact of climate change on the two crops studies in this work, we highlight the complex model for soybeans and the temperature model for maize. For soybeans, our results do not show significant differences between moderate and high-emission scenarios; however, for the complex model, we observed increased pressure on crop yield risk premiums. The complex model incorporates means and extreme indices, indicating and increasing 26.5 % in risk premiums over a 3 decades horizon. For maize, on the other hand, we observed a dominance of temperature means on crop yields, increasing risk premiums in 24.6 %, also in a horizon of 3 decades.

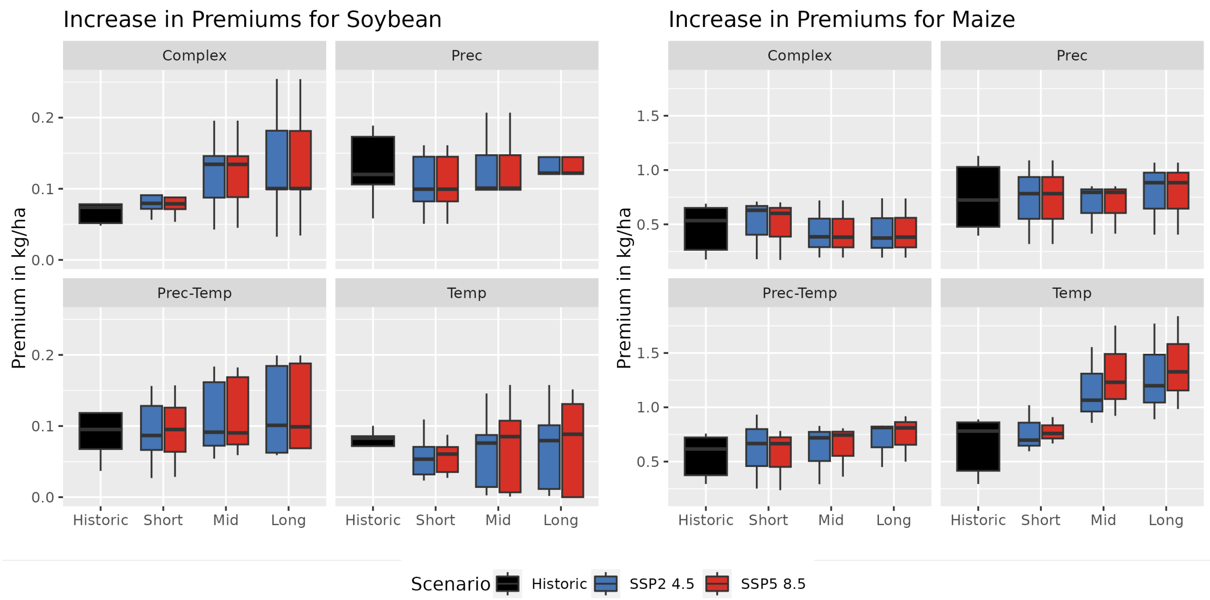


**Figure 4.9.** Projected additional impact of future global warming scenarios on Soybean and Maize due to increased precipitation means (Prec), temperature means (Temp), coupled precipitation and temperature means (Prec-Temp), and means and extreme events (Complex). Ensemble additional impact of climate change on Soybean as a percentage of mean of historic premiums (1980-2010), over short-term (2015-2044), mid-term (2045-2074) and long-term(2075-2100) under moderate-emissions scenario (SSP2-4.5) and high-emission scenario (SSP5-8.5)

Here, we presented the impacts of projected climate extremes; however, there are some important considerations that must be taken into consideration. Hazard analysis is based on historical data and matching with observed loss events for a rather limited amount of data. Risk analysis often requires long historical data to assess the probability of natural hazards and their impacts. Moreover, given the high amount of indices utilized, evaluating the statistics and probability of each one, as well their interaction, becomes unpractical. We try to overcome this limitation by adjusting the loss data to a probability distribution model within the impact model. This impact model is based on a random forest algorithm that evaluated the interaction between all the variables and makes it possible to evaluate the importance of the variable and the interaction among the variables.

The impact model also relies on the limited number of observed losses resulting from climate-related events. Linking extremes to losses is a key component to evaluating risk in agricultural systems. This includes taken into consideration monetary, non-monetary losses, direct and indirect impacts. This work provided insight into direct monetary losses to agricultural production, indicating impacts on risk management strategies for crop yield losses. Damages to properties, loss of livestock, changes in the market value of crops, and increased operational costs are also affected by the increase in climate extremes. In addition, other staple foods can be affected by climate change, such as rice, wheat, and fruits. Take into account other factors, such as access to labor, energy, and modes of production.

In addition to monetary losses, nonmonetary losses such as environmental degradation, health and population well-being, social disruption, food security, and adaptive capacity. The vulnerability of crops is based on the observed relationships that were obtained from historical data, and generating this information is a complex process that must require further analysis and expert adjustment. By using a fixed definition of vulnerability, we identified how increasing climate pressure can impact farmers, insurers, and governments; however, we acknowledge that this can be dynamic and changing over time when adaptation strategies for climate change are adopted. The exposure is also subject to dynamic factors. Observed trends have shown that cropland areas have increased in Brazil, which translates into increased exposure to climate. Changes in cropland area can also be affected.

In this work, we conceptualize that insurance can help farmers adapt to climate change, as insurance will provide additional protection against adverse climate conditions. However, there is a discussion on the problem of maladaptation. According to Wang et al. (2021), crop insurance has a negative impact on mean yields under climate change, and since farmers are protected, a larger area is exposed to climate extremes. The authors conclude that traditional crop insurance contributes as a disincentive to adaptation to climate change in agriculture. The same statement is supported by other observational works that integrate social, economical and ecological information to analyze insurance data (O'Hare et al., 2016; Müller et al., 2017). The aforementioned sources of maladaptation are closely related to land use planning and crop diversification. This means that a new paradox emerges, i.e., on one hand farmers need to safeguard income under dynamic risk (increasing risk), on the other hand, farmers must continue adopting adaptation measures, which also require investment. In this thesis, we presented an index insurance design for extreme climate that considered hazard identification, vulnerability assessment, and risk pricing. According to Reckien et al. (2023), the following steps must be incorporated in the insurance design to improve the index insurance as an adaptation measure: (i) ecosystem and ecosystem services, (ii) climate system (GHG emissions)*, (iii) social system, (iv) low-income populations, (v) women and girls, (vi) marginalized ethnic groups. Among these measures, we only included the second one. This indicates that the model presented in this work has many research opportunities that require more studies and improvements.

Even tough there are some limitations, identifying and quantifying changes in risk and risk reduction measures plays a critical role for demonstrating the importance of these measures and advancing climate science. Since this study involves evaluating the long-term impact of climate extremes, we were able to shed some light on the complexity and limitations of climate impact studies, including data collection, modeling, and assessment when studying climate resilience and insurance in agriculture, especially in the context of a changing climate.

## 4.4 Conclusion

In this study, we have analyzed the dynamics of multi-hazard risk in index insurance design and evaluation in the agricultural context of Brazil. Our approach involved considering the production of 452 municipalities (including the soybean producing municipalities and the 216 maize producing municipalities) under the climate projections of CMIP6 for both moderate and high emissions scenarios.

We evaluated multi-hazards using observed ERA5 data from 1980 to 2022 for soybeans and from 2003 to 2022 for maize considering precipitation means (Prec), temperature means (Temp), combination of precipitation and temperature means (Prec-Temp), and a combination of means and extremes (Complex) in a random forest modeling framework for predicting crop yield losses that was presented in Chapter 3. In particular, we have extended the applicability of this model to insurance policy frameworks, providing a means of quantifying and proposing mitigation measures for the financial risks associated with climate-induced agricultural losses. We observed that the complex model improved the risk reduction potential for soybeans and maize, indicating that the addition of climate extreme indicators improves the financial resilience of farmers.

To evaluate dynamic risk, we considered 19 global climate models from CLimate Models for BRAzil (CLIMBRA) (Ballarin et al., 2023). We proposed a selection model based on the inter-annual variability of precipitation and temperature, and we concluded that UKESM1.0, K.ACE and HadGEM3.GC3.1 were the most suited models for Brazil. The approach presented here aims to ensure that the selected global climate models represent changes in average and extreme climatic conditions well, but at the same time the models must have a reasonable ability to simulate past climate, with a focus on seasonality of temperature and precipitation, which is fundamental for determining agroclimatic zones.

The climate models selected for SSP 2 4.5 (moderate) and SSP 5 8.5 (high) show that there are critical trends in the averages and extremes. Projections of mean air temperature indicate an increase ranging from 2.5 °C to 6.3 °C between baseline (1980-2010) and future projections (2015-2100). Droughts (SPEI less than -1.5) are projected to increase 2 to 5 times the probability; however, they are located in the northwest of Paraná and in the states of São

Paulo, Mato Grosso do Sul, Minas Gerais and Goias, which indicates stronger warming at lower latitudes. This is corroborated by analysis of high and low temperature indices that represent an upward trajectory from 0.2 to 0.4 °C/month/decade under moderate emissions and 0.6 to 0.8 °C/month/decade under a high emission scenario. The trends in heavy rainfall also represent an upward trajectory of up to 0.2 days/month/decade that especially affects southern Brazil.

The dynamics of risk represents a potential increase in pressure for the insurance market. On average, risk premiums for soybeans were more sensitive to complex models, and projected climate extremes represent an increase in 26.5 % premiums. For maize, risk premiums were more sensitive to temperature models and climate projections represented an increase of 24 6 % in premiums.

In conclusion, this study demonstrates the importance of considering climate extremes in the design of index insurance to increase the risk reduction potential of financial losses in food production.

## References


Akaike, H. (1974). A new look at the statistical model identification problem. *IEEE Trans. Autom. Control*, 19:716.

Arvor, D., Dubreuil, V., Ronchail, J., Simões, M., e Funatsu, B. M. (2014). Spatial patterns of rainfall regimes related to levels of double cropping agriculture systems in mato grosso (brazil). *International Journal of Climatology*, 34(8):2622–2633.

Avila-Diaz, A., Benezoli, V., Justino, F., Torres, R., e Wilson, A. (2020). Assessing current and future trends of climate extremes across brazil based on reanalyses and earth system model projections. *Climate Dynamics*, 55(5-6):1403–1426.

Awange, J. L., Mpelasoka, F., e Goncalves, R. M. (2016). When every drop counts: Analysis of droughts in brazil for the 1901-2013 period. *Science of the Total Environment*, 566:1472–1488.

Ballarin, A. S., Sone, J. S., Gesualdo, G. C., Schwamback, D., Reis, A., Almagro, A., e Wendland, E. C. (2023). Climbra-climate change dataset for brazil. *Scientific Data*, 10(1):47.

Barrett, C. B. (2010). Measuring food insecurity. *Science*, 327(5967):825–828.

Battisti, R. e Sentelhas, P. C. (2014). New agroclimatic approach for soybean sowing dates recommendation: A case study. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 18:1149–1156.

Becker, W. R., Richetti, J., Mercante, E., Esquerdo, J. C. D. M., Silva Junior, C. A. d., Paludo, A., e Johann, J. A. (2020). Calendário agrícola de soja e milho baseado em imagens de satélite de moderada resolução para o sul do brasil. *Semina Ciências Agrárias*, 41(05, supl. 01):2419–2428.

Benso, M. R., Gesualdo, G. C., Silva, R. F., Silva, G. J., Castillo Rápalo, L. M., Navarro, F. A. R., Marques, P. A. A., Marengo, J. A., e Mendiondo, E. M. (2023). Design and evaluation of weather index insurance for multi-hazard resilience and food insecurity. *Natural Hazards and Earth System Sciences*, 23(4):1335–1354.

Black, F. e Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654.

Brunner, M. I., Gilleland, E., Wood, A., Swain, D. L., e Clark, M. (2020). Spatial dependence of floods shaped by spatiotemporal variations in meteorological and land-surface processes. *Geophysical Research Letters*, 47(13):e2020GL088000.

Bullock, J. M., Dhanjal-Adams, K. L., Milne, A., Oliver, T. H., Todman, L. C., Whitmore, A. P., e Pywell, R. F. (2017). Resilience and food security: rethinking an ecological concept. *Journal of Ecology*, 105(4):880–884.

Cesarini, L., Figueiredo, R., Monteleone, B., e Martina, M. L. (2021). The potential of machine learning for weather index insurance. *Natural Hazards and Earth System Sciences*, 21(8):2379–2405.

Chiang, F., Mazdiyasni, O., e AghaKouchak, A. (2021). Evidence of anthropogenic impacts on global drought frequency, duration, and intensity. *Nature communications*, 12(1):2754.

Cottrell, R. S., Nash, K. L., Halpern, B. S., Remenyi, T. A., Corney, S. P., Fleming, A., Fulton, E. A., Hornborg, S., Johne, A., Watson, R. A., et al. (2019). Food production shocks across land and sea. *Nature Sustainability*, 2(2):130–137.

Cuartas, L. A., Cunha, A. P. M. d. A., Alves, J. A., Parra, L. M. P., Deusdará-Leal, K., Costa, L. C. O., Molina, R. D., Amore, D., Broedel, E., Seluchi, M. E., et al. (2022). Recent hydrological droughts in brazil and their impact on hydropower generation. *Water*, 14(4):601.

Cunha, A. P. M., Zeri, M., Deusdará Leal, K., Costa, L., Cuartas, L. A., Marengo, J. A., Tomasella, J., Vieira, R. M., Barbosa, A. A., Cunningham, C., et al. (2019). Extreme drought events over brazil from 2011 to 2019. *Atmosphere*, 10(11):642.

de Souza Nóia, R. e Sentelhas, P. C. (2020). Yield gap of the double-crop system of main-season soybean with off-season maize in brazil. *Crop and Pasture Science*, 71(5):445–458.

Delignette-Muller, M. L. e Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, 64(4):1–34.

Deusdará-Leal, K., Cuartas, L., Zhang, R., Mohor, G., Carvalho, L., Nobre, C., Mendiondo, E., Broedel, E., Seluchi, M., e Alvalá, R. (2019). Implication of the new operation rules for cantareira system: Re-reading of the 2014/2015 water crisis. *Water Resour. Res.*

Dias, H. B. e Sentelhas, P. C. (2017). Evaluation of three sugarcane simulation models and their ensemble for yield estimation in commercially managed fields. *Field Crops Research*, 213:174–185.

Droogers, P. e Allen, R. G. (2002). Estimating reference evapotranspiration under inaccurate data conditions. *Irrigation and drainage systems*, 16:33–45.

Einmahl, J. H., Ferreira, A., de Haan, L., Neves, C., e Zhou, C. (2022). Spatial dependence and space–time trend in extreme events. *The Annals of Statistics*, 50(1):30–52.

Franch, B., Cintas, J., Becker-Reshef, I., Sanchez-Torres, M. J., Roger, J., Skakun, S., Sobrino, J. A., Van Tricht, K., Degerickx, J., Gilliams, S., et al. (2022). Global crop calendars of maize and wheat in the framework of the worldcereal project. *GIScience & Remote Sensing*, 59(1):885–913.

Geirinhas, J. L., Russo, A., Libonati, R., Sousa, P. M., Miralles, D. G., e Trigo, R. M. (2021). Recent increasing frequency of compound summer drought and heatwaves in southeast brazil. *Environmental Research Letters*, 16(3):034036.

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., e Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. *science*, 327(5967):812–818.

Hallegatte, S. (2014). Economic resilience: definition and measurement. *World Bank Policy Research Working Paper*.

Hasegawa, T., Sakurai, G., Fujimori, S., Takahashi, K., Hijioka, Y., e Masui, T. (2021). Extreme climate events increase risk of global food insecurity and adaptation needs. *Nature Food*, 2(8):587–595.

Haylock, M. R., Peterson, T. C., Alves, L. M., Ambrizzi, T., Anunciação, Y., Báez, J., Barros, V. R., Berlato, M., Bidegain, M., Coronel, G., et al. (2006). Trends in total and extreme south american rainfall in 1960–2000 and links with sea surface temperature. *Journal of climate*, 19(8):1490–1512.

Heino, M., Kinnunen, P., Anderson, W., Ray, D. K., Puma, M. J., Varis, O., Siebert, S., e Kummu, M. (2023). Increased probability of hot and dry weather extremes during the growing season threatens global crop yields. *Scientific Reports*, 13(1):3583.

Hernández-Rojas, L. F., Abrego-Perez, A. L., Lozano Martínez, F. E., Valencia-Arboleda, C. F., Diaz-Jimenez, M. C., Pacheco-Carvajal, N., e García-Cárdenas, J. J. (2023). The role of data-driven methodologies in weather index insurance. *Applied Sciences*, 13(8):4785.

Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual review of ecology and systematics*, 4(1):1–23.

IPCC (2022). *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge.

Júnior, R. d. S. N. e Sentelhas, P. C. (2019). Soybean-maize succession in brazil: Impacts of sowing dates on climate variability, yields and economic profitability. *European Journal of Agronomy*, 103:140–151.

Kim, W., Iizumi, T., e Nishimori, M. (2019). Global patterns of crop production losses associated with droughts from 1983 to 2009. *Journal of Applied Meteorology and Climatology*, 58(6):1233–1244.

Kshetri, N. (2021). Blockchain-based smart contracts to provide crop insurance for smallholder farmers in developing countries. *IT Professional*, 23(6):58–61.

Lam, M. R., Matanó, A., Van Loon, A. F., Odongo, R., Teklesadik, A. D., Wamucii, C. N., van den Homberg, M. J., Waruru, S., e Teuling, A. J. (2022). Linking reported drought impacts with drought indices, water scarcity, and aridity: the case of kenya. *EGUsphere*, pages 1–25.

Lesk, C., Coffel, E., Winter, J., Ray, D., Zscheischler, J., Seneviratne, S. I., e Horton, R. (2021). Stronger temperature–moisture couplings exacerbate the impact of climate warming on global crop yields. *Nature Food*, 2(9):683–691.

Marengo, J. A., Cunha, A. P., Cuartas, L. A., Deusdará Leal, K. R., Broedel, E., Seluchi, M. E., Michelin, C. M., De Praga Baião, C. F., Chuchón Angulo, E., Almeida, E. K., et al. (2021). Extreme drought in the brazilian pantanal in 2019–2020: characterization, causes, and impacts. *Frontiers in Water*, 3:639204.

Marengo, J. A., Seluchi, M. E., Cunha, A. P., Cuartas, L. A., Goncalves, D., Sperling, V. B., Ramos, A. M., Dolif, G., Saito, S., Bender, F., et al. (2023). Heavy rainfall associated with floods in southeastern brazil in november–december 2021. *Natural Hazards*, pages 1–28.

Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P. R., et al. (2022). *Global Warming of 1.5 C: IPCC special report on impacts of global warming of 1.5 C above pre-industrial levels in context of strengthening response to climate change, sustainable development, and efforts to eradicate poverty.* Cambridge University Press.

Matthews, H. D. e Wynes, S. (2022). Current global efforts are insufficient to limit warming to 1.5 c. *Science*, 376(6600):1404–1409.

McKee, T. B. (1995). Drought monitoring with multiple time scales. In *Proceedings of 9th Conference on Applied Climatology, Boston, 1995.*

McKee, T. B., Doesken, N. J., Kleist, J., et al. (1993). The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology, California, 1993.*

McKenna, S., Santoso, A., Gupta, A. S., Taschetto, A. S., e Cai, W. (2020). Indian ocean dipole in cmip5 and cmip6: characteristics, biases, and links to enso. *Scientific reports*, 10(1):11500.

Mechler, R. (2016). Reviewing estimates of the economic efficiency of disaster risk management: opportunities and limitations of using risk-based cost–benefit analysis. *Natural Hazards*, 81:2121–2147.

Minoli, S. (2023). *cropCalendars: Compute rule-base crop sowing and harvest dates.* R package version 0.1.0.

Minoli, S., Egli, D. B., Rolinski, S., e Müller, C. (2019). Modelling cropping periods of grain crops at the global scale. *Global and Planetary Change*, 174:35–46.

Moriondo, M., Giannakopoulos, C., e Bindi, M. (2011). Climate change impact assessment: the role of climate extremes in crop yield simulation. *Climatic change*, 104(3-4):679–701.

Müller, B., Johnson, L., e Kreuer, D. (2017). Maladaptive outcomes of climate insurance in agriculture. *Global Environmental Change*, 46:23–33.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., et al. (2021). Era5-land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13(9):4349–4383.

Nofer, M., Gomber, P., Hinz, O., e Schiereck, D. (2017). Blockchain. *Business & Information Systems Engineering*, 59:183–187.

Nóia Júnior, R. d. S., Deswarte, J.-C., Cohan, J.-P., Martre, P., van Der Velde, M., Lecerf, R., Webber, H., Ewert, F., Ruane, A. C., Slafer, G. A., et al. (2023). The extreme 2016 wheat yield failure in france. *Global Change Biology*, 29(11):3130–3146.

of Sciences Engineering, N. A., Medicine, et al. (2016). *Attribution of extreme weather events in the context of climate change*. National Academies Press.

O'Hare, P., White, I., e Connelly, A. (2016). Insurance as maladaptation: Resilience and the 'business as usual'paradox. *Environment and Planning C: Government and Policy*, 34(6):1175–1193.

Oligini, K. F., Batista, V. V., Lemes, E. S., Silva, É. J. d., e Adami, P. F. (2021). Sowing date and maturity groups on the economic feasibility of soybean-maize double summer crop system. *Revista de Economia e Sociologia Rural*, 59.

Pires, G. F., Abrahão, G. M., Brumatti, L. M., Oliveira, L. J., Costa, M. H., Liddicoat, S., Kato, E., e Ladle, R. J. (2016). Increased climate risk in brazilian double cropping agriculture systems: Implications for land use in northern brazil. *Agricultural and forest meteorology*, 228:286–298.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raucci, G. L., Lanna, R., da Silveira, F., e Capitani, D. H. D. (2019a). Development of weather derivatives: evidence from the brazilian soybean market. *Italian Review of Agricultural Economics*, 74(2):17–28.

Reckien, D., Magnan, A. K., Singh, C., Lukas-Sithole, M., Orlove, B., Schipper, E. L. F., e Coughlan de Perez, E. (2023). Navigating the continuum between adaptation and maladaptation. *Nature Climate Change*, pages 1–12.

Regoto, P., Dereczynski, C., Chou, S. C., e Bazzanela, A. C. (2021). Observed changes in air temperature and precipitation extremes over brazil. *International Journal of Climatology*, 41(11):5125–5142.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5):206–215.

Schulzweida, U. (2023). Cdo user guide.

Simonovic, S. P. e Peck, A. (2013). Dynamic resilience to climate change caused natural disasters in coastal megacities quantification framework. *British Journal of Environment and Climate Change*, 3(3):378.

Skees, J. R., Black, J. R., e Barnett, B. J. (1997). Designing and rating an area yield crop insurance contract. *American journal of agricultural economics*, 79(2):430–438.

Stott, P. (2016). How climate change affects extreme weather events. *Science*, 352(6293):1517–1518.

Surminski, S., Bouwer, L. M., e Linnerooth-Bayer, J. (2016). How insurance can support climate resilience. *Nature Climate Change*, 6(4):333–334.

Tian, B. e Dong, X. (2020). The double-itcz bias in cmip3, cmip5, and cmip6 models based on annual mean precipitation. *Geophysical Research Letters*, 47(8):e2020GL087232.

Van Wart, J., van Bussel, L. G., Wolf, J., Licker, R., Grassini, P., Nelson, A., Boogaard, H., Gerber, J., Mueller, N. D., Claessens, L., et al. (2013). Use of agro-climatic zones to upscale simulated crop yield potential. *Field crops research*, 143:44–55.

Velez, D. R., White, B. C., Motsinger, A. A., Bush, W. S., Ritchie, M. D., Williams, S. M., e Moore, J. H. (2007). A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction. *Genetic Epidemiology: the Official Publication of the International Genetic Epidemiology Society*, 31(4):306–315.

Vicente-Serrano, S. M., Beguería, S., e López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of climate*, 23(7):1696–1718.

Waha, K., Van Bussel, L., Müller, C., e Bondeau, A. (2012). Climate-driven simulation of global crop sowing dates. *Global Ecology and Biogeography*, 21(2):247–259.

Wang, R., Rejesus, R. M., e Aglasan, S. (2021). Warming temperatures, yield risk and crop insurance participation. *European Review of Agricultural Economics*, 48(5):1109–1131.

Xiong, H., Dalhaus, T., Wang, P., e Huang, J. (2020). Blockchain technology for agriculture: applications and rationale. *frontiers in Blockchain*, 3:7.

Zampieri, M., Ceglar, A., Dentener, F., e Toreti, A. (2017). Wheat yield loss attributable to heat waves, drought and water excess at the global, national and subnational scales. *Environmental Research Letters*, 12(6):064008.

# 5 CONCLUSIONS AND RECOMMENDATIONS

## 5.1 Conclusions

In a world in which the climatic conditions are rapidly changing and the boundaries of the planetary limit are being transgressed, resilience assumes a central position, as it addresses the interface between climate and food security. This interface entails the dual challenge of ensuring an adequate food supply for present and future generations while simultaneously mitigating the risk associated with climate extremes. In accordance with this problematic, the hypothesis of this thesis is that **the implementation of a multi-hazard risk index insurance strategy in Brazilian agro-producing regions can enhance the financial resilience of food production systems by mitigating the adverse effects of various climate extremes**. The findings of this research corroborate the aforementioned hypothesis, underscoring the need to improve the adoption of appropriate measures for effective climate risk management for food production as part of an integrated policy for climate change adaptation and mitigation plans. In this thesis, we evaluated multi-hazard risks include wet and dry and warm and cold conditions, among these risks, increasing extreme high temperatures in combination with dry conditions pose a severe threat for Brazil. In this context, index insurance shows the potential to effectively reduce financial vulnerability within agro-producing regions.

The first specific objective of this thesis was to **review and analyze index insurance design for multihazard risk in the context of food production and climate extremes**. In Chapter 2, we address this objective by conducting a systematic review of the literature to understand how to use multi-hazard risk for the design and evaluation of index insurance policies. Although acknowledging the importance of multiple sources of climate threats to food production, central attention has been paid in the literature to droughts, given their widespread impact on agriculture. We acknowledge the limited exploration of multihazard risk assessment. As response to this gap and the second specific objective of this thesis, which was to **propose a conceptual framework for multi-hazard risk transfer model for climate extremes**, we described a conceptual framework aggregating and contextualizing the literature providing examples of its use and how I can be applied. This conceptual framework also integrates methodologies that have been applied in the Water Adaptive Design and Innovation Laboratory (WADI-Lab) by Righetto et al. (2007); Guzmán et al. (2020); Mohor e Mendiondo (2017) and are being further applied to new studies in the context of insurance design for other applications, such as multi-hazard threats to water supply companies (da Silva et al., 2023a; Gesualdo et al., 2022), urban flooding (Navarro et al., 2021), multipurpose reservoirs, and others.

According to our systematic review of the literature, the selection and evaluation of multi-hazard risk for index insurance design requires more development. In light of this, we highlight the need for developing new tools for selecting climate information that is relevant for impact studies for food production tailoring indices for each region of Brazil. Thus, the second specific objective was to **build a learning model for assessing the impact of climate extremes on food production**. We answer this objective in **Chapter three** as we explored the multi-hazard risk assessment on food production proposing a random forest model coupled to an explanatory XGBoost model that allied indices that are relevant to capture extreme climate events such as droughts, heat waves, cold waves and excessive rainfall, soybean and maize second cycle crop yield data. Several datasets can be used for impact studies and we observed that municipal-level statistics from Brazilian Institute of Geography and Statistics and the Department (IBGE) of Rural Economics of Paraná (Deral) are highly correlated and are traditionally used for risk analysis. We also observed that the Global Dataset of Historical Yields, which is a gridded dataset, despite its weak correlation with yields from the municipal statistics, offers a good alternative for analyzing crop yield variability associated with climate. The highlight is the compound effect of moisture and temperatures in crop risk analysis of soybean-maize double-cropping systems; however, precipitation is the major driver of crop losses. Using an agnostic model evaluation approach, we unveiled the importance of each climate variable in the prediction of crop yields. Using this approach, we were able to identify thresholds for crop yield losses, for example, for Soybeans, the probability of crop yield losses is associated with precipitation accumulated in December lower than 80 mm. Monthly average temperatures in October and November higher than 25 °C are also an important driver of soybeans in Paraná. For maize second season, accumulated precipitation in May was below 100 mm in Paraná.

The fourth objective of this thesis was to **assess the impacts of climate change on food production and index insurance as a mitigation strategy**. In Chapter four, we offer a contribution to climate resilience studies. In this chapter, we presented a dynamic resilience approach to evaluate the impacts of climate change on the assessment of food risk. Our research was focused on extreme events that present changing conditions. First, we selected climate models downscaled to Brazil from the Climate Change Dataset for Brazil (CLIMBra). By using a concept of agro-climatic zones based on temperature and precipitation intra-annual variability, we selected the climate models that best simulated climate conditions for crop production according to types of seasonality. For Brazil, the recommended models are UKESM1.0, K.ACE, and HadGEM3.GC3.1. Using these models, we observed that the projected changes in temperature and precipitation point to the following trends: (i) increase in mean temperatures; (ii) decrease in low temperatures, which implies less risk of hails and frosts during maize second cycle growing season; (iii) increase in excessive rainfall, (iv) increase in drought frequency. These trends, or risk dynamics, are evaluated using a random forest model and used to predict future impacts on food production. We observed that climate risk in moderate and high-emission scenarios presents a great challenge for food production with an increasing risk of extreme losses due to increased climate extremes. In the context of changing risk, we propose an insurance mode to evaluate index insurance for historical and future scenarios. Our results highlight the impact of climate change on insurance pricing, indicating an increase of risk premium prices. However, in all scenarios, insurance reduced financial risks and improved income for farmers.

Finally, we conclude that this thesis met the general objective that was to **propose a multi-hazard risk transfer model as a tool to mitigate the impacts of climate extremes on food production under climate changes in Brazil**. In this thesis, we proposed a study to develop and evaluate multihazard index insurance policies, assessment of climate impact-drivers applied to food production, and promote climate resilience of food production under global warming scenarios. The findings presented in this thesis offer a scientific contribution to the fields of risk management and disaster risk reduction, as well as practical insights that can aid science-informed decision making and policy.

## 5.2 Recommendations

This thesis has several applications for risk management in agriculture, insurance policies, and climate resilience. The results obtained in this research offer guidelines for index insurance development and multi-hazard risk assessment; however, the complexity of risk and the interaction of different sectors were poorly explored. Further work should consider different risks and their complexity. In terms of applications for risk management in agriculture, the machine learning framework presented in Chapter 3 can be used to inform early warning systems by providing predictions on the severity of crop yield losses. Early warning systems can issue alerts that correspond to the severity of crop loss. However, this implication requires further development and study and might require the used of remotely sensed data to improve the spatial resolution of the alerts and the accuracy of the model. The crop risk analysis proposed in this thesis can also help to develop long-term adaptation plans and help identify and address regions that have different levels of climate resilience and those that are the most vulnerable to the impacts of climate change. Since we presented a data-driven approach, other risk-based alerts can be developed using the same modeling strategy generating data-driven methods for decision support tools.

Insurance, and especially index-based insurance, has the potential to mitigate the increasing risk of crop yield losses associated with climate change, however, the price of risk premiums can burden farmers. Climate change adjustment of risk premiums has potential implications for government insurance policies and should be integrated into programs such as the Brazilian Insurance Premium Subsidy Program (PSR).

Since some regions can face an increase in premiums more than others, it is strategic to invest in long-term plants to improve resilience. Further studies can be used to evaluate different strategies on a local scale. The insurance rating game proposed in Chapter 4 is an exercise that can be expanded for educational purposes since it follows the principles of serious games. This tool provides hands-on exercises for risk analysis. We provide a data set that encompasses 452 Brazilian municipalities and climate indices to characterize mean and extreme climate conditions. Other datasets can be explored in the context of a classroom for other types of risk analysis, such as hydrological risk. We highlight the use of the Catchments Attributes for Brazil (CABra) (Almagro et al., 2021) and Caravan (Kratzert et al., 2023). Educational games have been recommended for water resource management (Mariano e Alves, 2020), and the insurance model proposed in this thesis can be used in the context of learning concepts of risk management.

In conclusion, this thesis highlights the pressing need for addressing multi-hazard risk dynamics in index insurance design and evaluation. In this work, we demonstrate that index insurance has great potential for Brazil and indicated that, under the complexity of climate change scenarios, it becomes of paramount importance for safeguarding income for farmers mitigating the impact of extremes on food security.

## References

Almagro, A., Oliveira, P. T. S., Meira Neto, A. A., Roy, T., e Troch, P. (2021). Cabra: a novel large-sample dataset for brazilian catchments. *Hydrology and Earth System Sciences*, 25(6):3105–3135.

da Silva, G. J., Gesualdo, G. C., da Silva, P. G. C., Benso, M. R., Arias, D. A. G., Sabogal-Paz, L. P., Orsini, J. M., e Mendiondo, E. M. (2023a). Index insurance as a risk management strategy to hydrological extremes for water utilities — essopenarchive.org. `https://essopenarchive.org/doi/full/10.22541/essoar.167525153.37247927`. [Accessed 08-10-2023].

Gesualdo, G., Benso, M., Navarro, F., Rápalo, L., e Mendiondo, E. (2022). Index-based insurance to mitigate drought financial losses for water supply sector. Technical report, Copernicus Meetings.

Guzmán, D. A., Mohor, G. S., e Mendiondo, E. M. (2020). Multi-Year Index-Based Insurance for Adapting Water Utility Companies to Hydrological Drought: Case Study of a Water Supply System of the Sao Paulo Metropolitan Region, Brazil. *Water*, 12(11):2954.

Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., et al. (2023). Caravan-a global community dataset for large-sample hydrology. *Scientific Data*, 10(1):61.

Mariano, D. J. K. e Alves, C. d. M. A. (2020). The application of role-playing games and agent-based modelling to the collaborative water management in peri-urban communities. *RBRH*, 25.

Mohor, G. S. e Mendiondo, E. M. (2017). Economic indicators of hydrologic drought insurance under water demand and climate change scenarios in a Brazilian context. *Ecological Economics*, 140:66–78.

Navarro, F. A. R., Gesualdo, G. C., Ferreira, R. G., Rápalo, L. M. C., Benso, M. R., de Macedo, M. B., e Mendiondo, E. M. (2021). A novel multistage risk management applied to water-related disaster using diversity of measures: A theoretical approach. *Ecohydrology & Hydrobiology*, 21(3):443–453.

Righetto, J., Mendiondo, E., e Righetto, A. (2007). Modelo de Seguro para Riscos Hidrológicos. *Revista Brasileira de Recursos Hídricos*, 12(2):107–113.

# APPENDIX

## Appendix I

### Framework application: Illustrative example

An illustrative example was developed to demonstrate the multi-hazard risk path of the proposed framework. This covers all steps from problem definition and data collection to index calculation and loss evaluation for several cities and a specific crop. It is important to note that the methods used and the code created can be reused for different years, areas, countries, hazards, and crops.

This choice was motivated to prove the impact of selecting multi-hazard indices for designing weather index-based insurance for crop yields. More studies are needed to link crop yield to other aspects of food security, such as transportation, storage and retail.

In this example, we chose the 42 largest Soybean producing municipalities in the Brazilian state of Paraná for evaluation. According to Pereira et al. (2013), the study area is located in a region with low to very low geodiversity and low soil diversity. Fine-grained soil such as ultisols and latosols with high iron content are predominant in the region, followed by the presence of medium-grained soils such as cambisols (Bhering et al., 2009). These soils are generally classified as clayey and loamy, and their soil volumetric water content (VWC) tends to be high. According to Saxton e Rawls (2006), the Permanent Wilting Point (PWP) is, on average 22%, the field capacity (FC) is 37%, and the porosity/saturation is at 47%.

The production of soybeans represents an interesting object of study since it is an essential crop for oil and protein. Soybean has a significant economic impact in Brazil, wide geographical distribution, and a vulnerability to various hazards. Soybean crop yields in Paraná are mainly threatened by temperature variation, droughts, and excessive rainfall (da Silva et al., 2021).

We used 22 years of yield data of first-cycle soybean production from 1996/1997 to the 2019/2020 growing seasons. Crop data was retrieved from the official statistical yearbooks (Parana, 2021). The multi-hazard risk hypothesis was tested using the widely employed machine learning algorithm random forest (Breiman, 2001).

The conceptual framework illustrated in Figure 4 was applied to a case study for soybean production in South Brazil, following the methodology described in section 2.2. The main objective of this case study was to illustrate the main steps of the framework, focusing on multi-hazards risks (A2), testing the hypothesis of synergic interaction between hazards (B2), Stationary System State (C1), a loss estimated with index value (D2), single contracts (E1) and no risk pooling (F1). The following five key steps were used in the illustrative example:

**1st step - Hazard identification:** We selected thermal stress, drought, and excessive rainfall as the main threats to soybean production. The index selection was based on the indices found in the literature (Table 2.2) and the indices indicated by the CCl/CLIVAR/JCOMM Expert Team (ET) on Climate Change Detection and Indices (ETCCDI) (Peterson et al., 2001). The index selection focused on finding simple indicators based solely on precipitation and temperature. After an extensive examination, the following indices were considered: maximum daily rainfall event over the growing season (pmax), 3-month Standardized Precipitation Index (SPI), number of days where daily precipitation is higher than the 90th percentile over the growing season (TX90p);

**2nd step - Definition of loss thresholds:** Crop losses were chosen as the target variable because they can be used as a proxy for the impact of extreme weather occurrences. The crop yields were detrended following the linear procedure used in Bucheli et al. (2021) $\bar{y} = y_i + \beta \cdot (year_{end} - year_i)$, where $\bar{y}$ is the detrended crop yield series $y_i$ the raw crop yield data in the year i, $\beta$ is the linear regression coefficient of the equation $y_i = \alpha + \beta \cdot year_i$. The losses were then determined following the equation: $Loss = max(0, K - \bar{y}/K)$. The K variable is the crop yield threshold value. It can be understood as the threshold that divides unfavorable crop yields for farmers (values below K) and favorable crop yields (values above K);

**3rd step - Data clustering for evaluating the interaction between hazards:** The kmeans clustering method (MacQueen, 1967), a widely used clustering method, was implemented to understand the data better. The clustering was applied for four relevant variables: pmax, SPI, TX90p, and crop yield. The elbow method was used to define the optimal value of clusters (also referred to as hyperparameter $\kappa$). This is the most used method in the literature for defining  . The method was implemented in R Environment using the package stats (R Core Team, 2022);

**4th step - Crop loss prediction modeling:** Several models were tested. However, two crop loss prediction models were chosen to demonstrate the importance of multi-hazard risk modeling, following a regression model and using the random forest algorithm: (i) Multi-hazard model M1 drought and thermal stressed using SPI and TX90p as inputs (M1(SPI,TX90p)); and (ii) Multi-hazard model M2 excessive precipitation and thermal stressed model using SPI, TX90p, and pmax as inputs (M2(SPI,TX90p,pmax)). The multi-hazard model M1 was trained and validated using data from clusters 2 and 4 and the multi-hazard model M2 was trained and validated using data from clusters 6. The standard cross-validation method was applied, following the best practices for machine learning workflows presented in the literature. The models were built using the R-package randomForest (Liaw e Wiener, 2002);

**5th step - Risk pricing:** The risk analysis is performed to determine pure risk premiums using stochastic methods. Historical burn analysis was performed on detrended crop yields to determine reference pure risk premium values. Then, a stochastic analysis of premiums for multi-hazard models M1(SPI,TX90p) and M2(SPI,TX90p) was determined considering P = E[Loss]. The expectation of loss E[Loss] was determined using the generation of 50 synthetic scenarios of weather data. The synthetic weather data was simulated using a multi-site multi-variable (daily precipitation and temperature ) weather generator method. The stochastic simulation was performed using a wavelet-based algorithm that allows multi-site simulation. The simulation was implemented in R-Environment with the package PRSim (Brunner et al., 2021).

The cluster analysis using climate indices and crop yield losses allowed us to interpret the multi-hazard nature of the historical loss events. We identified 6 clusters that are described in Table A.1. Three clusters reasonably explained ca. 70% of soybean crop losses for the region and period studied. Cluster 2 represents years where losses were predominantly driven by precipitation deficit (Single hazard years).

Cluster 4 represents years where losses were driven by precipitation deficit and thermal stress (Multi-hazard year 1). Cluster 6 is associated with relatively normal years in SPI but with heavy rainfall events and higher temperatures (Multi-hazard year 2). The underlying structure of the other clusters (1, 3, and 5) can be related to other drivers of losses that were not considered in the present analysis.

The coupling of high temperatures and droughts have been a major cause of crop losses globally, and global warming is pointed to increased coupled thermal-moisture threats to food production (Lesk et al., 2021). On the other hand, excessive precipitation can increase soil moisture creating conditions for plant hypoxia, which means that plants have less access to oxygen and have a reduction in their energetic status (Brandão e Sodek, 2009). Then, plants become more vulnerable to other threats.

**Table A.1.** Description of each cluster identified

| **Cluster** | **Nº of obs** | **% of Losses** | $SPI$ | $pmax$ | $TX90p$ |
|---|---|---|---|---|---|
| 1 | 389 | 14.4% | 0.702 | 42.4 | 9.55 |
| **2** | **153** | **86.3%** | **-0.941** | **36.1** | **13.4** |
| 3 | 162 | 19.1% | -0.320 | 37.9 | 22.4 |
| **4** | **106** | **96.2%** | **-1.340** | **39.8** | **33.4** |
| 5 | 110 | 27.3% | 1.390 | 76.5 | 11.1 |
| **6** | **46** | **95.7%** | **-0.357** | **70.9** | **22.5** |

The illustrative example highlights the multi-hazard nature effects of extreme weather events on crop yield losses. We used cluster analysis to identify what hazards were dominant each year that a crop loss event occurred. This results using a single hazard approach or assuming independence among hazards can be an oversimplification. The cluster method used for assessing multi-hazard events represent an option that can be used to visualize how to apply the proposed framework (2.3), which may improve decision-making in terms of index selection and vulnerability assessment. Improving hazard identification by categorizing historic data allows insurance designers having new insights in comparison to traditional methods, which conduct statistical analysis of past crop losses associated with a single hazard index.

We summarize the multi-hazard risk analysis for a specific city (Toledo), which is essential for soybean production and is severely impacted by extreme weather occurrences in Figure 5 it is possible to observe that: (i) Toledo had considerable losses on the multi-hazard periods (especially 2012); (ii) both models presented satisfying results for predicting crop losses on the different periods; (iii) the models identified different aspects of the data, implying that a model ensemble might produce the best results.; (iv) although multi-hazard model M2(SPI,pmax,TX90p) better suited the data (presenting a lower mean absolute error), multi-hazard model M1(SPI,TX90p) better predicted the worst year (2012), providing another evidence that an ensemble approach could present better results; (v) The sum of losses estimated with models M1 and M2 (Figure A.1b) presented the lowest overall error in comparison to the observed crop loss, providing further evidence of the importance of using ensembles for predicting crop loss probability.

The study case conducted in this subsection illustrates one possible application of the framework, considering several analyses and visualizations that a stakeholder could use to

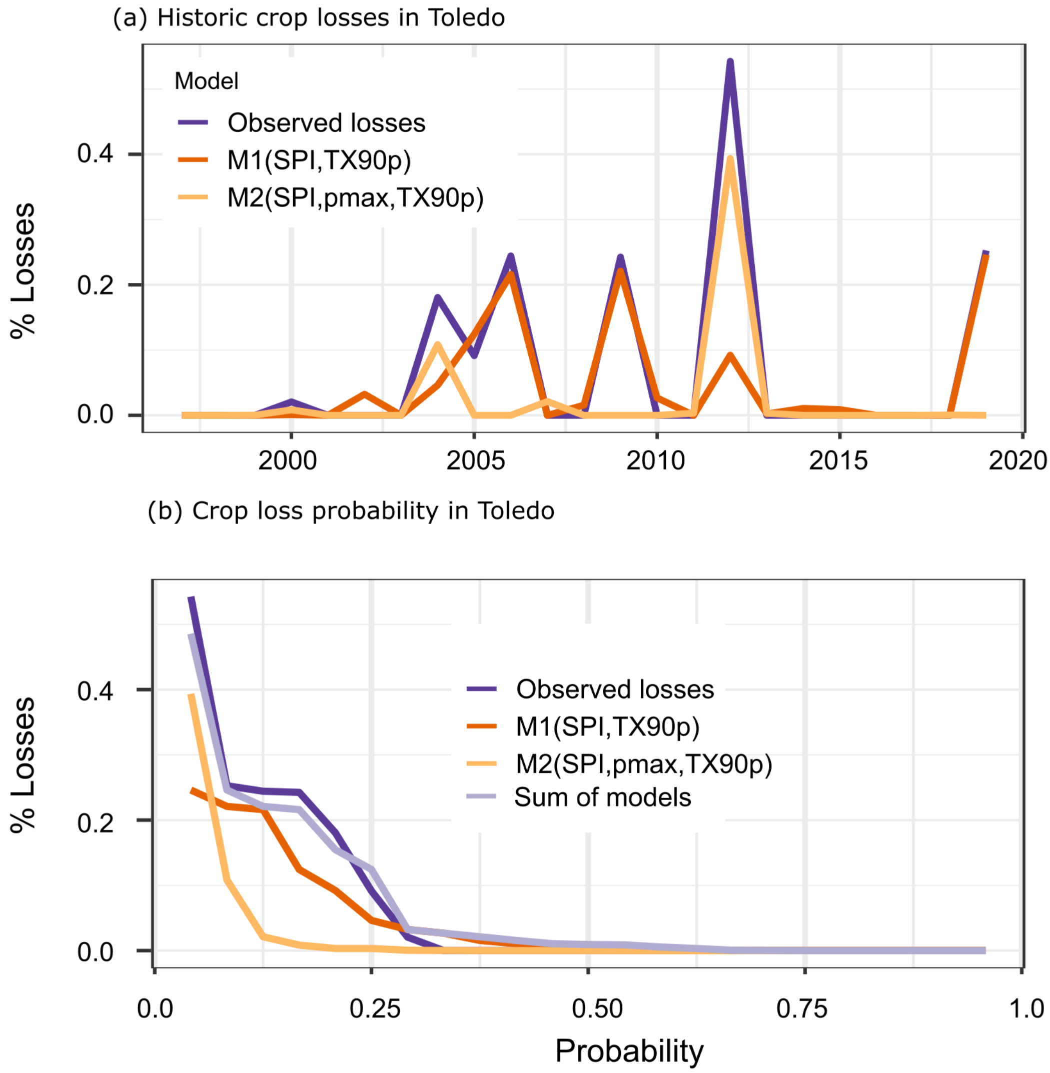


**Figure A.1.** Risk analysis module applied to the city of Toledo-PR, Brazil (1,197.016 km$^2$) for demonstrating the multi-hazard model M1 (SPI,TX90p); multi-hazard model M2 (SPI,pmax,TX90p), the sum of the two models; and observed crop losses in the studied period; (a) simulation of crop losses and (b) crop loss probability in the studied period including the sum of the losses estimated with two loss models

understand better the impacts of extreme weather events over time on agricultural productivity, considering both the historical values and the crop loss probability. This information could improve insurance policy design and a better understanding of different regions' situations. Generating charts such as the ones illustrated in Figure A.1 for multiple regions on a dashboard would allow for a better overview of the impacts of weather events on different crops and regions and be used to improve decision making.

**Table A.2.** Summary of pure risk premiums in terms of percentage of expected crop yields

| | Min | Mean | Max |
|---|---|---|---|
| Historical Burn Analysis | 2.745 % | 5.873 % | 9.722 % |
| Model 1 Synthetic scenario generation | 2.894 % | 3.253 % | 3.630 % |
| Model 2 Synthetic scenario generation | 0.519 % | 0.997 % | 2.506 % |

## Appendix II

### Detrending of Crop Yields

The detrending crop yield process is shown in Fig. A.2 for soybeans and in Fig. A.3.

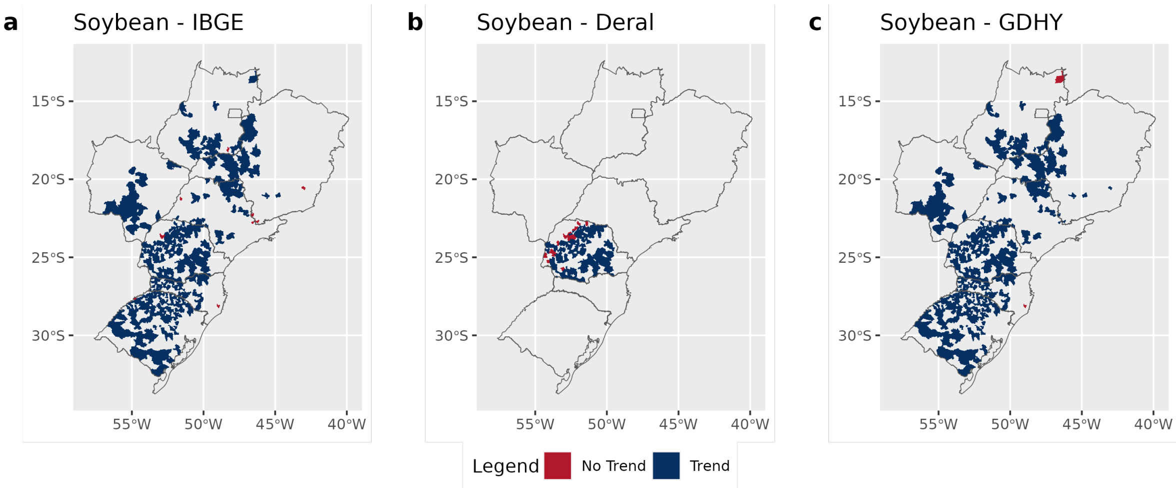


**Figure A.2.** Technological trends for Soybean considering different datasets

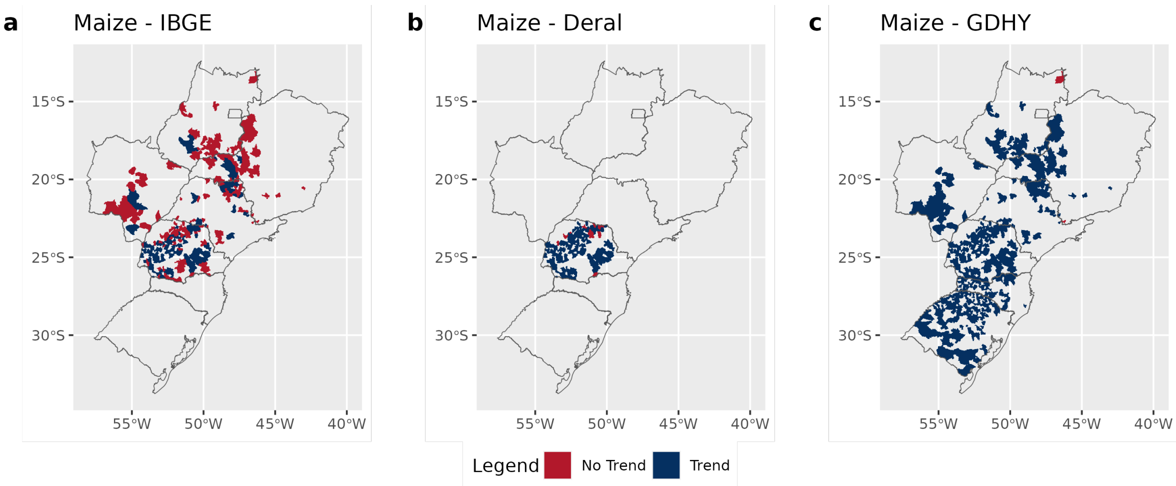


**Figure A.3.** Technological trends for Maize considering different datasets

## Appendix III

### Crop yield Zonal statistics

```r
#require(devtools)
#install_github("ipeaGIT/geobr")
#install.packages(raster)
#install.packages(terra)
#install.packages(tidyverse)

library(geobr)
library(terra)
library(raster)
library(tidyverse)

# Reads shapefile from IBGE using the package geobr

br <- read_municipality()


path <- '/home/marcosrb/Downloads/soybean/' # folder with soybean data nc4 format
path_maize <- '/home/marcosrb/Downloads/maize_second/' # folder with maize data nc4 format


aggregate_to_municipality <- function(path,cities){
  z <- vect(cities) # transform sf object to spat vector
  lapply(list.files(path = path,pattern = "nc4"),function(x){
    r1 <- rast(paste0(path,x)) #  read nc4 crop yield data in raster format
    # adjust raster extension to match -180 to 180 format
    x1 <- crop(r1, extent(180, 360, -90, 90))
    x2 <- crop(r1, extent(0, 180, -90, 90))
    ext(x1) <- c(-180, 0, -90, 90)
    ext(x2) <- c(0, 180, -90, 90)
    m <- merge(x1, x2)
    # create dataframe with crop yield at municipality level´
    dat <- data.frame(
      city = z$name_muni,
      state = z$code_state,
      yield = terra::zonal(m,z,"mean", na.rm=TRUE)$var, # zonal statistics
      year = as.numeric(str_sub(x,7,10))
    )
  })
}

gdhy.soy <- aggregate_to_municipality(path,br)
gdhy.soy <- do.call("rbind",gdhy.soy)
gdhy.soy

gdhy.maize <- aggregate_to_municipality(path_maize,br)
gdhy.maize <- do.call("rbind",gdhy.maize)
gdhy.maize
```

## Appendix IV

## Download of ERA5 Land data using GEE

```

var roi = ee.FeatureCollection("users/marcosbenso/soybean")

print(roi)

var empty = ee.Image().byte();
var outline = empty.paint({
  featureCollection: roi,
  color: 1,
  width: 2
})

Map.addLayer(outline, {palette: '#000000'});
Map.centerObject(roi,7);

var startDate = ee.Date('1961-01-01')
var endDate = ee.Date('2023-01-01')

var collection = ee.ImageCollection("ECMWF/ERA5_LAND/HOURLY")
        .select("temperature_2m")
        .filterDate(startDate, endDate);

var numberOfDays = endDate.difference(startDate, 'days')
var min = ee.ImageCollection(
  ee.List.sequence(0, numberOfDays.subtract(1))
    .map(function (dayOffset) {
      var start = startDate.advance(dayOffset, 'days')
      var end = start.advance(1, 'days')
      return collection
        .filterDate(start, end)
        .min()
    })
)

var max = ee.ImageCollection(
  ee.List.sequence(0, numberOfDays.subtract(1))
    .map(function (dayOffset) {
      var start = startDate.advance(dayOffset, 'days')
      var end = start.advance(1, 'days')
      return collection
        .filterDate(start, end)
        .max()
    })
)

print(collection,"collection");


```

```
function date (image) {
  return image.set({date : image.date().format('yyyy-MM-dd')})
}


var dataset_date = max.map(date)

print(dataset_date);

var feats = min.map(function(image){
  var stats = image.reduceRegions({
    collection: roi,
    reducer: ee.Reducer.mean(),
    scale: 11132,
    crs: 'EPSG:4326'
  })

  stats = stats.map(function(f) {return f.set({date: image.get('date')})})
  return stats.copyProperties(image,['system:index'])
})

print(feats,"feats");

//Empilhando a coleção

print(feats);

var dataset_mean = feats.flatten()
.select(['name_mn','code_mn','date','mean']);


Export.table.toDrive({
  collection: dataset_mean,
  description: 'temperature_2m',
  fileNamePrefix: 'temperature_2m',
  fileFormat: 'CSV',
  selectors: ['name_mn','code_mn','date','mean']
});
```

## Appendix V

### Comparison between random forest models

| ID | State | .y. | group1 | group2 | n1 | n2 | p.adj | p.adj.signif |
|---|---|---|---|---|---|---|---|---|
| Deral | PR | Rsquared | Complex | Temp | 100 | 100 | 0.00e+00 | **** |
| Deral | PR | Rsquared | Prec | Temp | 100 | 100 | 0.00e+00 | **** |
| Deral | PR | Rsquared | Prec-Temp | Temp | 100 | 100 | 0.00e+00 | **** |
| GDHY | RS | Rsquared | Complex | Temp | 100 | 100 | 0.00e+00 | **** |
| GDHY | RS | Rsquared | Prec | Temp | 100 | 100 | 0.00e+00 | **** |
| GDHY | RS | Rsquared | Prec-Temp | Temp | 100 | 100 | 0.00e+00 | **** |
| GDHY | SC | Rsquared | Complex | Prec-Temp | 100 | 100 | 8.00e-03 | ** |
| GDHY | SC | Rsquared | Prec | Temp | 100 | 100 | 2.80e-02 | * |
| GDHY | SC | Rsquared | Prec-Temp | Temp | 100 | 100 | 5.00e-03 | ** |
| GDHY | PR | Rsquared | Complex | Prec | 100 | 100 | 4.56e-04 | *** |
| GDHY | PR | Rsquared | Complex | Prec-Temp | 100 | 100 | 1.00e-03 | ** |
| GDHY | PR | Rsquared | Prec | Temp | 100 | 100 | 1.10e-06 | **** |
| GDHY | PR | Rsquared | Prec-Temp | Temp | 100 | 100 | 4.00e-07 | **** |
| GDHY | GO | Rsquared | Complex | Temp | 100 | 100 | 2.00e-03 | ** |
| IBGE | RS | Rsquared | Complex | Prec | 100 | 100 | 3.00e-03 | ** |
| IBGE | RS | Rsquared | Complex | Temp | 100 | 100 | 0.00e+00 | **** |
| IBGE | RS | Rsquared | Prec | Temp | 100 | 100 | 0.00e+00 | **** |
| IBGE | RS | Rsquared | Prec-Temp | Temp | 100 | 100 | 0.00e+00 | **** |
| IBGE | SC | Rsquared | Complex | Temp | 100 | 100 | 0.00e+00 | **** |
| IBGE | SC | Rsquared | Prec | Temp | 100 | 100 | 0.00e+00 | **** |
| IBGE | SC | Rsquared | Prec-Temp | Temp | 100 | 100 | 0.00e+00 | **** |
| IBGE | PR | Rsquared | Complex | Temp | 100 | 100 | 0.00e+00 | **** |
| IBGE | PR | Rsquared | Prec | Temp | 100 | 100 | 0.00e+00 | **** |
| IBGE | PR | Rsquared | Prec-Temp | Temp | 100 | 100 | 0.00e+00 | **** |
| IBGE | SP | Rsquared | Prec | Temp | 100 | 100 | 2.70e-02 | * |
| IBGE | SP | Rsquared | Prec-Temp | Temp | 100 | 100 | 8.00e-03 | ** |
| IBGE | MS | Rsquared | Complex | Prec | 100 | 100 | 3.00e-02 | * |
| IBGE | MS | Rsquared | Complex | Temp | 100 | 100 | 2.20e-06 | **** |
| IBGE | MS | Rsquared | Prec | Prec-Temp | 100 | 100 | 5.38e-04 | *** |
| IBGE | MS | Rsquared | Prec-Temp | Temp | 100 | 100 | 0.00e+00 | **** |
| IBGE | MG | Rsquared | Complex | Temp | 100 | 100 | 1.30e-02 | * |
| IBGE | MG | Rsquared | Prec | Prec-Temp | 100 | 100 | 3.00e-03 | ** |
| IBGE | MG | Rsquared | Prec | Temp | 100 | 100 | 4.40e-06 | **** |

**Table A.3.** T-test for Randon Forest models for Soybean

| ID | State | .y. | group1 | group2 | n1 | n2 | p.adj | p.adj.signif |
|---|---|---|---|---|---|---|---|---|
| Deral | PR | Rsquared | Complex | Prec-Temp | 100 | 100 | 1.60e-02 | * |
| GDHY | PR | Rsquared | Complex | Prec | 100 | 100 | 0.00e+00 | **** |
| GDHY | PR | Rsquared | Complex | Prec-Temp | 100 | 100 | 3.00e-07 | **** |
| GDHY | PR | Rsquared | Complex | Temp | 100 | 100 | 0.00e+00 | **** |
| GDHY | SP | Rsquared | Complex | Prec | 100 | 100 | 3.01e-05 | **** |
| GDHY | SP | Rsquared | Prec | Prec-Temp | 100 | 100 | 2.40e-02 | * |
| GDHY | SP | Rsquared | Prec | Temp | 100 | 100 | 3.00e-03 | ** |
| GDHY | MS | Rsquared | Complex | Prec | 100 | 100 | 0.00e+00 | **** |
| GDHY | MS | Rsquared | Complex | Prec-Temp | 100 | 100 | 0.00e+00 | **** |
| GDHY | MS | Rsquared | Complex | Temp | 100 | 100 | 0.00e+00 | **** |
| GDHY | MG | Rsquared | Complex | Prec | 100 | 100 | 2.00e-02 | * |
| GDHY | MG | Rsquared | Prec | Temp | 100 | 100 | 3.60e-02 | * |
| GDHY | GO | Rsquared | Complex | Prec | 100 | 100 | 0.00e+00 | **** |
| GDHY | GO | Rsquared | Complex | Prec-Temp | 100 | 100 | 0.00e+00 | **** |
| GDHY | GO | Rsquared | Complex | Temp | 100 | 100 | 6.66e-05 | **** |
| GDHY | GO | Rsquared | Prec-Temp | Temp | 100 | 100 | 1.90e-02 | * |
| IBGE | SP | Rsquared | Prec | Temp | 100 | 100 | 3.90e-02 | * |
| IBGE | MS | Rsquared | Complex | Prec | 100 | 100 | 5.00e-06 | **** |
| IBGE | MS | Rsquared | Complex | Prec-Temp | 100 | 100 | 2.00e-03 | ** |
| IBGE | MG | Rsquared | Complex | Prec | 100 | 100 | 0.00e+00 | **** |
| IBGE | MG | Rsquared | Complex | Prec-Temp | 100 | 100 | 0.00e+00 | **** |
| IBGE | MG | Rsquared | Prec | Temp | 100 | 100 | 0.00e+00 | **** |
| IBGE | MG | Rsquared | Prec-Temp | Temp | 100 | 100 | 2.40e-06 | **** |
| IBGE | GO | Rsquared | Complex | Prec | 100 | 100 | 1.00e-03 | ** |
| IBGE | GO | Rsquared | Complex | Temp | 100 | 100 | 1.70e-02 | * |
| IBGE | GO | Rsquared | Prec | Temp | 100 | 100 | 0.00e+00 | **** |
| IBGE | GO | Rsquared | Prec-Temp | Temp | 100 | 100 | 9.36e-04 | *** |

**Table A.4.** T-test for Randon Forest models for Maize

# Appendix VI

## Variable Importance

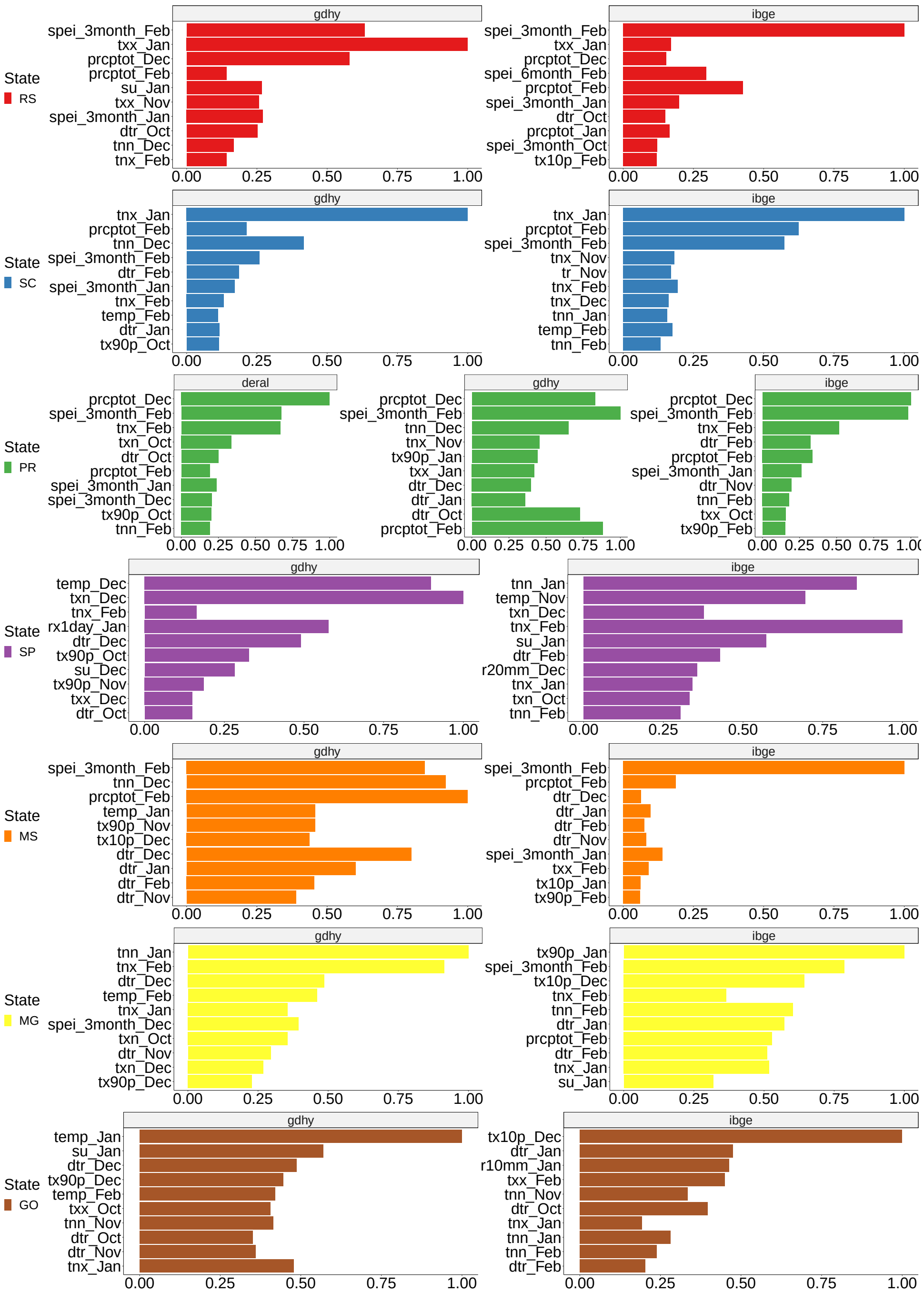


**Figure A.4.** Variable Importance Analysis in Random Forest for Soybean Yield Prediction Across Brazilian States. The figure illustrates the variable importance scores obtained from Random Forest models applied to three distinct datasets: Deral, IBGE, and GDHY, encompassing seven Brazilian states - RS, SC, PR, SP, MS, MG, and GO

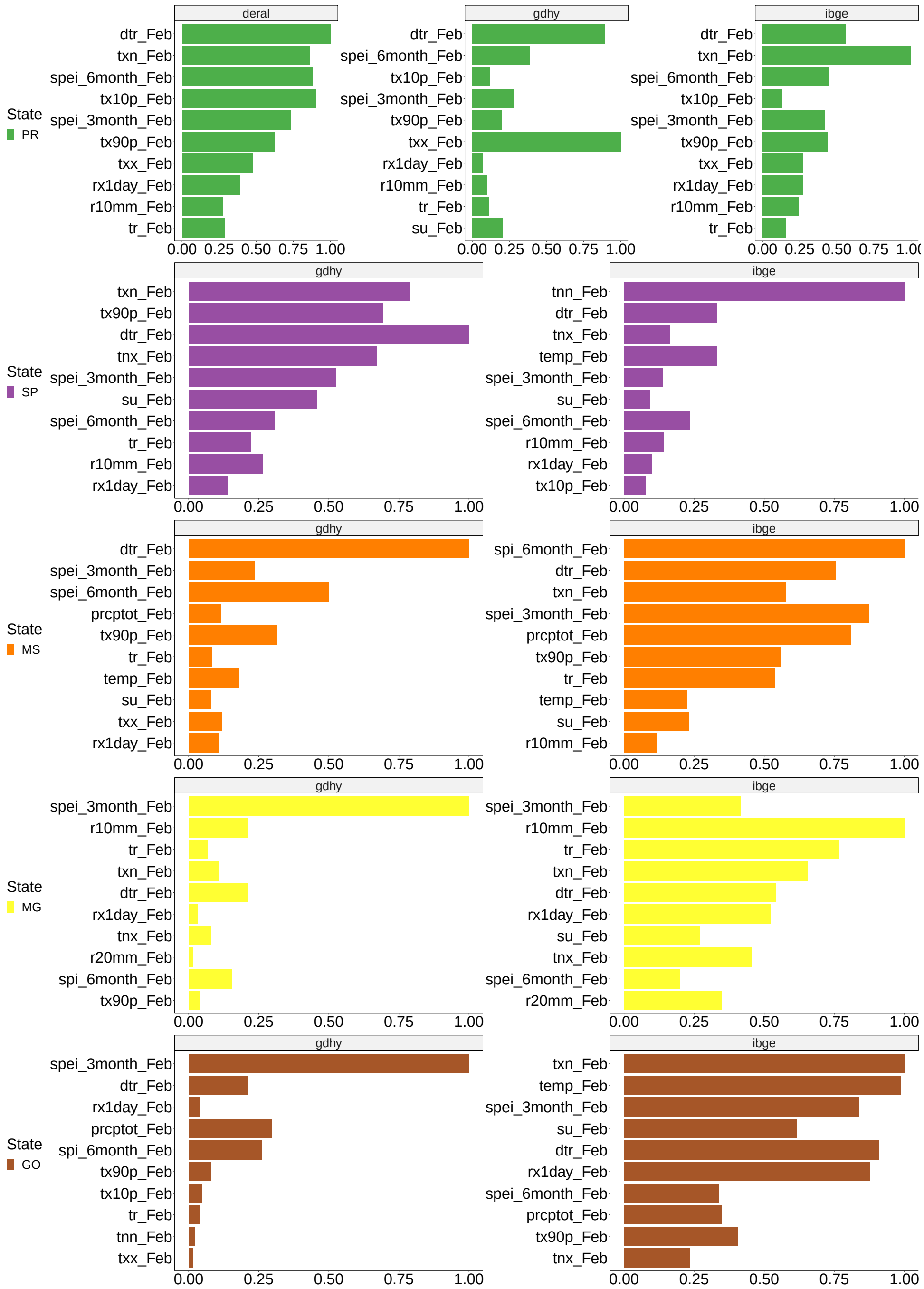


**Figure A.5.** Variable Importance Analysis in Random Forest Models for Maize Yield Prediction Across Brazilian States. The figure illustrates the variable importance scores obtained from Random Forest models applied to three distinct datasets: Deral, IBGE, and GDHY, encompassing seven Brazilian states - PR, SP, MS, MG, and GO

## References

Bhering, S. B., dos SANTOS, H. G., Bognola, I., Cúrcio, G., CARVALHO JUNIOR, W. d., CHAGAS, C. d. S., Manzatto, C., Áglio, M., e SILVA, J. d. S. (2009). Mapa de solos do estado do paraná, legenda atualizada. In *Embrapa*. In: CONGRESSO BRASILEIRO DE CIÊNCIA DO SOLO, 32., 2009, Fortaleza. O solo e ....

Brandão, A. D. e Sodek, L. (2009). Nitrate uptake and metabolism by roots of soybean plants under oxygen deficiency. *Brazilian Journal of Plant Physiology*, 21:13–23.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Brunner, M. I., Slater, L., Tallaksen, L. M., e Clark, M. (2021). Challenges in modeling and predicting floods and droughts: A review. *Wiley Interdisciplinary Reviews: Water*, 8(3):e1520.

Bucheli, J., Dalhaus, T., e Finger, R. (2021). The optimal drought index for designing weather index insurance. *European Review of Agricultural Economics*, 48(3):573–597.

Choudhary, V., D'Alessandro, S., Giertz, A., Suit, K., Johnson, T., Baedeker, T., e Caballero, R. (2016). Agricultural sector risk assessment: methodological guidance for practitioners. *World Bank Group. http://documents. worldbank. org/curated/en/586561467994685817/Agricultural-sector-risk-assessmentmethodological-guidance-forpractitioners*.

da Silva, E. H. F. M., Antolin, L. A. S., Zanon, A. J., Junior, A. S. A., de Souza, H. A., dos Santos Carvalho, K., Junior, N. A. V., e Marin, F. R. (2021). Impact assessment of soybean yield and water productivity in brazil due to climate change. *European Journal of Agronomy*, 129:126329.

Lesk, C., Coffel, E., Winter, J., Ray, D., Zscheischler, J., Seneviratne, S. I., e Horton, R. (2021). Stronger temperature–moisture couplings exacerbate the impact of climate warming on global crop yields. *Nature Food*, 2(9):683–691.

Liaw, A. e Wiener, M. (2002). Classification and regression by randomforest. *R News*, 2(3):18–22.

Lundberg, S. e et al. (2023). A game theoretic approach to explain the output of any machine learning model. [Accessed 30-10-2023].

MacQueen, J. (1967). Classification and analysis of multivariate observations. In *5th Berkeley Symp. Math. Statist. Probability*, pages 281–297.

Mechler, R. (2016). Reviewing estimates of the economic efficiency of disaster risk management: opportunities and limitations of using risk-based cost–benefit analysis. *Natural Hazards*, 81:2121–2147.

Mechler, R., Bouwer, L. M., Linnerooth-Bayer, J., Hochrainer-Stigler, S., Aerts, J. C., Surminski, S., e Williges, K. (2014). Managing unnatural disaster risk from climate extremes. *Nature Climate Change*, 4(4):235–237.

Parana (2021). Levantamento da produção agropecuária. `https://www.agricultura.pr.gov.br/deral/ProducaoAnual`.

Pereira, D. I., Pereira, P., Brilha, J., e Santos, L. (2013). Geodiversity assessment of paraná state (brazil): an innovative approach. *Environmental management*, 52(3):541–552.

Peterson, T., Folland, C., Gruza, G., Hogg, W., Mokssit, A., e Plummer, N. (2001). *Report on the activities of the working group on climate change detection and related rapporteurs.* World Meteorological Organization Geneva.

R Core Team (2022). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Sacks, W. J., Deryng, D., Foley, J. A., e Ramankutty, N. (2010). Crop planting dates: an analysis of global patterns. *Global ecology and biogeography*, 19(5):607–620.

Saxton, K. E. e Rawls, W. J. (2006). Soil water characteristic estimates by texture and organic matter for hydrologic solutions. *Soil science society of America Journal*, 70(5):1569–1578.

# ANNEX

## Annex I

### Crop production in Brazil

**Table B.5.** Crop production of soybean in 2021

| state | crop | production | area | yield | production accumulated | area_acumulated |
|---|---|---|---|---|---|---|
| MT | **Soybean** | 26.188% | 26.698% | 3.39 | 26.19% | 26.698% |
| RS | **Soybean** | 15.134% | 15.586% | 3.23 | 41.32% | 42.28% |
| PR | **Soybean** | 14.233% | 14.090% | 3.45 | 55.56% | 56.37% |
| GO | **Soybean** | 10.119% | 9.838% | 3.40 | 65.67% | 66.21% |
| MS | **Soybean** | 9.061% | 8.618% | 3.43 | 74.74% | 74.83% |
| MG | **Soybean** | 5.176% | 4.939% | 3.61 | 79.91% | 79.77% |
| BA | **Soybean** | 5.069% | 4.338% | 3.14 | 84.98% | 84.11% |
| SP | **Soybean** | 3.101% | 3.061% | 3.18 | 88.08% | 87.17% |
| TO | Soybean | 2.718% | 2.989% | 3.08 | 90.80% | 90.16% |
| MA | Soybean | 2.402% | 2.612% | 3.13 | 93.20% | 92.77% |
| PI | Soybean | 2.011% | 2.123% | 3.08 | 95.21% | 94.89% |
| SC | Soybean | 1.722% | 1.758% | 3.38 | 96.93% | 96.65% |
| PA | Soybean | 1.655% | 1.924% | 2.86 | 98.59% | 98.57% |
| RO | Soybean | 1.002% | 1.022% | 3.21 | 99.59% | 99.60% |
| DF | Soybean | 0.231% | 0.204% | 3.90 | 99.82% | 99.80% |
| RR | Soybean | 0.131% | 0.146% | 3.12 | 99.95% | 99.95% |
| AC | Soybean | 0.015% | 0.016% | 2.88 | 99.97% | 99.96% |
| AP | Soybean | 0.013% | 0.017% | 2.64 | 99.98% | 99.98% |
| AM | Soybean | 0.007% | 0.008% | 3.33 | 99.99% | 99.99% |
| AL | Soybean | 0.007% | 0.008% | 2.85 | 100.00% | 100.00% |
| CE | Soybean | 0.003% | 0.004% | 2.86 | 100.00% | 100.00% |
| PE | Soybean | 0.000% | 0.001% | 2.00 | 100.00% | 100.00% |
| ES | Soybean | 0.000% | 0.000% | NA | 100.00% | 100.00% |
| PB | Soybean | 0.000% | 0.000% | NA | 100.00% | 100.00% |
| RJ | Soybean | 0.000% | 0.000% | NA | 100.00% | 100.00% |
| RN | Soybean | 0.000% | 0.000% | NA | 100.00% | 100.00% |
| SE | Soybean | 0.000% | 0.000% | NA | 100.00% | 100.00% |

**Table B.6.** Crop production of maize off-season in 2021

| state | crop | production | area | yield | production accumulated | area_acumulated |
|---|---|---|---|---|---|---|
| MT | **Maize** | 36.232% | 29.653% | 5.42 | 36.23% | 29.65% |
| GO | **Maize** | 12.153% | 9.612% | 4.90 | 48.38% | 39.26% |
| PR | **Maize** | 11.902% | 14.586% | 4.62 | 60.29% | 53.85% |
| MG | **Maize** | 7.674% | 6.093% | 4.59 | 67.96% | 59.94% |
| MS | **Maize** | 6.125% | 10.980% | 2.98 | 74.09% | 70.92% |
| RS | **Maize** | 4.962% | 3.983% | 5.22 | 79.05% | 74.91% |
| SP | **Maize** | 4.194% | 3.480% | 5.07 | 83.24% | 78.39% |
| BA | **Maize** | 2.770% | 3.154% | 1.86 | 86.01% | 81.54% |
| MA | Maize | 2.563% | 2.467% | 2.38 | 88.57% | 84.01% |
| PI | Maize | 2.425% | 2.665% | 2.02 | 91.00% | 86.67% |
| SC | Maize | 2.268% | 1.730% | 5.85 | 93.27% | 88.40% |
| TO | Maize | 1.706% | 1.805% | 4.04 | 94.97% | 90.21% |
| RO | Maize | 1.532% | 1.626% | 3.50 | 96.51% | 91.83% |
| PA | Maize | 1.269% | 1.849% | 2.27 | 97.78% | 93.68% |
| SE | Maize | 0.839% | 0.895% | 4.53 | 98.61% | 94.58% |
| CE | Maize | 0.468% | 2.848% | 0.75 | 99.08% | 97.42% |
| DF | Maize | 0.366% | 0.306% | 5.40 | 99.45% | 97.73% |
| RR | Maize | 0.129% | 0.109% | 5.10 | 99.58% | 97.84% |
| AC | Maize | 0.120% | 0.174% | 2.52 | 99.70% | 98.01% |
| AL | Maize | 0.092% | 0.277% | 1.58 | 99.79% | 98.29% |
| PE | Maize | 0.075% | 0.934% | 0.66 | 99.87% | 99.22% |
| PB | Maize | 0.054% | 0.441% | 0.58 | 99.92% | 99.67% |
| ES | Maize | 0.047% | 0.070% | 2.96 | 99.97% | 99.74% |
| RN | Maize | 0.015% | 0.232% | 0.28 | 99.98% | 99.97% |
| RJ | Maize | 0.010% | 0.010% | 3.05 | 99.99% | 99.98% |
| AM | Maize | 0.007% | 0.015% | 2.37 | 100.00% | 99.99% |
| AP | Maize | 0.001% | 0.007% | 0.89 | 100.00% | 100.00% |

## Annex II

### Crop calendars

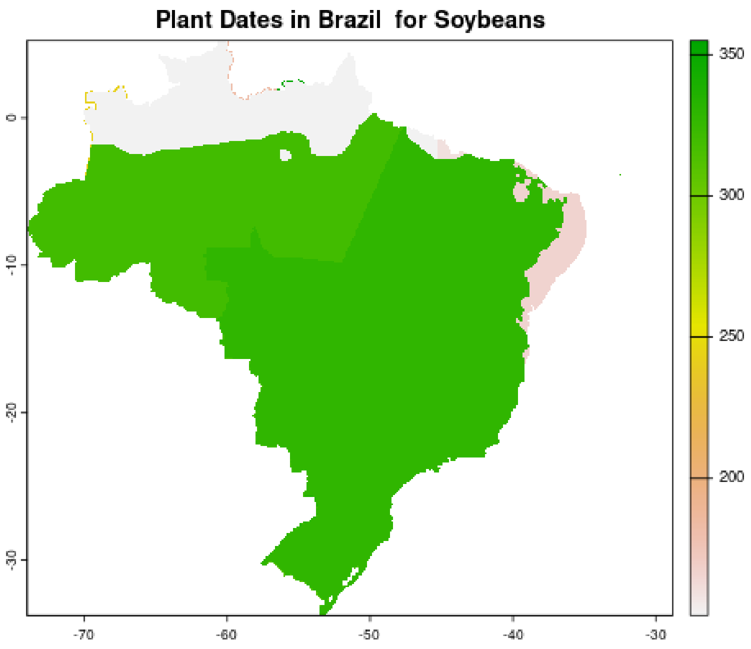


**Figure B.6.** Plant dates in Brazil for soybean. Values represent the day of the year according to Sacks et al. (2010)

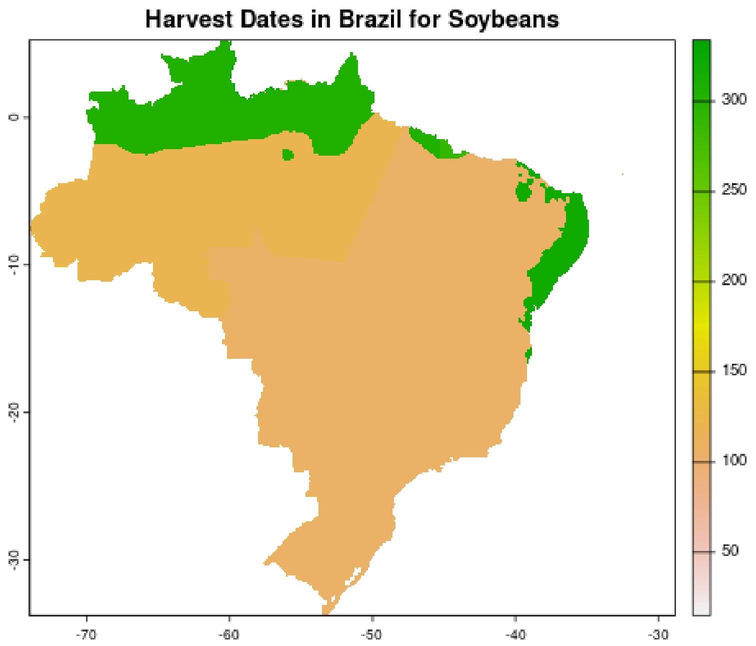


**Figure B.7.** Harvest dates in Brazil for soybean. Values represent the day of the year according to Sacks et al. (2010)

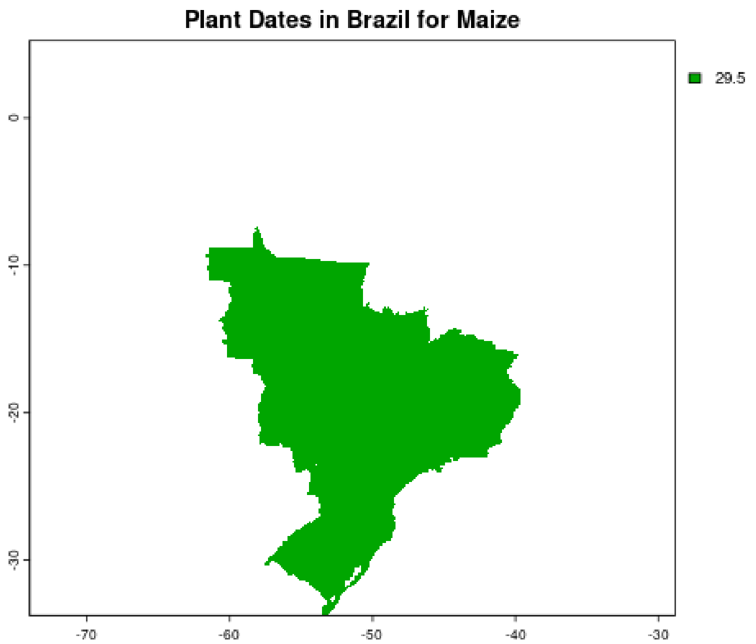


**Figure B.8.** Plant dates in Brazil for maize second cycle. Values represent the day of the year according to Sacks et al. (2010)

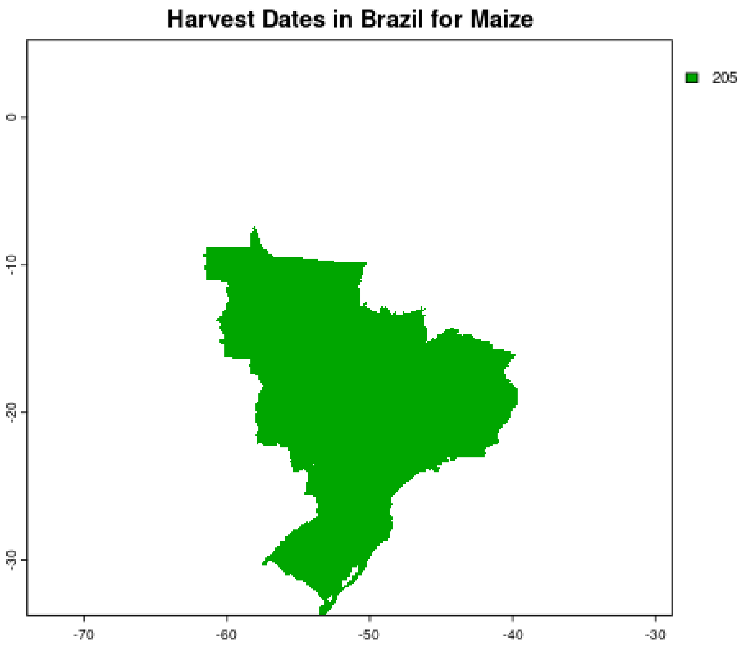


**Figure B.9.** Harvest dates in Brazil for maize second cycle. Values represent the day of the year according to Sacks et al. (2010)

EESC • USP