Supplement of Nat. Hazards Earth Syst. Sci., 25, 1387–1404, 2025
https://doi.org/10.5194/nhess-25-1387-2025-supplement
© Author(s) 2025. CC BY 4.0 License.

Natural Hazards and Earth System Sciences
Open Access

Supplement of

A data-driven framework for assessing climatic impact drivers in the context of food security

Marcos Roberto Benso et al.

Correspondence to: Marcos Roberto Benso (marcosbenso@gmail.com)

The copyright of individual parts of the supplement might differ from the article licence.

S1 Introduction

This supplemental document contains details of data pre-processing and provides additional visual aids for the main manuscripts. Readers might benefit from further details on how crop yield data were obtained and processed and further evaluation of the different crop yield datasets used.

S2 Crop yield data

Data from Brazilian Institute of Geography and Statistics

Crop yield data is a crucial 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 from IBGE can be assessed in the Multidimensional Statistics Database (BME). One of the main challenges of this dataset is that data represents annual statistics and does not represent different cycles. The double cropping system is widely adopted in the study area, therefore representing a potential bottleneck. However, IBGE has started to collect maize in the first and second cycles since 2003. This matches the period when maize's second cycle is intensified in Brazil.

Data from Parana state statistical yearbooks

The Department 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 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 data from Deral. This dataset is derived from the Gross Value of Production, derived from the price and the quantity of production of 30 crops.

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 and Sakai (2020) and Iizumi et al. (2014).

The dataset was aggregated to the municipal level using zonal statistics in the terra package (Hijmans, 2023) in R Studio.

In the literature, three major problems have been reported regarding the quality of crop yield data for risk analysis, namely the presence of outliers, technological trends, and heteroskedasticity. Removal of outliers is a complex problem, as we are dealing with extreme events. The definition of an outlier must be carefully taken to remove valuable 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 normality assumption can be supported. Therefore, for 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 crop yields (Liu and Ker, 2020). This effect is well documented in the agronomic literature and increases the averages and leads to changes in non-constant variance, i.e., heteroskedasticity (Tolhurst and 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 indicates 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 at the year t. The residuals ϵ_t are considered to be the detrended crop yields.

We tested heteroskedasticity in the data using the Pagan-Breusch test (Breusch and 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 heteroskedastic. We compute the normalized residuals y_t^n for two cases in this case. The first case is when the errors are proportional to the yield level. Then, we obtain the y_t^n considering the proportional error ε_t, calculated by dividing the error term from the LOESS prediction function ϵ_t by the yield predicted by the function. Lastly, the proportional errors are multiplied by the yields observed in 2021. With this procedure, the past yields are expressed in terms of 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 corrected for trends, heteroskedasticity, and outliers, the next step is adjusting for a distribution function. Statistical modeling of crop yields is relevant for risk management because we do not have a time series long enough to empirically evaluate risk (Liu and Ker, 2020) empirically. 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. The flexibility to assume different shapes according to the parameters c(α, β), the shape and rate parameters, respectively. For the gamma distribution, when α 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.

To eliminate potential outliers, we excluded values considering each year and immediate region within the state. This was done because he hypothesizes that within the immediate region, the crop yields should be similar. The hypothesis of high correlation of crop yields within studied regions was confirmed for the soybean (Fig. S1) and maize (Fig S2) data, indicating that this strategy is adequate. Since soybeans have a longer record, the correlations were more stable in the immediate regions and tended to have higher values.

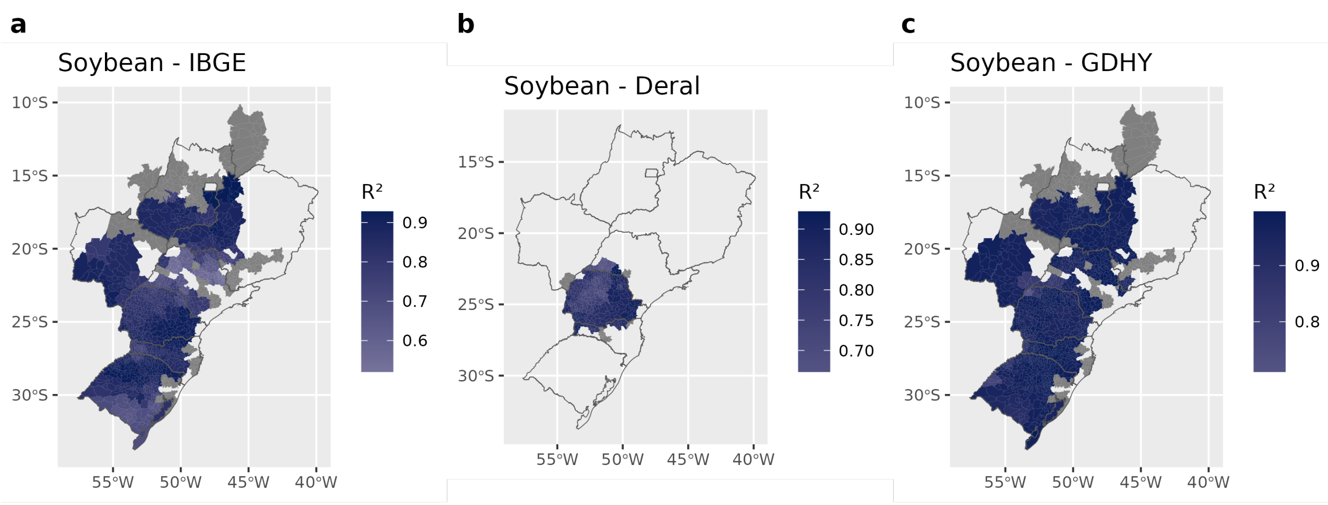

Figure S1. Spatial correlation of municipal soybean crop yields for each dataset, IBGE, Deral and GDHY

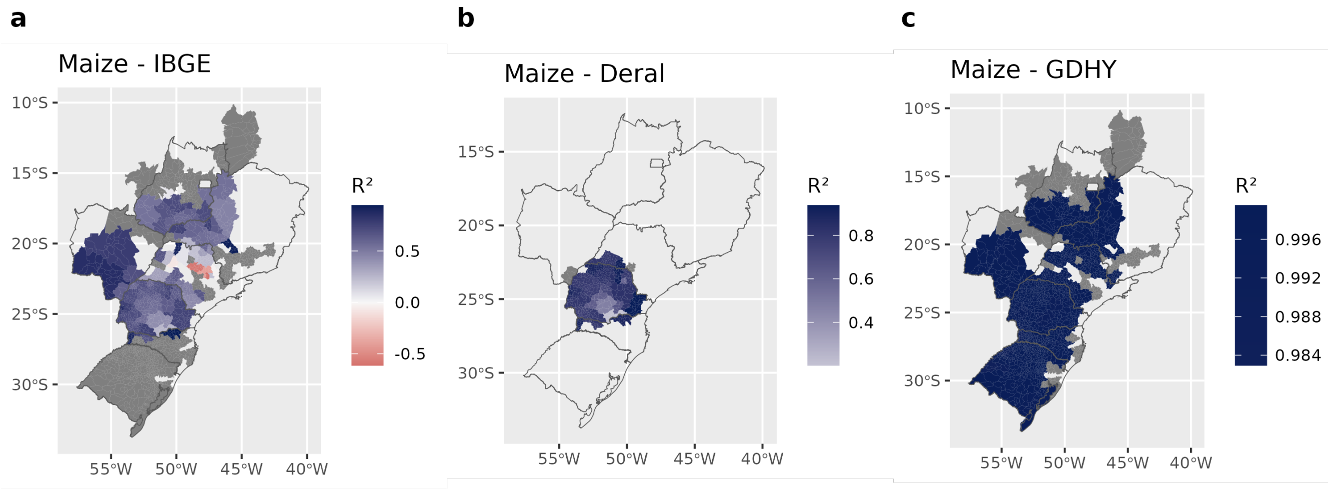

Figure S2. Spatial correlation of municipal maize crop yields for each dataset, IBGE, Deral and GDHY

The comparison of the datasets used in this study is vital 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 improves the simulation of soil moisture (Sinnathamby et al., 2017). For water resources management, using higher quality crop yield data has improved global knowledge on the water-food-energy nexus (Ai and Hanasaki, 2023; Wang et al., 2023).

We compared crop yields at the municipal level in Brazil. As observed in Figures S3 and S4, IBGE and Parná 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 eliminating data improved the dataset's quality. 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 there is a noticeable increase in croplands, which can influence the estimation of GHDY. Also, the exact location of the planted area within each municipality can vary from year to year.

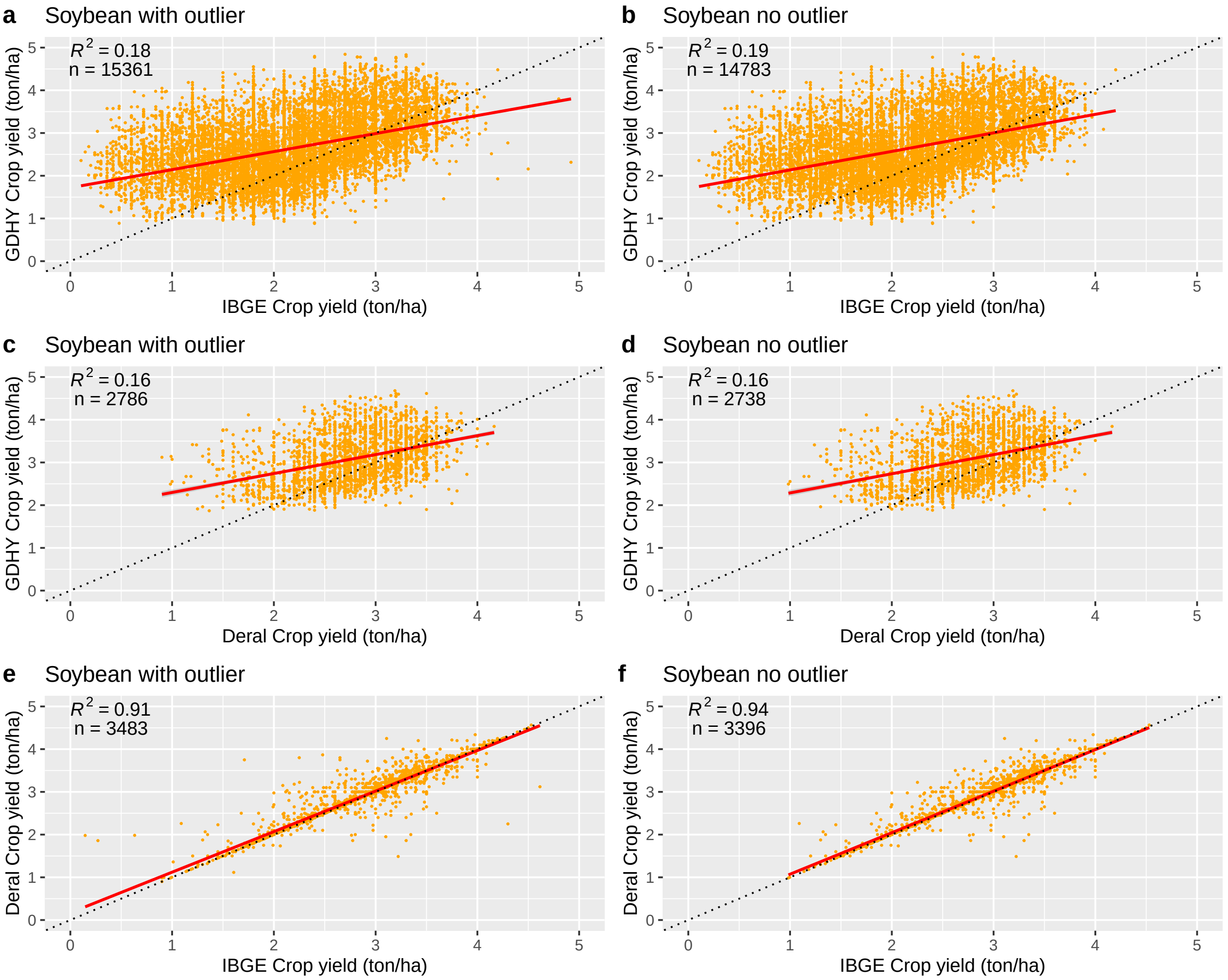

Figure S3. Correlation analysis of three different datasets for soybean crop yields: (Global dataset of historical yields ($yield_{gdhy}$), Brazilian Institute for Geography and Statistics ($yield_{IBGE}$), and Paraná Department of Rural Economy ($yield_{PR}$)

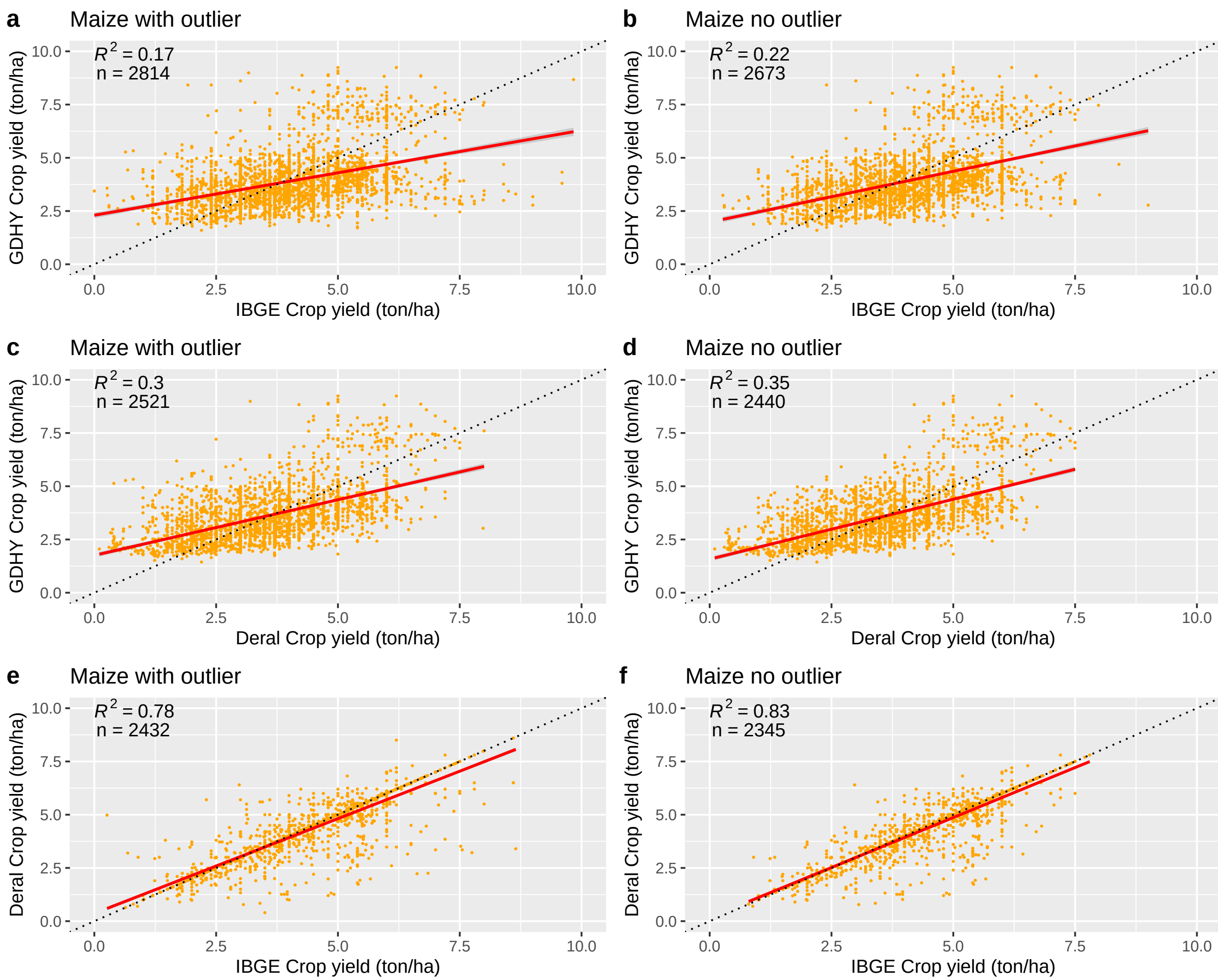

Figure S4. Correlation analysis of three different datasets for maize second cycle crop yields: (Global dataset of historical yields ($yield_{gdhy}$), Brazilian Institute for Geography and Statistics ($yield_{IBGE}$), and Paraná Department of Rural Economy ($yield_{PR}$)

In order to evaluate risk, testing and removing trends is a fundamental step to remove the effects of technology advances on the data (Harri et al., 2011). The 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 in our study presented significant positive monotonic trends considering p values less than 0.05. For maize, on the other hand, significant positive monotonic trends were observed in the majority of municipalities. However, they 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 not having long-term data for the second maize cycle.

Since most data presented positive trends, we applied a LOESS model for all the municipalities. The residuals of the LOESS models were then tested for heteroscedasticity. Other studies evaluated the presence of heteroscedasticity 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) at farm-level studies in São Paulo.

The presence of heteroscedasticity represents systematic changes in crop yield data Yang et al. (1992) and to the best of our knowledge, this is the first study that considers and evaluates spatial characteristics of heteroscedasticity of municipal level in the long term.

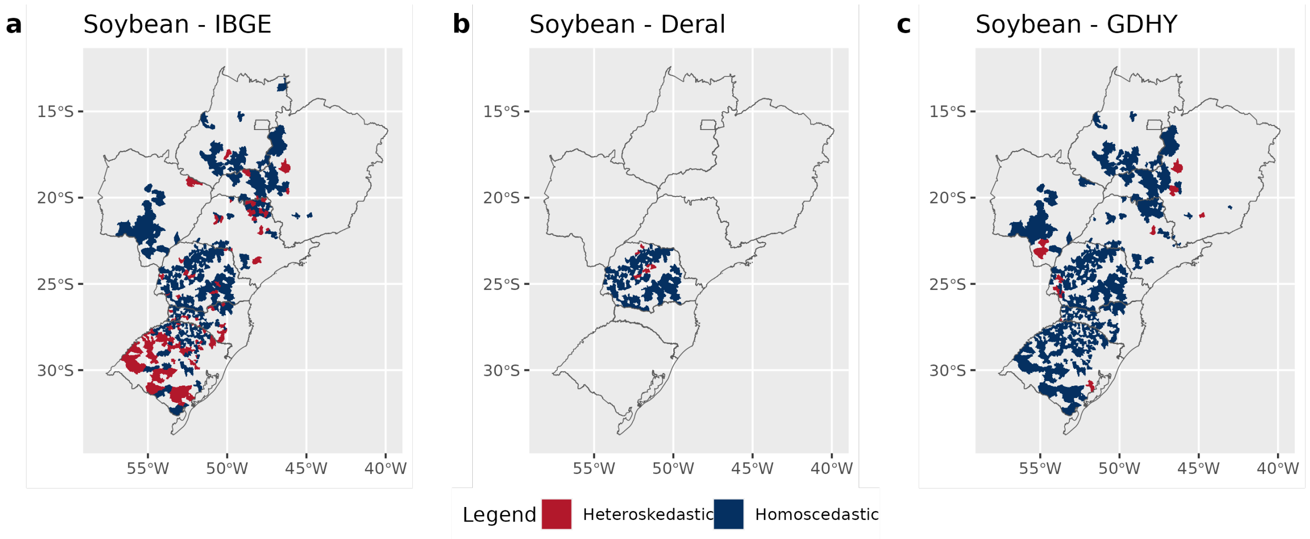

Figure S5. Heteroskedasticity Test Results for soybean crop yields in Brazilian Municipalities

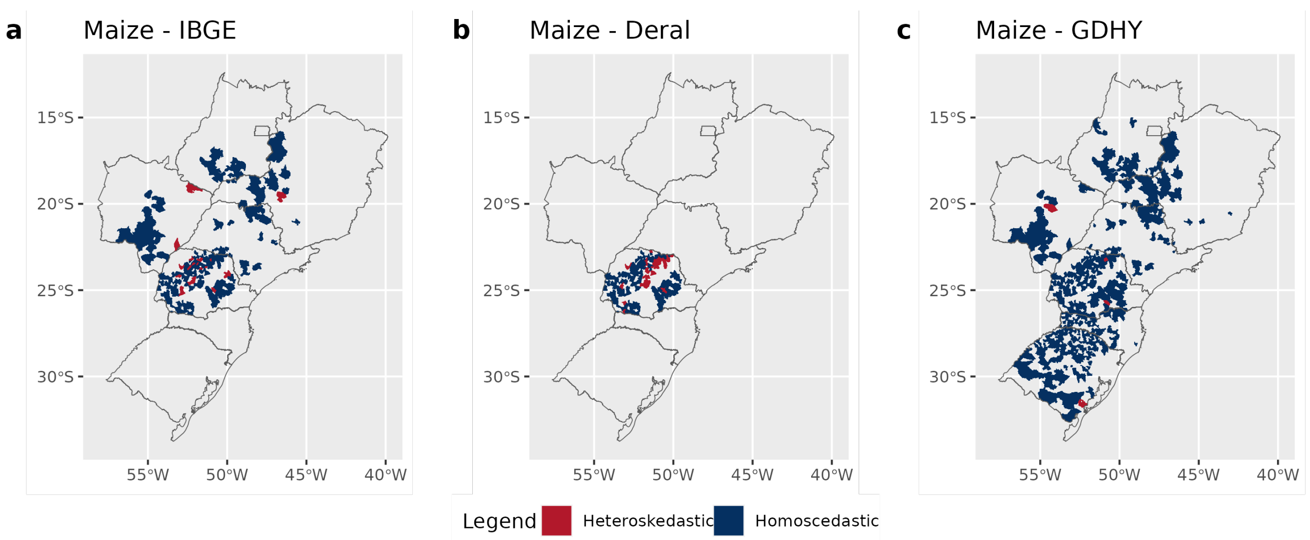

Figure S6. Heteroskedasticity Test Results for maize crop yields in Brazilian Municipalities

S3 Crop yield risk monitoring

To illustrate the relationship between extreme climate impacts on food production and climate indices, we highlight the main loss events from 1997 to 2021 in the study area in Figure S7. Several distinct periods of crop yield losses emerged during the study period, which required further analysis. The notable events among these were 2010 for soybeans, with an average of 16% of crop yield losses, and 2021 for maize, with an average of 40%, which impacted agricultural productivity in the region.

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 might not have sufficient cases of severe failure to be trained adequately and might underestimate losses. The second concern is related to the decision of what to do with these anomalous events. Possible solutions are to use it for training, testing, or removing it from the dataset. We opted to maintain 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. As our aim was to assess the effects of extreme climate events, we opted to retain all of these extreme events within the dataset.

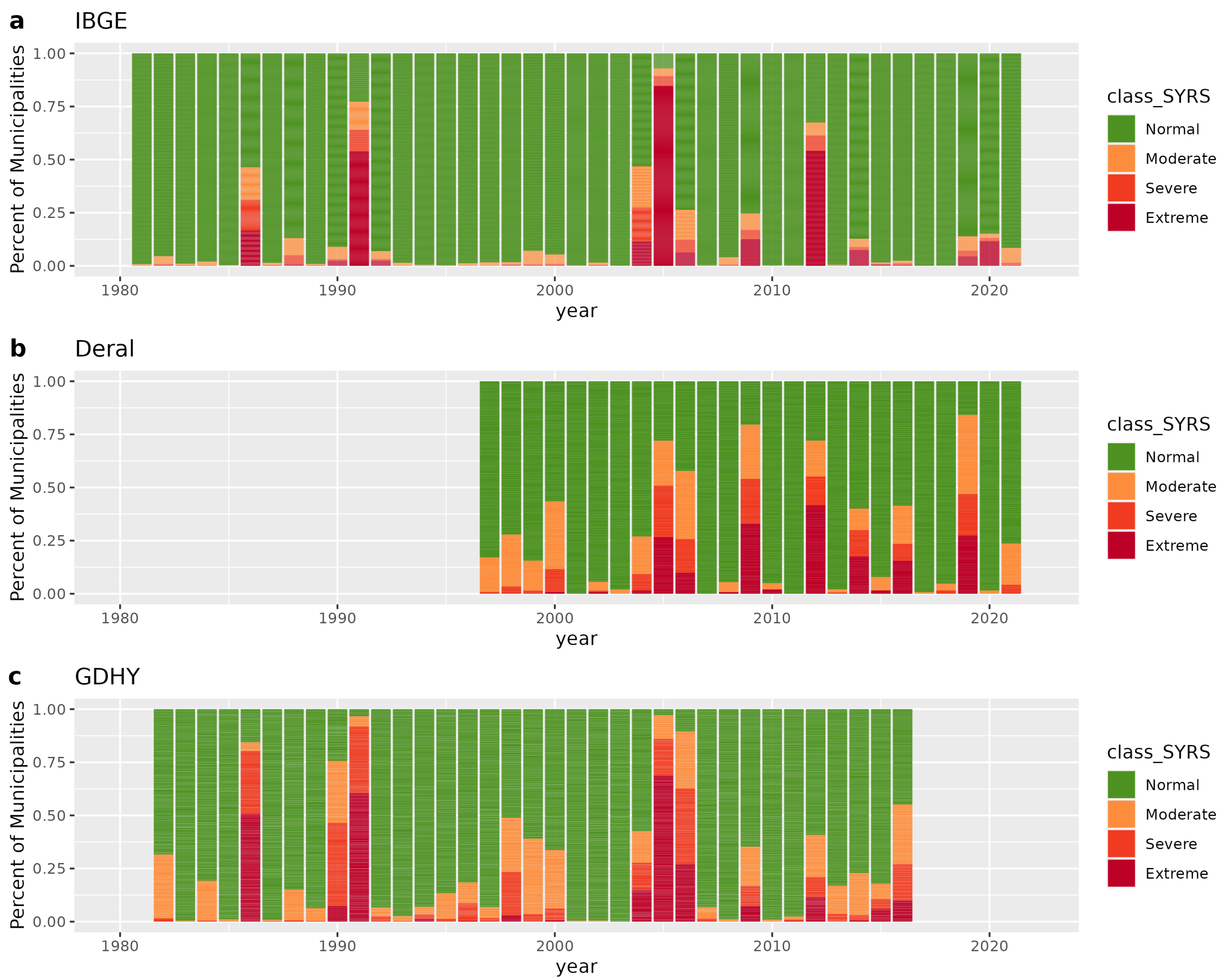

Figure S7. Temporal Variation of Soybean Crop Yields Across Risk Classes for IBGE, Deral and GDHY Datasets. The year-to-year distribution of crop yields is categorized into four risk classes: 'Normal,' 'Moderate,' 'Severe,' and 'Extreme.'

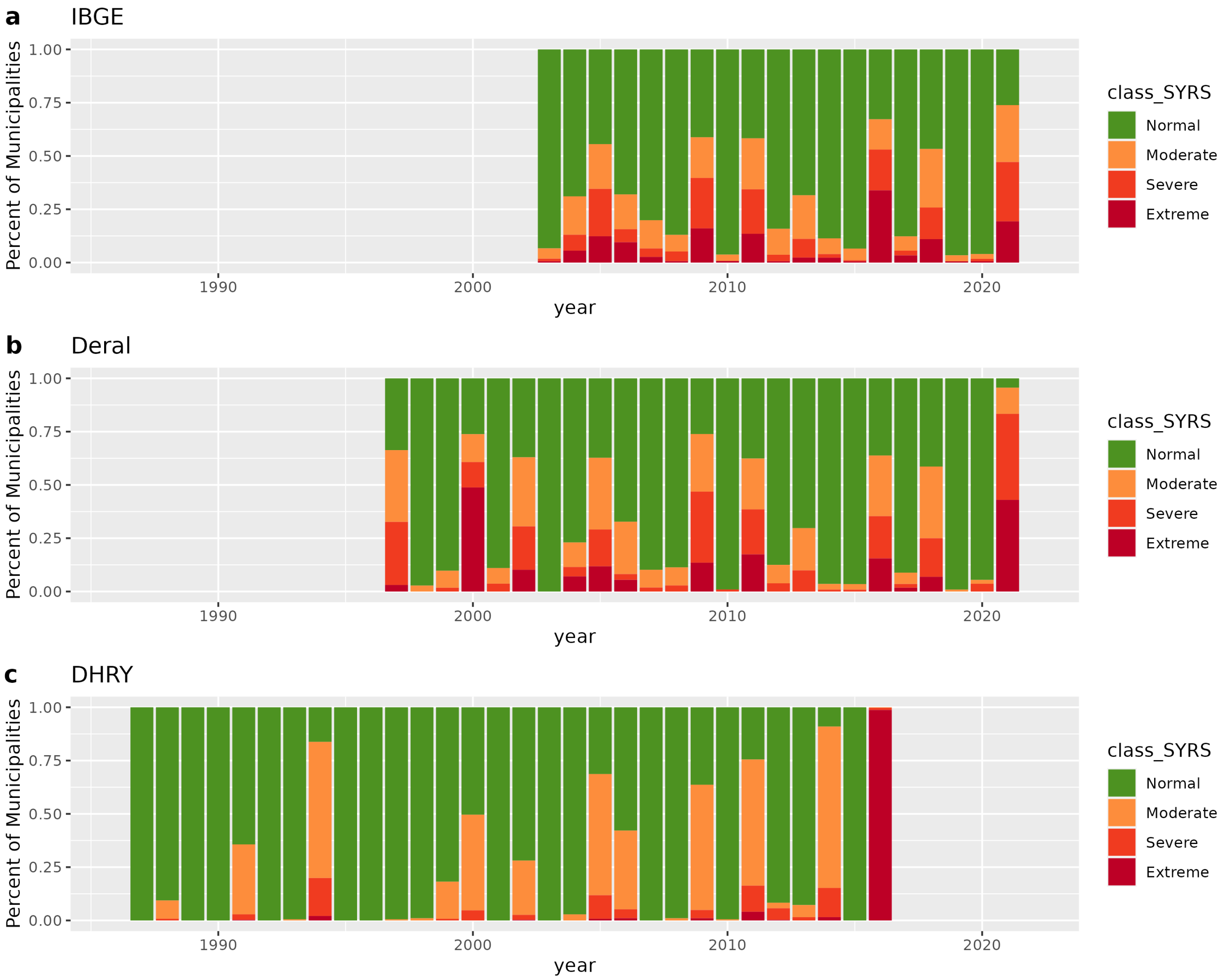

Figure S8. Temporal Variation of Maize Crop Yields Across Risk Classes for IBGE, Deral and GDHY Datasets. The year-to-year distribution of crop yields is categorized into four risk classes: 'Normal,' 'Moderate,' 'Severe,' and 'Extreme.'

S4 SHAP Algorithm

In Fig. S9, 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. The difference between base prediction and the model output, also called SHAP Value, is explained with SHAP explanations.

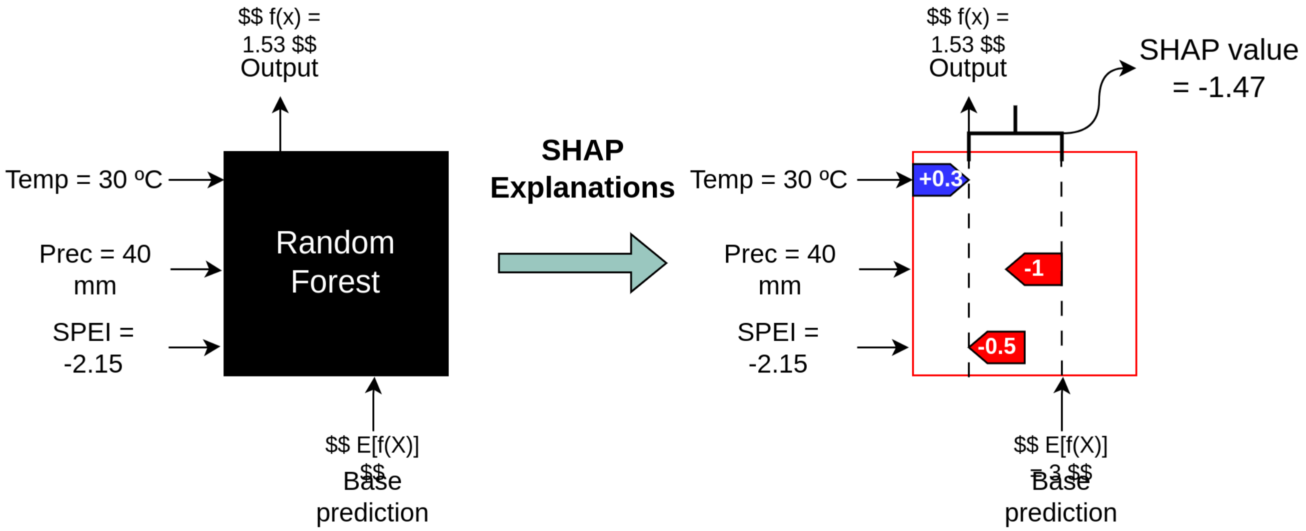

Figure S9. Demonstration of SHAP Explanations. Adapted from Lundberg (2023)

S5 Indentifying key climatic impact-drivers

The analysis presented in Table SS1 reveals that variables such as tnx (indicative of extreme heat), spei_3month (a drought index), and prcptot (precipitation totals) are consistently identified as the most significant variables in various regions. These indices are crucial for predicting the target variable, most likely crop yield losses, as they represent key climate impact drivers. In particular, extreme heat and drought appear to have a profound effect on agricultural results.

The significance of climate variables varies across geographic regions. For instance, in Paraná (PR), factors like tnx_Feb (extreme heat in February) and spei_3month_Feb (February drought) are particularly critical, highlighting heat stress and drought as major concerns for this area during this time. Conversely, in Minas Gerais (MG), variables such as temp_Jan (mean air temperature in January) and tnn_Jan (extreme cold in January) are more significant, indicating that both heat and cold extremes must be taken into account in this region.

Certain months, particularly February and January, prominently appear in multiple significant variables, emphasizing the seasonal effects of climate impacts. For example, February is notably a crucial month for extreme heat and drought events in various areas. This time-related aspect is vital for comprehending when climate factors most significantly influence agricultural systems, providing insights into seasonal patterns that can guide agricultural planning decisions.

The table offers valuable insights into the climate hazards that significantly impact the model's predictions. Variables related to extreme heat, like tnx, txn, and txx, consistently rank highest across different regions, demonstrating that heat stress is a key factor influencing the target variable. Additionally, drought indicators, such as spei_3month, and total precipitation (e.g., prcptot) also hold high importance, underscoring the role of water availability in outcomes such as crop yields. This underscores the urgent need to manage both heat and water stress in agricultural systems.

The results from this Random Forest model uncover several significant trends. Firstly, it emphasizes the importance of temperature extremes and drought indices, indicating that fluctuations in temperature and water stress are key determinants of the target outcome. Furthermore, the model accounts for spatial and temporal variations, which is essential for enhancing predictive precision across various regions and time frames.

	Overall	Variables	rank	UF	dataset	month	hazards
1	100.00000	tnx_Feb	1	PR	Deral	Feb	extreme_heat
2	60.99664	spei_3month_Feb	2	PR	Deral	Feb	drought
3	45.89551	prcptot_Dec	3	PR	Deral	Dec	mean_precipitation
4	34.99172	su_Feb	4	PR	Deral	Feb	extreme_heat
5	31.89870	txn_Oct	5	PR	Deral	Oct	extreme_heat
6	26.03781	dtr_Oct	6	PR	Deral	Oct	mean_air_temp
7	16.10180	spei_3month_Dec	7	PR	Deral	Dec	drought
8	15.23229	dtr_Nov	8	PR	Deral	Nov	mean_air_temp
9	15.15879	dtr_Dec	9	PR	Deral	Dec	mean_air_temp
10	13.13012	dtr_Feb	10	PR	Deral	Feb	mean_air_temp
69	100.00000	temp_Jan	1	MG	GDHY	Jan	mean_air_temp
70	90.78860	tnn_Jan	2	MG	GDHY	Jan	extreme_heat
71	78.32961	tnx_Feb	3	MG	GDHY	Feb	extreme_heat
72	55.61309	txn_Jan	4	MG	GDHY	Jan	extreme_heat
73	50.20335	spei_3month_Dec	5	MG	GDHY	Dec	drought
74	46.13138	dtr_Dec	6	MG	GDHY	Dec	mean_air_temp
75	44.31361	temp_Dec	7	MG	GDHY	Dec	mean_air_temp
76	39.82485	tnx_Jan	8	MG	GDHY	Jan	extreme_heat
77	37.49046	dtr_Nov	9	MG	GDHY	Nov	mean_air_temp
78	34.23006	txn_Oct	10	MG	GDHY	Oct	extreme_heat
137	100.00000	temp_Jan	1	MS	GDHY	Jan	mean_air_temp
138	68.94557	spei_3month_Feb	3	MS	GDHY	Feb	drought
139	66.48501	dtr_Oct	4	MS	GDHY	Oct	mean_air_temp
140	64.68782	tnx_Feb	5	MS	GDHY	Feb	extreme_heat
141	63.31379	tx10p_Jan	6	MS	GDHY	Jan	cold_spell
142	63.07101	su_Jan	7	MS	GDHY	Jan	extreme_heat
143	61.21219	tnx_Jan	8	MS	GDHY	Jan	extreme_heat
144	55.96900	prcptot_Feb	9	MS	GDHY	Feb	mean_precipitation
145	54.61341	txx_Feb	10	MS	GDHY	Feb	extreme_heat
210	100.00000	temp_Nov	1	PR	GDHY	Nov	mean_air_temp
211	94.07476	temp_Oct	2	PR	GDHY	Oct	mean_air_temp
212	91.89831	txn_Oct	3	PR	GDHY	Oct	extreme_heat
213	65.60651	tnn_Oct	4	PR	GDHY	Oct	extreme_heat
214	53.73975	tr_Oct	5	PR	GDHY	Oct	extreme_heat
215	49.21974	dtr_Oct	6	PR	GDHY	Oct	mean_air_temp

216	47.51726	tnx_Nov	7	PR	GDHY	Nov	extreme_heat
217	46.70260	temp_Dec	8	PR	GDHY	Dec	mean_air_temp
218	45.26252	tnx_Feb	9	PR	GDHY	Feb	extreme_heat
219	44.67547	txx_Feb	10	PR	GDHY	Feb	extreme_heat
275	100.00000	tnx_Jan	1	RS	GDHY	Jan	extreme_heat
276	86.92816	prcptot_Dec	2	RS	GDHY	Dec	mean_precipitation
277	72.28343	r10mm_Dec	3	RS	GDHY	Dec	heavy_precipitation
278	68.07999	txx_Jan	4	RS	GDHY	Jan	extreme_heat
279	52.90353	dtr_Nov	5	RS	GDHY	Nov	mean_air_temp
280	49.35403	spei_3month_Feb	6	RS	GDHY	Feb	drought
281	45.35339	prcptot_Feb	7	RS	GDHY	Feb	mean_precipitation
282	42.93565	tnx_Nov	8	RS	GDHY	Nov	extreme_heat
283	38.94912	prcptot_Nov	9	RS	GDHY	Nov	mean_precipitation
284	38.70593	prcptot_Jan	10	RS	GDHY	Jan	mean_precipitation
348	100.00000	tnx_Jan	1	SC	GDHY	Jan	extreme_heat
349	25.86169	tx90p_Oct	2	SC	GDHY	Oct	extreme_heat
350	25.42701	dtr_Feb	3	SC	GDHY	Feb	mean_air_temp
351	22.95067	spei_3month_Feb	4	SC	GDHY	Feb	drought
352	22.75066	txn_Dec	5	SC	GDHY	Dec	extreme_heat
353	20.97155	prcptot_Feb	6	SC	GDHY	Feb	mean_precipitation
354	18.95401	tnn_Dec	7	SC	GDHY	Dec	extreme_heat
355	15.68545	txx_Feb	8	SC	GDHY	Feb	extreme_heat
356	14.64066	spei_6month_Oct	9	SC	GDHY	Oct	drought
357	13.28630	spei_3month_Jan	10	SC	GDHY	Jan	drought
417	100.00000	txn_Dec	1	SP	GDHY	Dec	extreme_heat
418	64.72215	temp_Dec	2	SP	GDHY	Dec	mean_air_temp
419	63.21173	tnn_Dec	3	SP	GDHY	Dec	extreme_heat
420	53.35772	tnx_Dec	4	SP	GDHY	Dec	extreme_heat
421	46.85076	dtr_Dec	5	SP	GDHY	Dec	mean_air_temp
422	38.25983	tx90p_Oct	7	SP	GDHY	Oct	extreme_heat
423	37.62061	txx_Dec	8	SP	GDHY	Dec	extreme_heat
424	36.21886	tnn_Jan	9	SP	GDHY	Jan	extreme_heat
425	34.00213	su_Dec	10	SP	GDHY	Dec	extreme_heat
489	100.00000	dtr_Feb	1	MG	IBGE	Feb	mean_air_temp
490	90.38603	dtr_Jan	2	MG	IBGE	Jan	mean_air_temp

491	59.19799	tnx_Jan	3	MG	IBGE	Jan	extreme_heat
492	58.55139	tnn_Feb	4	MG	IBGE	Feb	extreme_heat
493	56.02843	prcptot_Feb	5	MG	IBGE	Feb	mean_precipitation
494	55.44643	txn_Feb	6	MG	IBGE	Feb	extreme_heat
495	48.38260	spei_3month_Feb	7	MG	IBGE	Feb	drought
496	48.15441	prcptot_Nov	8	MG	IBGE	Nov	mean_precipitation
497	47.54246	dtr_Oct	9	MG	IBGE	Oct	mean_air_temp
498	47.39682	temp_Feb	10	MG	IBGE	Feb	mean_air_temp
555	100.00000	spei_3month_Feb	1	MS	IBGE	Feb	drought
556	62.18494	prcptot_Feb	2	MS	IBGE	Feb	mean_precipitation
557	27.14332	spi_6month_Feb	3	MS	IBGE	Feb	drought
558	25.20665	spei_6month_Feb	4	MS	IBGE	Feb	drought
559	17.57680	tnx_Feb	5	MS	IBGE	Feb	extreme_heat
560	15.85621	txx_Feb	6	MS	IBGE	Feb	extreme_heat
561	14.87482	spei_3month_Jan	7	MS	IBGE	Jan	drought
562	13.31633	dtr_Jan	8	MS	IBGE	Jan	mean_air_temp
563	12.22819	dtr_Dec	9	MS	IBGE	Dec	mean_air_temp
564	11.46425	temp_Jan	10	MS	IBGE	Jan	mean_air_temp
628	100.00000	spei_3month_Feb	1	PR	IBGE	Feb	drought
629	58.20304	prcptot_Dec	2	PR	IBGE	Dec	mean_precipitation
630	55.22292	tnx_Feb	3	PR	IBGE	Feb	extreme_heat
631	27.02189	tnx_Nov	4	PR	IBGE	Nov	extreme_heat
632	24.09338	dtr_Feb	5	PR	IBGE	Feb	mean_air_temp
633	23.54219	prcptot_Feb	6	PR	IBGE	Feb	mean_precipitation
634	21.53629	dtr_Nov	7	PR	IBGE	Nov	mean_air_temp
635	16.68468	spei_6month_Feb	8	PR	IBGE	Feb	drought
636	16.61503	tnn_Oct	9	PR	IBGE	Oct	extreme_heat
637	16.27090	txx_Feb	10	PR	IBGE	Feb	extreme_heat
695	100.00000	spei_3month_Feb	1	RS	IBGE	Feb	drought
696	78.11165	prcptot_Feb	2	RS	IBGE	Feb	mean_precipitation
697	42.98244	prcptot_Jan	3	RS	IBGE	Jan	mean_precipitation
698	36.42602	prcptot_Dec	4	RS	IBGE	Dec	mean_precipitation
699	31.35147	r10mm_Dec	5	RS	IBGE	Dec	heavy_precipitation
700	29.18250	dtr_Oct	6	RS	IBGE	Oct	mean_air_temp
701	19.06189	spei_3month_Jan	7	RS	IBGE	Jan	drought
702	18.24222	r10mm_Jan	8	RS	IBGE	Jan	heavy_precipitation
703	14.84210	prcptot_Oct	9	RS	IBGE	Oct	mean_precipitation

704	14.39600	tnx_Jan	10	RS	IBGE	Jan	extreme_heat
770	100.00000	tnx_Jan	1	SC	IBGE	Jan	extreme_heat
771	56.02377	temp_Jan	2	SC	IBGE	Jan	mean_air_temp
772	48.16173	spei_3month_Feb	3	SC	IBGE	Feb	drought
773	46.39182	prcptot_Feb	4	SC	IBGE	Feb	mean_precipitation
774	31.39655	tnx_Feb	5	SC	IBGE	Feb	extreme_heat
775	25.19294	tnn_Feb	6	SC	IBGE	Feb	extreme_heat
776	23.85294	tnn_Jan	7	SC	IBGE	Jan	extreme_heat
777	22.63478	temp_Nov	8	SC	IBGE	Nov	mean_air_temp
778	16.09298	r10mm_Feb	9	SC	IBGE	Feb	heavy_precipitation
779	15.54663	txn_Nov	10	SC	IBGE	Nov	extreme_heat
838	100.00000	txn_Dec	1	SP	IBGE	Dec	extreme_heat
839	96.21386	tnn_Dec	2	SP	IBGE	Dec	extreme_heat
840	96.17796	temp_Nov	3	SP	IBGE	Nov	mean_air_temp
841	91.15508	tnn_Oct	4	SP	IBGE	Oct	extreme_heat
842	89.22131	temp_Oct	5	SP	IBGE	Oct	mean_air_temp
843	85.13383	temp_Jan	6	SP	IBGE	Jan	mean_air_temp
844	77.75324	temp_Dec	7	SP	IBGE	Dec	mean_air_temp
845	76.73861	txn_Nov	8	SP	IBGE	Nov	extreme_heat
846	76.36698	txn_Oct	9	SP	IBGE	Oct	extreme_heat
847	75.43840	tnn_Jan	10	SP	IBGE	Jan	extreme_heat

Table S1. 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

Table SS2 offers a ranking of climate variables impacting Maize second cycle various regions across Brazil, sorted by state, month, and type of hazard. It emphasizes extreme climate conditions such as severe heat, droughts, cold snaps, and average air temperature. The data is divided based on the overall influence of these variables in different months, illustrating how different regions (denoted by their state abbreviations) are affected.

In the state of Paraná (PR), April precipitation (prcptot_Apr) held the top rank, followed by extreme heat conditions in May (tnn_May), and these variables markedly impacted the agricultural outcomes in both months. Furthermore, mean air temperatures in April and May (temp_Apr, temp_May), along with July precipitation (prcptot_Jul), were also key factors, indicating the significant role of both rainfall and temperature.

Conversely, in Goiás (GO), the highest-ranking variable was the diurnal temperature range in April (dtr_Apr). Significant factors also included extreme heat in May (tnn_May) and high precipitation levels in February (prcptot_Feb) and May (prcptot_May). This suggests that variations in both precipitation and temperature significantly influence agricultural and climate vulnerability in Goiás during these periods.

In Minas Gerais (MG), the total precipitation for May (prcptot_May) was identified as the most critical variable, followed by the average temperature in March (temp_Mar) and the extreme maximum temperature in August (tnx_Ago). Additionally, the drought conditions in June (spei_3month_Jun) and April's diurnal temperature range (dtr_Apr) were significant in influencing the state's climate impact.

Mato Grosso do Sul (MS) is significantly impacted by February's average temperature (temp_Feb), with severe heat (su_Feb) and drought conditions in June (spei_6month_Jun) being key factors. The diurnal temperature range in August (dtr_Ago) and precipitation levels in April (prcptot_Apr) highlight that both heat and variations in precipitation shape the climate patterns in the state.

For Rio Grande do Sul (RS), the diurnal temperature range in May (dtr_May) emerges as the most crucial variable, with drought conditions in March (spi_6month_Mar) and elevated precipitation in August (prcptot_Ago) and July (prcptot_Jul) also playing significant roles. This highlights the critical impact of the interplay between temperature and precipitation extremes in RS.

Lastly, São Paulo (SP) exhibits a significant trend where extreme temperatures in August (tnx_Ago) and heavy rainfall in July (prcptot_Jul) are the primary climate hazards, accompanied by other temperature indicators such as diurnal temperature range and extreme heat events. This focus on these factors underscores the vital impact of heat and precipitation patterns on the environmental and agricultural issues faced by São Paulo.

	Overall	Variables	rank	UF	dataset	month	hazards
1	100.00000	prcptot_Apr	1	PR	Deral	Apr	mean_precipitation
2	82.46489	tnn_May	2	PR	Deral	May	extreme_heat
3	75.13381	temp_May	3	PR	Deral	May	mean_air_temp
4	71.49528	prcptot_Jul	4	PR	Deral	Jul	mean_precipitation
5	70.02075	txn_May	5	PR	Deral	May	extreme_heat
6	67.53607	temp_Apr	6	PR	Deral	Apr	mean_air_temp
7	64.57218	tnx_Ago	7	PR	Deral	Ago	extreme_heat
8	63.82222	dtr_Ago	8	PR	Deral	Ago	mean_air_temp
9	61.67738	tnn_Feb	9	PR	Deral	Feb	extreme_heat
10	61.62358	temp_Ago	10	PR	Deral	Ago	mean_air_temp
98	100.00000	dtr_Apr	1	GO	GDHY	Apr	mean_air_temp
99	97.50680	tnn_May	2	GO	GDHY	May	extreme_heat
100	65.78049	txn_May	3	GO	GDHY	May	extreme_heat
101	64.64102	tnn_Jun	4	GO	GDHY	Jun	extreme_heat
102	59.29969	prcptot_Feb	5	GO	GDHY	Feb	mean_precipitation
103	38.93645	tnx_Jun	6	GO	GDHY	Jun	extreme_heat
104	36.49084	prcptot_May	7	GO	GDHY	May	mean_precipitation
105	27.97299	txn_Jun	8	GO	GDHY	Jun	extreme_heat
106	21.80135	temp_Jun	9	GO	GDHY	Jun	mean_air_temp
107	19.81397	tnx_Ago	10	GO	GDHY	Ago	extreme_heat
192	100.00000	prcptot_May	1	MG	GDHY	May	mean_precipitation
193	78.72920	temp_Mar	2	MG	GDHY	Mar	mean_air_temp
194	75.93930	dtr_May	3	MG	GDHY	May	mean_air_temp
195	53.84657	tnx_Ago	5	MG	GDHY	Ago	extreme_heat
196	50.52855	dtr_Apr	6	MG	GDHY	Apr	mean_air_temp
197	45.21166	txn_Mar	7	MG	GDHY	Mar	extreme_heat
198	44.15792	dtr_Mar	8	MG	GDHY	Mar	mean_air_temp
199	41.63691	txx_Ago	10	MG	GDHY	Ago	extreme_heat
284	100.00000	temp_Feb	1	MS	GDHY	Feb	mean_air_temp
285	65.96125	su_Feb	2	MS	GDHY	Feb	extreme_heat
286	55.30782	tnx_Feb	3	MS	GDHY	Feb	extreme_heat
287	40.40994	spei_6month_Jun	4	MS	GDHY	Jun	drought
288	39.67969	txn_May	5	MS	GDHY	May	extreme_heat
289	31.04435	tx90p_May	7	MS	GDHY	May	extreme_heat
290	26.72179	dtr_Ago	8	MS	GDHY	Ago	mean_air_temp
291	24.86532	prcptot_Apr	9	MS	GDHY	Apr	mean_precipitation

292	24.72938	tnn_May	10	MS	GDHY	May	extreme_heat
389	100.00000	tnx_Ago	1	PR	GDHY	Ago	extreme_heat
390	61.89264	temp_Ago	2	PR	GDHY	Ago	mean_air_temp
391	43.43037	txn_May	3	PR	GDHY	May	extreme_heat
392	42.71941	temp_May	4	PR	GDHY	May	mean_air_temp
393	29.23238	tnn_May	5	PR	GDHY	May	extreme_heat
394	21.88527	tnx_May	6	PR	GDHY	May	extreme_heat
395	21.65946	dtr_Apr	7	PR	GDHY	Apr	mean_air_temp
396	15.10926	dtr_May	8	PR	GDHY	May	mean_air_temp
397	13.37548	dtr_Ago	9	PR	GDHY	Ago	mean_air_temp
398	13.15006	tnn_Mar	10	PR	GDHY	Mar	extreme_heat
487	100.00000	dtr_May	1	RS	GDHY	May	mean_air_temp
488	82.31632	spi_6month_Mar	2	RS	GDHY	Mar	drought
489	68.54319	prcptot_Ago	3	RS	GDHY	Ago	mean_precipitation
490	64.07012	dtr_Ago	4	RS	GDHY	Ago	mean_air_temp
491	59.57734	tnn_Jun	5	RS	GDHY	Jun	extreme_heat
492	53.79602	spei_3month_Jul	6	RS	GDHY	Jul	drought
493	49.04170	prcptot_Jul	8	RS	GDHY	Jul	mean_precipitation
494	46.64403	tx10p_May	9	RS	GDHY	May	cold_spell
495	45.34317	tx10p_Apr	10	RS	GDHY	Apr	cold_spell
588	100.00000	dtr_Apr	1	SC	GDHY	Apr	mean_air_temp
589	96.40534	dtr_Ago	2	SC	GDHY	Ago	mean_air_temp
590	96.37887	dtr_May	3	SC	GDHY	May	mean_air_temp
591	87.34563	r10mm_Ago	5	SC	GDHY	Ago	heavy_precipitation
592	84.21535	txx_May	6	SC	GDHY	May	extreme_heat
593	77.93737	tnn_Jun	7	SC	GDHY	Jun	extreme_heat
594	77.68189	prcptot_Ago	8	SC	GDHY	Ago	mean_precipitation
595	77.21550	spi_6month_Mar	9	SC	GDHY	Mar	drought
596	75.58487	spei_3month_Ago	10	SC	GDHY	Ago	drought
687	100.00000	tnx_Ago	1	SP	GDHY	Ago	extreme_heat
688	32.19566	prcptot_Jul	2	SP	GDHY	Jul	mean_precipitation
689	23.34099	prcptot_Feb	3	SP	GDHY	Feb	mean_precipitation
690	20.77636	dtr_Jul	4	SP	GDHY	Jul	mean_air_temp
691	20.63356	txx_Ago	5	SP	GDHY	Ago	extreme_heat
692	19.63010	txn_Mar	6	SP	GDHY	Mar	extreme_heat

693	17.00114	dtr_Apr	7	SP	GDHY	Apr	mean_air_temp
694	16.21477	tnn_Feb	8	SP	GDHY	Feb	extreme_heat
695	12.99909	tnn_Mar	10	SP	GDHY	Mar	extreme_heat
787	100.00000	tnn_Feb	1	GO	IBGE	Feb	extreme_heat
788	97.29698	txx_Jul	2	GO	IBGE	Jul	extreme_heat
789	95.93036	txx_May	3	GO	IBGE	May	extreme_heat
790	85.09398	tnx_May	4	GO	IBGE	May	extreme_heat
791	82.40128	temp_May	5	GO	IBGE	May	mean_air_temp
792	77.77108	dtr_Apr	6	GO	IBGE	Apr	mean_air_temp
793	77.47015	prcptot_Apr	7	GO	IBGE	Apr	mean_precipitation
794	76.32820	tnx_Apr	8	GO	IBGE	Apr	extreme_heat
795	74.88640	temp_Jul	9	GO	IBGE	Jul	mean_air_temp
887	100.00000	tx90p_Apr	1	MG	IBGE	Apr	extreme_heat
888	79.61648	temp_Apr	2	MG	IBGE	Apr	mean_air_temp
889	62.98098	spei_3month_Jun	3	MG	IBGE	Jun	drought
890	62.76512	tx90p_May	4	MG	IBGE	May	extreme_heat
891	61.25962	spei_3month_Jul	5	MG	IBGE	Jul	drought
892	60.48007	tnx_Apr	6	MG	IBGE	Apr	extreme_heat
893	58.55813	dtr_Ago	7	MG	IBGE	Ago	mean_air_temp
894	58.21883	dtr_Jul	8	MG	IBGE	Jul	mean_air_temp
895	57.59278	dtr_Apr	9	MG	IBGE	Apr	mean_air_temp
896	57.09095	prcptot_May	10	MG	IBGE	May	mean_precipitation
982	100.00000	temp_Apr	1	MS	IBGE	Apr	mean_air_temp
983	93.00263	tnn_Jun	2	MS	IBGE	Jun	extreme_heat
984	54.44418	su_Apr	3	MS	IBGE	Apr	extreme_heat
985	52.53396	tnx_Jun	4	MS	IBGE	Jun	extreme_heat
986	49.29613	prcptot_Feb	5	MS	IBGE	Feb	mean_precipitation
987	46.85822	tx90p_Apr	6	MS	IBGE	Apr	extreme_heat
988	36.94694	temp_Jun	7	MS	IBGE	Jun	mean_air_temp
989	35.53650	prcptot_Apr	8	MS	IBGE	Apr	mean_precipitation
990	28.44460	txx_May	9	MS	IBGE	May	extreme_heat
991	27.73175	dtr_Feb	10	MS	IBGE	Feb	mean_air_temp
1088	100.00000	txn_Feb	1	PR	IBGE	Feb	extreme_heat
1089	98.37288	temp_Apr	2	PR	IBGE	Apr	mean_air_temp
1090	96.48529	txn_Apr	3	PR	IBGE	Apr	extreme_heat
1091	92.21233	temp_Mar	4	PR	IBGE	Mar	mean_air_temp

1092	90.94547	txx_Apr	5	PR	IBGE	Apr	extreme_heat
1093	88.86999	temp_May	6	PR	IBGE	May	mean_air_temp
1094	88.01913	tnx_Mar	7	PR	IBGE	Mar	extreme_heat
1095	87.86547	tnx_Ago	8	PR	IBGE	Ago	extreme_heat
1096	85.80402	txn_Mar	9	PR	IBGE	Mar	extreme_heat
1097	84.62325	temp_Feb	10	PR	IBGE	Feb	mean_air_temp
1186	100.00000	txx_Ago	1	SP	IBGE	Ago	extreme_heat
1187	64.92772	tr_Ago	2	SP	IBGE	Ago	extreme_heat
1188	62.51299	temp_Jul	3	SP	IBGE	Jul	mean_air_temp
1189	43.94459	tnx_Ago	4	SP	IBGE	Ago	extreme_heat
1190	27.14802	txn_Apr	5	SP	IBGE	Apr	extreme_heat
1191	26.54685	temp_Ago	6	SP	IBGE	Ago	mean_air_temp
1192	23.82096	txx_Jul	7	SP	IBGE	Jul	extreme_heat
1193	23.60973	spei_3month_May	8	SP	IBGE	May	drought
1194	22.30100	dtr_Feb	9	SP	IBGE	Feb	mean_air_temp

Table S2. 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

Ai, Z. and Hanasaki, N.: Simulation of crop yield using the global hydrological model H08 (crp. v1), Geoscientific Model Development, 16, 3275–3290, 2023.

Breusch, T. S. and Pagan, A. R.: A simple test for heteroscedasticity and random coefficient variation, Econometrica: Journal of the econometric society, pp. 1287–1294, https://doi.org/10.2307/1911963, 1979.

Cleveland, W. S., Grosse, E., and Shyu, W. M.: Local regression models, in: Statistical models in S, pp. 309–376, Routledge, 2017.

de Geografia e Estatística, I. B.: Produção Agrícola Municipal 2022, http://www.sidra.ibge.gov.br/bda/pesquisas/pam, 2022.

Harri, A., Coble, K. H., Ker, A. P., and Goodwin, B. J.: Relaxing heteroscedasticity assumptions in area-yield crop insurance rating, American Journal of Agricultural Economics, 93, 707–717, https://doi.org/10.1093/ajae/aar009, 2011.

Hijmans, R. J.: terra: Spatial Data Analysis, https://CRAN.R-project.org/package=terra, r package version 1.7-29, 2023.

Iizumi, T. and Sakai, T.: The global dataset of historical yields for major crops 1981–2016, Scientific Data, 7, 97, https://doi.org/10.1038/s41597-020-0433-7, 2020.

Iizumi, T., Yokozawa, M., Sakurai, G., Travasso, M. I., Romanenkov, V., Oettli, P., Newby, T., Ishigooka, Y., and Furuya, J.: Historical changes in global yields: major cereal and legume crops from 1982 to 2006, Global ecology and biogeography, 23, 346–357, https://doi.org/10.1111/geb.12120, 2014.

Liu, Y. and Ker, A. P.: 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, 194–203, https://doi.org/10.1017/aae.2019.40, 2020.

Lundberg, S.: A game theoretic approach to explain the output of any machine learning model, https://github.com/shap/shap, [Accessed 30-10-2023], 2023.

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

McLeod, A.: Kendall: Kendall Rank Correlation and Mann-Kendall Trend Test, https://CRAN.R-project.org/package=Kendall, r package version 2.2.1, 2022.

Monfreda, C., Ramankutty, N., and Foley, J. A.: Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000, Global biogeochemical cycles, 22, https://doi.org/10.1029/2007GB002947, 2008.

Ozaki, V. A., Goodwin, B. K., and Shirota, R.: Parametric and nonparametric statistical modelling of crop yield: implications for pricing crop insurance contracts, Applied Economics, 40, 1151–1164, https://doi.org/10.2307/1907187, 2008.

Rodrigues, M., Corá, J. E., Castrignanò, A., Mueller, T. G., and Rienzi, E.: A Spatial and Temporal Prediction Model of Corn Grain Yield as a Function of Soil Attributes., Agronomy Journal, pp. 1878–1887, https://doi.org/10.2134/agronj2012.0456, 2013.

Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A., Boote, K. J., Folberth, C., Glotter, M., Khabarov, N., et al.: 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, 3268–3273, https://doi.org/10.1073/pnas.1222463110, 2014.

Sacks, W. J., Deryng, D., Foley, J. A., and Ramankutty, N.: Crop planting dates: an analysis of global patterns, Global ecology and biogeography, 19, 607–620, https://doi.org/10.1111/j.1466-8238.2010.00551.x, 2010.

Sinnathamby, S., Douglas-Mankin, K. R., and Craige, C.: Field-scale calibration of crop-yield parameters in the Soil and Water Assessment Tool (SWAT), Agricultural water management, 180, 61–69, https://doi.org/10.1016/j.agwat.2016.10.024, 2017.

Tolhurst, T. N. and Ker, A. P.: On technological change in crop yields, American Journal of Agricultural Economics, 97, 137–158, https://doi.org/stable/24477005, 2015.

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

Wang, J., Wei, J., Shan, W., and Zhao, J.: Modeling the water-energy-food-environment nexus and transboundary cooperation opportunity in the Brahmaputra River Basin, Journal of Hydrology: Regional Studies, 49, 101 497, https://doi.org/10.1016/j.ejrh.2023.101497, 2023.

Yang, S.-R., Koo, W. W., and Wilson, W. W.: Heteroskedasticity in crop yield models, Journal of Agricultural and Resource Economics, pp. 103–109, https://www.jstor.org/stable/40986743, 1992.

Zhu, Y., Goodwin, B. K., and Ghosh, S. K.: Modeling yield risk under technological change: Dynamic yield distributions and the US crop insurance program, Journal of Agricultural and Resource Economics, pp. 192–210, https://www.jstor.org/stable/23243141, 2011.