BIS Working Papers

No 1278

Soybean yield prediction in Argentina using climate data

by Emiliano Basco, Diego Elías, Maximiliano Gómez Aguirre and Luciana Pastore

Monetary and Economic Department

July 2025

JEL classification: Q10, Q12, C13, C32, C33

Keywords: soybean yields, Argentina, forecasting, model selection

This publication is available on the BIS website (www.bis.org).

ISSN 1020-0959 (print)
ISSN 1682-7678 (online)

Soybean yield prediction in Argentina using climate data*

Emiliano Basco†, Diego Elías†, Maximiliano Gómez Aguirre†, Luciana Pastore†

July 2025

Abstract

Agriculture, and especially soybean production, has a critical role in Argentina's economy, as a major contributor to GDP and export revenue. This paper studies the impact of climate variability on soybean yields in Argentina using a novel department-level dataset spanning 1980–2023. We estimate a fixed effects spatial error model (SEM) to quantify the long-run effects of weather shocks—measured by extreme heat, precipitation, and ENSO phases—while controlling for economic and technological factors such as seed technology and relative prices. Our results show that extreme heat significantly reduces yields, while moderate rainfall boosts them up to a nonlinear threshold. El Niño phases increase yields, whereas La Niña events are detrimental. Technological adoption and favorable price signals also enhance productivity. These findings highlight the importance of accounting for both climatic and spatial dynamics when analyzing agricultural outcomes. The model provides a strong empirical basis for forecasting soybean yields and informing policy decisions under increasing climate uncertainty. These models can be employed as effective tools for anticipating yield outcomes under different climate scenarios and utilized in climate-related stress exercises. This work provides valuable insights for policymaking decisions, contributing to prepare for potential economic impacts stemming from climate risks on Argentina's agricultural sector.

Keywords: Soybean Yields, Argentina, Forecasting, Model Selection

JEL Code: Q10, Q12, C13, C32, C33

---

*We are grateful to Tamma Carleton and Torsten Ehlers for their valuable comments and suggestions, which have greatly improved this work. We also thank Pablo Eduardo Abbate, Research Director at the Instituto Nacional de Tecnología Agropecuaria, and Cristian Russo, Chief of GEA at the Bolsa de Comercio de Rosario, for their insightful contributions and sectoral expertise. We appreciate the feedback received from participants at the 14th BIS CCA Research Conference, Bogota, 3 December and at the Research Department's internal seminar at the Central Bank of Argentina (BCRA). The views expressed in this paper are those of the authors and do not necessarily reflect those of the BCRA or its authorities.

† Economic Research, BCRA. E-mail: ebasco@bcra.gob.ar, delias@bcra.gob.ar, mauricio.gomezaguirre@bcra.gob.ar and luciana.pastore@bcra.gob.ar

1. Introduction

Agriculture is a fundamental sector within the Argentine economy. In 2021, primary agriculture and agrifood value chains contributed an estimated 16% to the nation's GDP (World Bank, 2024). Agricultural exports account for around 60% of the country's total exports, emphasizing the sector's importance in generating foreign currency and supporting economic stability.

The soybean complex is the main agricultural production chain in Argentina, characterized by a strong export profile. It involves the industrialization of primary grain production and represents the country's largest export chain, accounting for nearly 28% of total exports in 2022, surpassing both the cereal and automotive sectors. Soybeans are the second most important agricultural crop, following corn, averaging a production of 46 million tons annually over the last five campaigns (2018-2022). Another important aspect of the soybean sector is its volatility, since it experiences high production variability from year to year compared to other crops.

Argentina plays a critical role in global agricultural markets, being the third largest producer and exporter of soybeans worldwide and the top exporter of soybean oil and soybean meal. Oilseed exports are a significant component of Argentina's trade, representing over 31% of total exports in recent years (e.g., 2021 and 2022). This sector's significant contribution reflects its importance for both domestic economic stability and global food markets.

Climate events such as droughts in the Pampas Region have severe repercussions, decreasing yields of essential crops such as soybeans, corn, and wheat, and consequently affecting their entire supply chains. The impacts of these declines also influence multiple sectors within the economy. Reductions in agricultural production caused by droughts have a spillover effect, with implications for upstream and downstream industries.

The agricultural sector's vulnerability to droughts also has broader economic consequences. A downturn in crop yields reduces fiscal revenue, lowers international reserves, and puts pressure on the exchange rate, impacting Argentina's economic resilience.

For example, the recent drought in 2022-2023 had a significant impact on the yield and production of Argentina's main crops: wheat, corn, and soybeans. The effects were felt directly in agricultural production and, through the agribusiness chain, extended to other sectors. The substantial decline in exports also created a shock to government revenue and the balance of payments due to its influence on the exchange rate and international reserves.

Two major climatic phenomena that impact moisture conditions and can cause droughts in the productive regions are La Niña and El Niño. La Niña results from cooling along the equatorial Pacific Ocean, leading to lower-than-average rainfall and causing droughts of varying severity across South American regions. In contrast, El Niño is a climate event marked by the warming of the equatorial Pacific Ocean, which tends to bring above-average precipitation in Argentina. In this paper we consider the influence of El Niño and La Niña indicators, along with the rest of the climate information.

In this paper, we examine the impact of climate variability on soybean yields in Argentina by combining a unique department level dataset covering 1980–2023 with daily weather observations—temperatures and precipitation—aligned to the crop's phenological calendar. We augment these climatic controls with economic and technological variables, including international soybean and fertilizer prices, transgenic seed adoption rates, and land use change indicators. We employ spatial panel econometric techniques to capture both local fixed effects and spatially correlated unobservables, improving our understanding of how climate and regional interdependencies shape yield outcomes.

Our empirical strategy unfolds in two stages. First, we estimate a fixed effects spatial error model (SEM) panel to quantify the long run relationship between weather shocks—measured by growing degree days, cumulative hours above 30 °C, precipitation, and El Niño and La Niña indicators—and soybean yields, while controlling for economic and technological covariates. We find that extreme heat consistently dampens yields, moderate precipitation increases them up to a nonlinear threshold, and lagged El Niño anomalies (warm phases) enhance yields while La Niña's dry conditions are detrimental. Technological progress, as proxied by transgenic seed use and favorable price ratios, further elevates yields, and the highly significant spatial error term confirms the importance of unobserved, regionally correlated factors.

In the second stage, we translate these insights into forecasting exercises by constructing different time series models based on the available weather information. We evaluate forecast performance via rolling window exercises and the Giacomini–White test, benchmarking against the USDA's monthly yield forecasts with same data availability. We find early information models that systematically outperform USDA forecasts, while those models with information closer to the end of the growing season, our forecasts are as reliable as USDA ones.

These findings carry important policy implications. By demonstrating that climate driven models can deliver accurate yield projections months before harvest, we provide a tool for anticipating foreign exchange pressures and informing monetary and fiscal policy responses. Our framework also lays the groundwork for constructing climate stress scenarios—simulating droughts, heat waves, or ENSO events—to evaluate potential economic vulnerabilities in Argentina's agricultural sector. Finally, because the methodology relies on spatially disaggregated yield and weather data, it can be readily adapted to other crops or regions, underscoring the broader value of high resolution climate data in modern agricultural risk management and policy planning.

This paper contributes to the understanding in the relationship between climate variability and soybean yields in Argentina. We introduce a large data set that leverages georeferenced, delegation-level records to uncover both local and regional climate–yield relationships and to account for spatial interdependencies among production zones. This new data set allows us to get a spatial error panel model that reveals how extreme heat, nonlinear precipitation, and ENSO phases affect soybean yields, while capturing technological gains. Finally, we provide statistically robust and accurate forecasts benchmarked against USDA projections.

The paper is structured as follows: first, we present a literature review related to the influence of climate on agriculture. In Section 3, we present some concepts that are crucial to understanding soybean production in Argentina, along with our data sources, treatment of the data, and the calculations involved in constructing the dataset. Section 4 contains an overview of the econometric methodology employed, and we present the results of our analyses. The last part concludes and discusses the implications of our findings and recommendations for future research.

2. Related Literature

A number of studies examine the effects of climate variations on crop yields. For example, Schlenker and Roberts (2009) studied the impact of variations of temperature on corn, soybeans, and cotton yields in the United States, using a panel data approach with nonlinear temperature effects. They found that extreme heat has significant negative impacts on yields, and defined heat thresholds for each crop, above which productivity is reduced.

Deschênes and Greenstone (2007) explored the effects of annual variations in temperature and precipitations to estimate their influence on agricultural profits, using county level panel data in the United States. They estimated long-run effects and found heterogeneous and overall small outcomes across the US.

Chen et al. (2013) conducted an analysis for corn and soybean yields in China, using spatial panel econometric techniques. They modeled the behavior of profit-maximizing farmers, and estimated changes on yields using climate variables, socioeconomic variables and variables representing farmers' adaptation behaviors. They also found nonlinear and asymmetric relationships between yields and climate variables.

Miao, Kanna and Huang (2016) examined the effects of climate variables and crop prices on corn and soybean yields and acreage in the United States, using a large spatial panel dataset. They found that prices have statistically significant effects on corn yields, and that the potential effect of climate change on production is negative but highly heterogeneous, depending on climate scenarios and models.

For Argentina, Cornejo and Ahumada (2021) analyzed the long-term relationships between climatic, technological, and economic factors and crop yields. They found that soybean yields adjust to technological improvements and that high temperatures have a negative effect. They also found evidence of CO_2 fertilization. Cornejo (2021), using information on climate variables published in advance and with a frequency higher than soybean yields, found that there are forecasting gains when considering the maximum temperature during the plant's growth cycle. Additionally, when using precipitation data, the model based on annual data outperforms the others.

These studies present different methodologies to assess the relationship between climate and agricultural productivity. Although technological advancements have improved overall yields, climate events, particularly droughts, continue to have a negative influence on agricultural production.

3. Descriptive Analysis and Data

3.1. Farmers production function

Our study employs a production function as a framework to estimate the effects of climate conditions on yields. This method can be used to isolate the impact of weather on specific crops, while it is also necessary to capture the behavior of farmers and take into account for the range of compensatory responses they perform in reaction to climate variability.

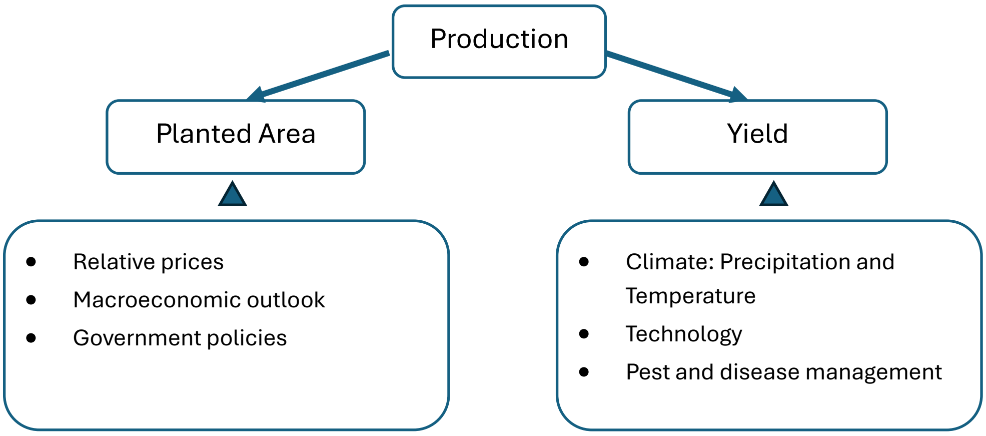
Soybean production, in particular, is determined by two key factors: the area planted and the yields achieved. The decision regarding the area to be cultivated is made in advance, based on expectations of profitability and risk. Producers take into account crop prices, macroeconomic conditions, and potential funding constraints.

On the other hand, yield variability depends on multiple factors, including the application of inputs (fertilizers, pesticides), the adoption of advanced agricultural technologies, and the climatic conditions during the growing season. Price variables can also influence farmers' decisions on input use and production intensity.

Our study focuses on yield variability as it is the variable which is more affected by climate shocks. While changing weather patterns, temperature and precipitation play a fundamental role in determining productivity, farmers decide the planted area before knowing the full extent of weather conditions.

Nevertheless, while climate conditions are a central focus of our study, we must also consider other determinants of yield variation, such as market prices, technological adoption, and land-use changes. By integrating these factors into our empirical framework, we aim to provide a more comprehensive understanding of how weather events influence agricultural production in a setting where adaptation strategies are critical.

Figure 1: Farmers' Production decision



3.2. Soybean Phenological Calendar

The phenological calendar in agriculture describes the different stages of plant development. For example, planting, germination and harvesting are part of the growing cycle of soybean plants. Each crop and region have a unique phenological calendar. This helps understand how climatic conditions influence plant development at critical stages, ultimately affecting yields.

For first-crop soybeans in Argentina, the sowing period runs up until November, the germination and grain-filling period goes from November to February or March, depending on if it's first or second crop soybeans. Harvesting starts in March in some regions, and extends through June.

Monitoring and understanding these stages are essential for both producers and researchers, as this knowledge allows them to anticipate potential climatic impacts on crop yields. Climatic events during these key periods can significantly influence agricultural production.

Table 1: Soybean 1st Crop Calendar

	Jul t-1	Aug t-1	Sep t-1	Oct t-1	Nov t-1	Dec t-1	Jan t	Feb t	Mar t	Apr t	May t	Jun t
Pre-Sowing	x	x										
Sowing			x	x								
Germination-Grain filling					x	x	x	x				
Harvest									x	x	x	x

For this study we consider years starting in July and ending in June, for example for the production harvested in 2022, we consider the climate variables from July 2021 to June 2022.

3.3. Production data

The Secretariat of Agriculture, Livestock, and Fisheries (SAGyP) publishes an annual series of agricultural statistics by crop, campaign, province, and department of the Argentine Republic. They include data on production and sown and harvested areas, permitting yield calculations.

Regarding soybean cultivation, information has been available since 1971. Starting with the 2000/01 campaign, the information on soybeans is detailed according to whether it corresponds to first or second crop. However, we consider only the information corresponding to 1981 onwards, due to the unavailability of daily climate variables before 1980.

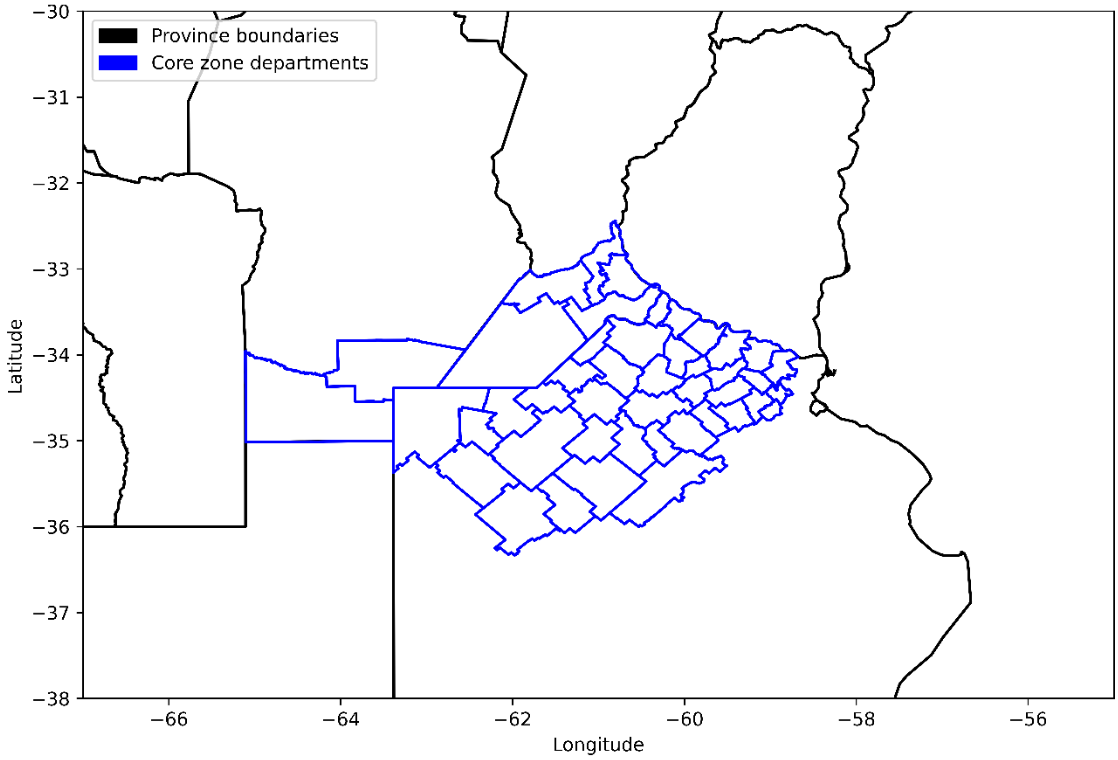
We analyze total soybean production nationwide, capturing total production in all provinces. Yield variations reflect diversity in climate, soil, and phenological calendars.

The territorial division from the SAGyP, known as delegations network is used as a reference. Departments are the second-order political division available in Argentina. Delegations group departments in broader zones, sometimes covering an entire province. This division does not comprise the entire country, as it does not include the Patagonian region. However, it does include all the departments that inform soybean production in the period studied. For this study, 32 of the 40 delegations available are considered, where the delegations excluded have negligible soybean production, and most years no soybean production overall[1].

On the other hand, the core zone is isolated, which includes parts of the provinces of Buenos Aires, Santa Fe, and Córdoba in the Pampas region. These are the best lands for soybean production (Figure 2). This area is critical within the humid Pampas region and covers 45 departments of high relevance in Buenos Aires, Córdoba, and Santa Fe provinces. This subdivision was used to construct climate indicators used in the time series models.

[1] The delegations area division is used to construct a panel dataset, as will be detailed below.

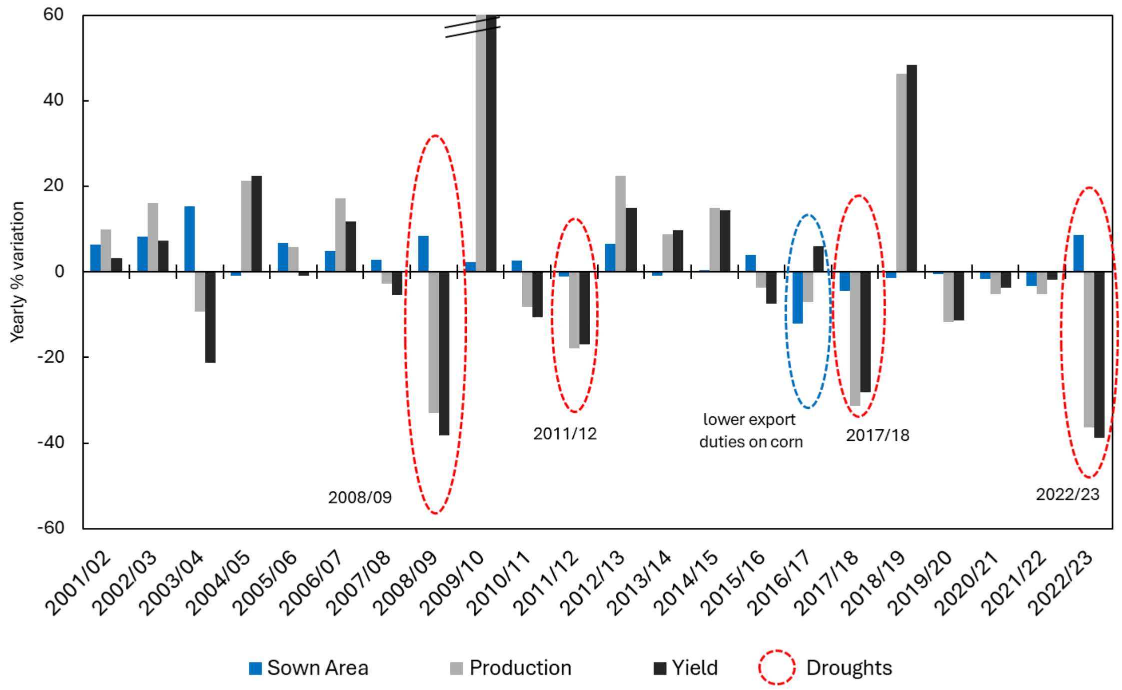
Figure 2: Map of the soybean core zone region



Source: Instituto Geográfico Nacional (IGN)

Drought years have led to heightened yield volatility, which, in turn, has had significant macroeconomic implications. These fluctuations in yield impact participants across the soybean supply chain and, ultimately, the country's GDP.

Figure 3: Soybean sown area, total production and yield yearly variations



Source: Secretariat of Agriculture, Livestock, and Fisheries (SAGyP)

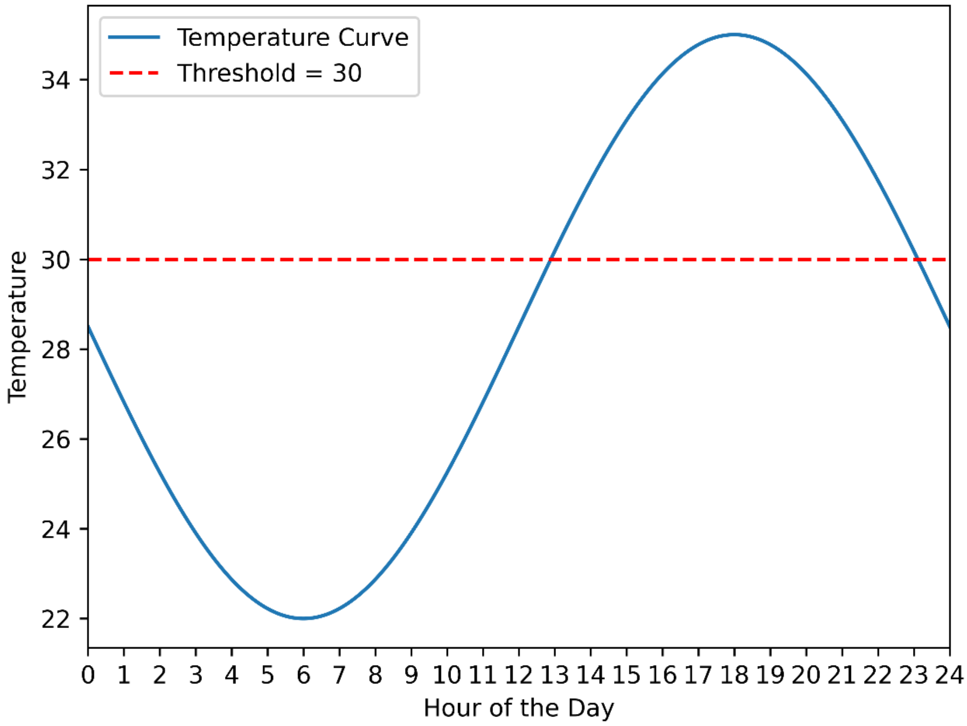
3.4. Historical weather data

The meteorological data used in this study is obtained from the National Meteorological Service (SMN) of Argentina, comprising daily records of precipitation, minimum temperatures, and maximum temperatures for each SMN station. This dataset starts in January 1980.

Given the need for monthly data in the analysis, the daily meteorological records are aggregated to generate monthly series for each station. This includes calculating monthly minimum and maximum temperatures and the cumulative precipitation for each month.

For maximum temperatures, we construct an additional variable that takes into account the amount of time the crops have been theoretically exposed to extreme heat. To achieve this, we fit a sine curve to daily minimum and maximum temperatures (Baskerville & Emin, 1969). This allows a calculation of time above a certain threshold for each day considered. Since soybean yields have been found to be sensitive to temperatures above 30ºC (Schlenkler & Roberts, 2009), we compute the hours above that temperature for each month from November to March.

Figure 4: Fitted sine curve for a day with minimum and maximum temperatures 22ºC and 35ºC

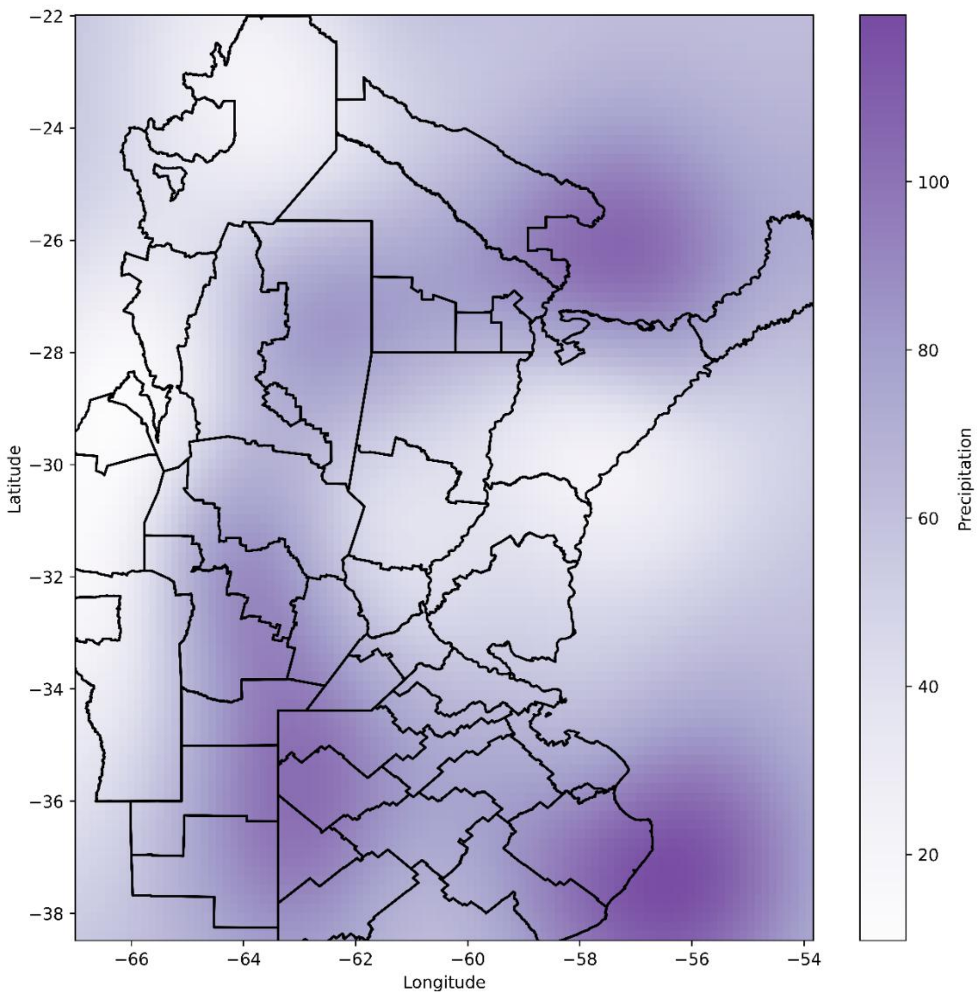


We calculate an additional temperature indicator, Growing Degree Days (GDD). This indicator, widely used in agriculture estimations, measures heat units during the growing season of the plant. It is usually computed as the sum of truncated degrees between two bounds (Schlenker and Roberts, 2009), defined as the theoretical beneficial heat segment for the plant. In this case we use 30º Celsius as the boundary for harmful heat, and 8º Celsius as a floor for heat computation. The GDD data is estimated using daily maximum and minimum temperatures, considering the fitted sine curve in Figure 4.

For each SMN station, we incorporate spatial coordinates and link them to the monthly data, in order to perform spatial interpolation. We use ordinary kriging to estimate the spatial distribution of meteorological variables across the study area, resulting in monthly grids for each variable (minimum temperature, maximum temperature, maximum temperature over 30ºC and precipitation) over the study period. Each grid covers the study area, showing values between meteorological stations based on spatial correlation.

As an illustration, Figure 5 displays the kriging results for November 2022, with the computed precipitation values across the grid. The map of delegations is shown for reference, where the irregularities in some delegations' borders show the departments without any history of soybean production. Department-level geographical data were obtained from the National Geographic Institute (Instituto Geográfico Nacional), which provides a CSV file containing department boundaries in spatial coordinates.

Figure 5: Spatial interpolation of monthly precipitation- November 2022

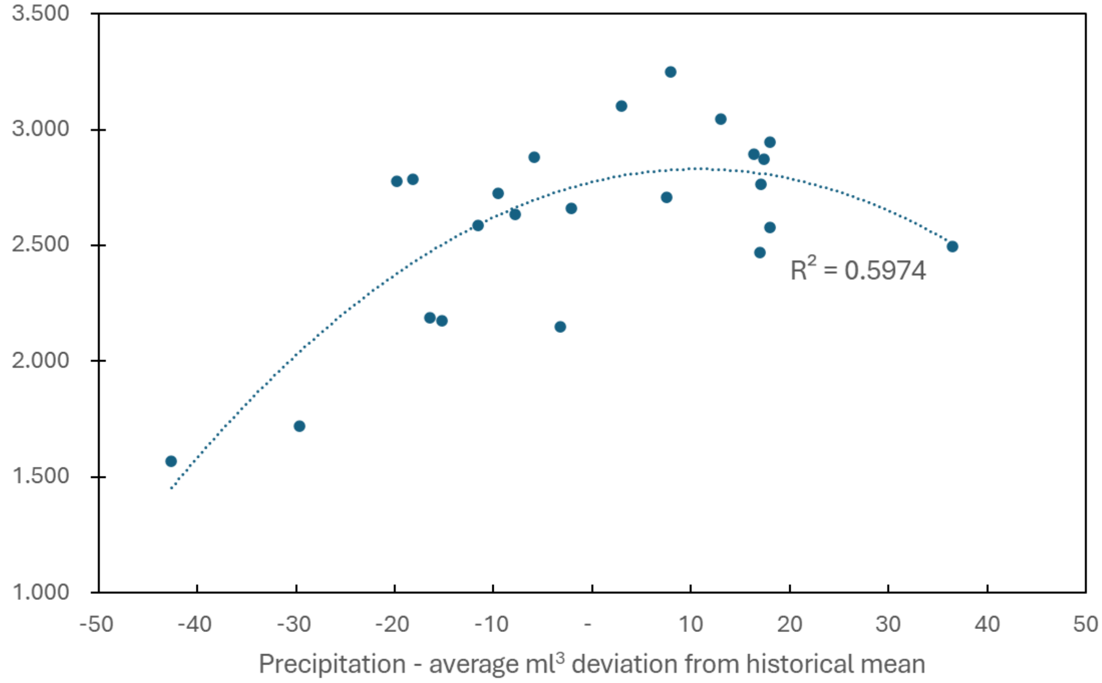


Sources: SMN, IGN, own calculations

To link the meteorological data with political divisions, a cross-referencing process is employed. Initial attempts to match the kriging grids with department boundaries were made, but some smaller departments were not sufficiently represented due to the grid cell size exceeding certain department areas. Instead, broader delegation divisions are selected as the primary spatial units for analysis. A Python program is used to cross-reference the data, matching the grid points to their respective delegations. A panel dataset is then constructed, consisting of 43 years of data across all delegations.

In preliminary findings, a significant relationship was observed between precipitation deviations for each growing season (relative to historical averages) and yields. This suggests that variations in precipitation strongly influence crop yields (Figure 6). Additionally, precipitation contributions appear to show diminishing marginal returns, with a point beyond which further precipitation no longer raises yields.

Figure 6: average precipitation deviation from historical mean vs soybean yields (tn per hectare)



3.5. La Niña / El Niño

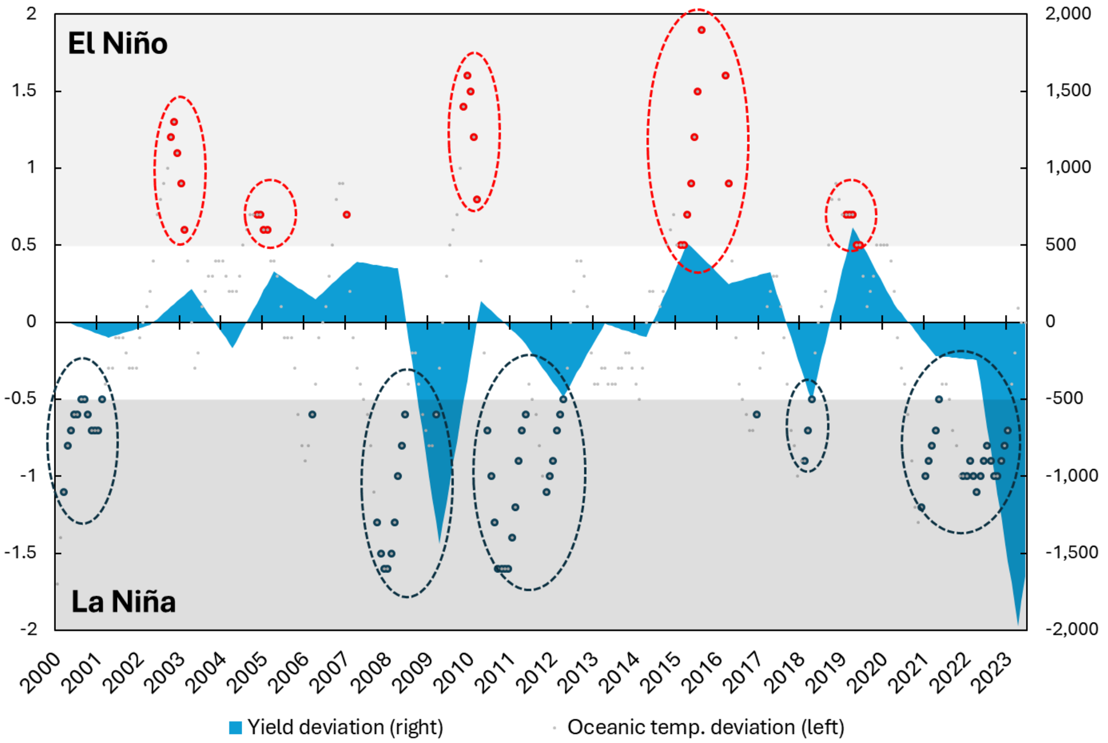
Understanding the effects of El Niño and La Niña is crucial when analyzing Argentina's climate conditions and their impact on soybean production. These phenomena significantly influence precipitation patterns in various regions worldwide, which, in turn, affect crop yields. Although these shifts can vary slightly between El Niño events, the most intense changes typically remain consistent across specific regions and seasons.

The El Niño–Southern Oscillation (ENSO) is monitored through sea surface temperature (SST) anomalies in the equatorial Pacific Ocean. The primary measure used is the Oceanic Niño Index (ONI), which averages SST anomalies in the Niño 3.4 region (170°W–120°W).

El Niño conditions occur when SST anomalies exceed 0.5°C for five consecutive months, indicating a warming phase. La Niña conditions are characterized by SST anomalies below –0.5°C, signaling a cooling phase. These climate anomalies disrupt global rainfall patterns, affecting various agricultural regions worldwide.

Unlike other soybean-producing regions, Argentina experiences reduced rainfall during La Niña years, leading to drought conditions that significantly impact crop yields. Historical observations confirm a negative correlation between La Niña events and soybean production in Argentina, as shown in Figure 7. This underscores the climatic vulnerability of Argentine agriculture to ENSO fluctuations.

Figure 7: Relationship between changes in soybean yield and El Niño/La Niña episodes



We also explore whether early signals of La Niña events—typically available before the sowing season—have any measurable impact on soybean planting decisions, particularly in terms of total planted area. The evidence suggests that, in the Argentine context, anticipated La Niña conditions do not significantly affect the area sown with soybeans. As a result, the primary channel through which La Niña affects soybean production in Argentina is through its adverse effects on yields. This reinforces the relevance of focusing on yield responses—rather than acreage—as the key margin of adjustment when modeling climate impacts and building predictive models.[2]

A key economic mechanism that could mitigate the impact of lower yields is price adjustment, which may dampen the broader macroeconomic effects. If La Niña–induced droughts were to significantly reduce global soybean supply, international prices could rise, partially offsetting farmers' income losses. This, in turn, could help stabilize export revenues, consumption, and investment. However, this adjustment mechanism is less straightforward in the case of Argentina. The main reason is that La Niña does not affect all major soybean-producing regions equally. While Argentina tends to suffer from droughts and yield losses, Brazil and parts of the United States often experience favorable growing conditions, leading to increased output in those regions

[2] In meetings with specialists from the agricultural sector, they were specifically asked whether the risk of El Niño/La Niña events was taken into account in planting plans (such as the choice of crop mix, the more or less intensive use of inputs, etc.), as this could introduce some form of endogeneity. Their response was negative.

(see Figure 8). As a result, global supply may not fall significantly, and price increases are limited. This is consistent with the evidence that the ENSO index does not systematically anticipate international soybean prices.

Figure 8: El Niño and La Niña Teleconnections Map (Lenssen, Goddard & Mason, 2020)

Figure 8.a. El Niño

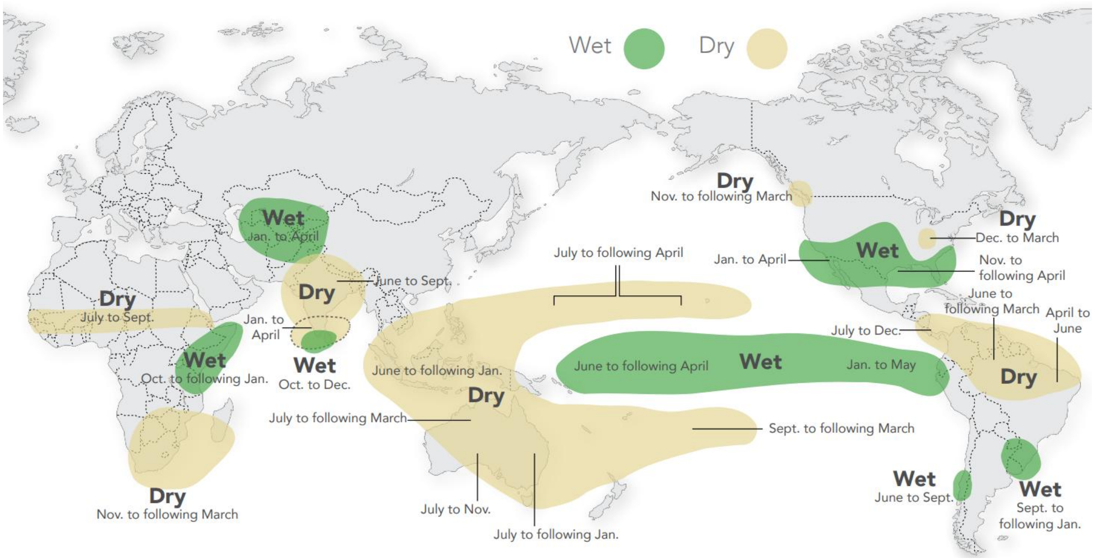
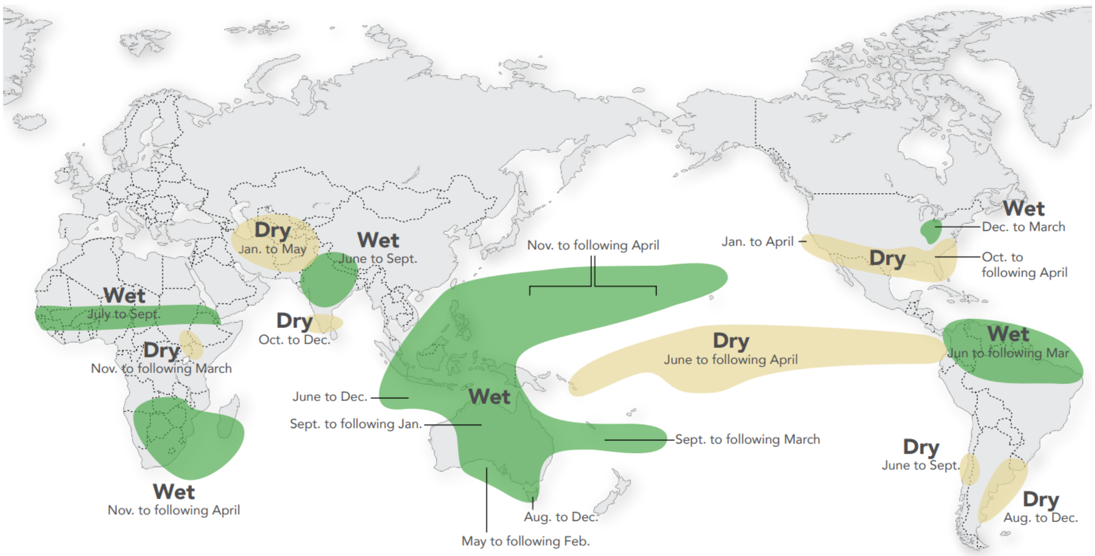


Figure 8.b. La Niña



3.6. Technology and other variables

In addition to climate effects, it is necessary to consider the context of soybean cultivation in Argentina. This section describes some factors that allowed its expansion in volume and territory, which have led to the incorporation of specific variables in the models.

Commercial soybean production began in Argentina in the 1970s. According to Cadenazzi (2009), soybean was initially introduced as a second crop after wheat, as its ease of management and adaptability represented a more profitable rotation, replacing the traditional agriculture/livestock rotation.

In the 1980s, soil erosion effects were observed in the Pampas region due to intensive agricultural activities. However, by the mid-1990s, soybean production expanded with technological advancements that also allowed for the use of lower-quality land both in the core zone and outside of it, expanding the agricultural frontier and increasing yields.

Internationally, a sustained increase in soybean demand was observed, driven by population growth and the need for food, as well as the growing demand for other uses. Better prices in the global market also stimulated supply. As mentioned earlier, before planting begins, producers will define the area to be cultivated, taking into account the market context and expectations. For this reason, the international price of soybeans and its recent volatility were incorporated as variables.

The introduction of transgenic soybean in 1996, specifically RR soybean, marked a turning point in agricultural production in Argentina. This variety of soybeans is resistant to the broad-spectrum herbicide glyphosate, which eliminates all weeds. This translates into cost reductions as it is easy to apply.

The practice of no-till farming also spread during this period. It involves eliminating plowing while the residue from the previous harvest conserves moisture and serves as fertilizer. Sowing is carried out with specially designed machines, with minimal soil disturbance. This method reduces fieldwork time and soil erosion.

No-till farming, along with the use of RR soybean and glyphosate, complement each other since the former leads to an increase in weed quantity. These techniques, along with the development of new agrochemicals such as herbicides, pesticides, and fertilizers, as well as the development of specific machinery, contributed to the expansion of soybean production, as they increase yields and reduce production costs. This makes it feasible to produce in areas that were previously not viable, resulting in increases in both yields and cultivated areas.

We incorporated the price and volatility of Diammonium Phosphate (DAP), a fertilizer commonly used for soybean production, as a signal of the variable costs for agricultural producers. Additionally, a variable representing the proportion of the planted area where transgenic varieties were established was included as an indicator of the technological advancement in the production sector, as this represents a shift in the soybean production model. Information on technological advancement in each delegation was not available.

With respect to land use, following Chen et al. (2013), we constructed an indicator of land use change, computing the year-on-year reduction in the planted area of crops other than soybean inside each delegation. This indicates the portion of cultivable land that can be made available to soybean production each year, and eventually have an effect on yields.

4. Methodological Approach and Results

4.1. Strategy for Model Selection

In order to develop a tool with robust predictive capacity to predict soybean yields in Argentina, we explore different specifications using delegation-level data from across the entire country. These specifications differ based on the type of variable used, the frequency of the variables incorporated, the combination of frequencies (monthly, annual, and quarterly), and the construction of indicators from the same variable. This generates a multiplicity of data structures that can be used to predict yield, each of which requires different estimation methods.

Our starting point is a panel data model designed to understand the underlying drivers of yield variation. While it is crucial to understand how climatic variables directly affect soybean yields in Argentina, these factors operate within an economic and financial context that can mitigate or exacerbate climatic effects. By detecting the most relevant variables in both the climate and the economic and financial aspects—and by examining their specific distributions—we are able to inform the configuration of alternative time series model specifications.

4.2. Panel Data Model: Specification and Results

In this section, we present a detailed analysis of our panel regression methods and results. First, we test the panel unit root hypothesis, and then we proceed to estimate the proposed model using the fixed effects estimator for a stochastic production function. All panel unit root tests[3]—across multiple specifications—consistently reject the null hypothesis of a unit root, indicating that the series are stationary in levels and do not require differencing prior to estimation.

Second, we estimate the proposed model using the fixed effects estimator for a stochastic production function. This approach relies in the production function mentioned in Section 3.1, along with production incentive variables, to achieve impact estimates by varying one or more input variables, such as temperature, precipitation, the sea surface temperature indicator of the Pacific Ocean, and the use of incentive variables like soybean prices and input costs along with their volatility.

[3] Levin, Lin & Chu (LLC) test, Im test, Pesaran & Shin (IPS) and Dickey Fuller

We utilize the fixed effects model for our panel data for two reasons. The main reason is that the fixed effects model allows us to estimate a unit-specific effect for each delegation in the model. Additionally, the fixed effects model does not require the restrictive assumption that the specific effect of the delegation is independent of the included covariates, as is the case with the random effects model (See Appendix 2 for the Hausman test results). Also, we are using the total population (all the soybean producing delegations), which prevents us from needing a random effects estimation to generalize sample data to a broader population.

The dependent variable for this model is the logarithm of soybean yields for each delegation (as described above). We estimate the following regression:

$$\log Y_{it} = Z_{it}\beta + c_i + \varepsilon_{it} \tag{1}$$

$$\varepsilon_{it} = \rho \sum_j W_{i,j}\varepsilon_{jt} + \phi_{it} \tag{2}$$

Where $\log Y_{it}$ indicates log crop yield in delegation i and year *t*. The term $Z_{it}\beta$ includes weather variables, a time trend and quadratic time trend, and other technological, land use change and economic variables. The term c_i is included to account for the delegations' fixed effects, and ε_{it} is the error term.

Following Chen et al. (2013) and Schlenker et al. (2006), we allow for spatial correlation in the error term. In equation 2, ρ is the spatial correlation factor, $W_{i,j}$ is a spatial weighting matrix that identifies neighbors for each delegation, and ϕ_{it} are the error terms that are independently normally distributed with E=0 and variance= σ^2.

Panel models that include an interaction effect in the error term -called SEM (spatial error models)-, indicate that units might have a similar behavior because of shared unobserved characteristics (Elhorst, 2017). In this case, yield might be influenced by factors that are not included in the model, such as regional policies, soil quality or seed varieties used, that are spatially correlated over delegations. We test for the significance of the spatial error term and found it to be statistically significant (see Appendix 3).

The model specification requires complete data for all delegations; consequently, we exclude seven delegations with missing or zero values from the estimation. The excluded delegations add up to 2% of the total production on average.

Table 2: SEM with spatial fixed effects.

	Coefficient	Std. Error	z	P>|z|
Trend	-2,323	0,907	-2,560	0,010
Quadratic trend	0,001	0,000	2,570	0,010
El Niño temperature anomaly (-1)	0,030	0,006	4,590	0,000
Transgenic Soybean	0,328	0,087	3,760	0,000
Ratio of soybean price/fertilizer price (-1)	0,119	0,081	1,470	0,140
LUC indicator	0,087	0,051	1,700	0,088
GDD	-0,000	0,000	-2,540	0,011
GDD over 30ºC	-0,000	0,000	-2,280	0,023
Precipitation	0,002	0,000	7,530	0,000
Precipitation squared	-0,000	0,000	-5,470	0,000
Spatial correlation	0,608	0,028	21,640	0,000
Number of groups = 32				
Panel length = 35				

The SEM panel estimation results are shown in Table 2, and the summary of the effects found is:

- **GDD and GDD over 30ºC:** These variables represent the growing degree days and cumulated hours with temperatures above 30°C during the growing season respectively. Coefficients are both negative, small (due to scaling of the variables), and statistically significant. A negative coefficient associated with the variable GDD over 30ºC suggests that temperatures above 30 degrees negatively impact yields. The coefficient for the variable GDD is also negative, suggesting that the effect of high temperatures is dominant over the summer soybean growing season, having a negative effect on yields overall.

- **Precipitation and Precipitation squared**: These variables reflect the cumulative precipitation and the squared cumulative precipitation between July and March along the same campaign. The coefficients associated with these variables (positive and negative, respectively) show a nonlinear relationship between precipitation and yields: more precipitation increases yields, but beyond a certain threshold, the impact starts to increase at a decreasing rate.

- **El Niño temperature anomaly**: This variable reflects an expanded yearly Oceanic Niño Index (ONI), that shows the presence of the El Niño and La Niña events along with their intensity. The variable is lagged one year, to show the conditions prior and during the early stages of each campaign. The coefficient is positive and statistically significant, suggesting that warm conditions (associated with the El Niño events) are favorable for yield. Conversely, negative values of ONI indexes indicate La Niña events that promote dry conditions over the soybean producing areas of Argentina, and would have a negative effect on yields.

Regarding trend, market and technology variables, we incorporated the following:

- **Trend and Quadratic trend:** The linear and quadratic trends represent long-term effects. The negative coefficient on the trend indicates a slight decrease in yield over time; while the positive quadratic trend suggests that the rate of change in yield may stabilize or even slightly increase in recent years due to factors not included in the estimation.

- **Ratio of soybean price/fertilizer price:** It is the ratio between the soybean price variable and the price of DAP (Diammonium Phosphate fertilizer) on the year prior to sowing. The positive coefficient observed in this ratio indicates potential increases in farmer benefits, which in turn lead to higher yields. This is likely due to farmers trying to increase production per area in response to their perception of higher future benefits.

- **LUC indicator:** The land use change indicator shows the year-on-year reduction in the planted area of crops other than soybean inside each delegation. The resulting coefficient is, although small, indicative of the fact that reductions in areas dedicated to other crops are being shifted to soybean planting.

- **Transgenic Soybean:** The impact of the percentage of transgenic soybean use yields a positive and significant coefficient, indicating that as the use of transgenic soybean in the productive area has increased, it has contributed to improved yields regardless of climatic impacts.

4.3. Time series: model specification and data structure

Using insights from the panel approach, we construct time-series models to evaluate different strategies for predicting Argentina's aggregate soybean yield.

The independent variable remains the logarithm of soybean yields. For the explanatory variables, we generate a time series dataset based on the panel data, considering different aggregation methods. First, we average the delegations' climate variables over the spatially interpolated data used in the panel model for the whole soybean areas. Second, we generate another dataset by averaging directly all soybean-producing zone's meteorological stations records. Lastly, we create another dataset with climatic variables measured in the core zone only, in order to generate a leading indicator. The idea is to explore a uniform region in terms of climate, production, and management.

We then test different combinations of variables. Our first set of time series models (Group G1) uses the same variables as the panel data estimation, where the weather variables were averaged over the whole soybean-producing zone. In the second step, we identify the optimal model—based on entropy—using variables from the panel-model for key month of the growing-season calendar. This yields different time-series models, each corresponding to the weather information available at successive stages (December, January, February, and March).

To provide robustness to the comparison, we also incorporate a second group of models designed to capture different features of the phenomenon at hand. This second set of models (Group G2) contains combinations of variables different from those on group G1, and also takes into account the availability of data for different months. The explanatory variables are the total panel averages for all the soybean producing region.

The third group (Group G3) takes into account different formulations using variables with annual, monthly, and quarterly frequencies, as is the case with ENSO. This mix of frequencies requires the use of different estimation methods such as MIDAS (Mixed Data Sampling[4]). Models in group G3 use as explanatory variables the weather averages using indicators measured within the core zone only. The independent variable is the same as the previous groups, the logarithm of the total soybean yield in Argentina.

4.4. Forecast comparison

Our model selection strategy is driven by the goal of anticipating or outperforming the forecast results published by the US Department of Agriculture (USDA), a globally recognized authority in agricultural production forecasting. Therefore, we consider two dimensions: the ability to anticipate and the ability to predict.

In this context we use the term "anticipate" referring to the comparison between two models with information available at different points in time. A model with inputs that are more recent or closer in time to the event being predicted is expected to have greater predictive capacity than a model with inputs that are more distant in time. When a model with information that is more distant in time predicts a contemporary variable better than a model with more recent information, we say that the latter anticipates the former; even if these two models are not able to differentiate their predictive capacity (they predict the same thing).

On the other hand, if two models' inputs are available at the same time, the model that better predicts the target variable will be considered to have better forecasting power and therefore have better predictive capacity.

To generate the forecasts, once the panel and time series models were defined, we divided the dataset into training and testing segments, a process often used in time-series analysis as setting

[4] See Ghysels, Santa-Clara and Valkanov (2006)

an "estimation window." Within this window, each model is trained to make predictions, projecting ‘h’ steps ahead.

For example, we estimate the model using data from 1989 to 2016 and we obtain a prediction for the soybean yields one year ahead. Then we expand the sample to 1989 to 2017, and we forecast 2018. Iterating this process results in a of RMSE for each model, to be compared with the USDA predictions’ RMSE.

To compare the models, we use the Giacomini and White (2006) test, which allows us to evaluate, compare, and prioritize different models based on their predictive performance. This tool only indicates which model is the best within a set, but it does not provide information on how good the model is in absolute terms. To overcome this limitation, we use the USDA forecasts as a benchmark for comparison.

Our benchmark is derived from a subset of the USDA’s series of predictions in their monthly reports. While USDA issues monthly Oilseeds reports, throughout the year, we take into account the reports published in January to April (the harvest starts in April and lasts until June). The reports are made available in the first days of each month, therefore we consider them to incorporate information up to the end of the previous month. For example, for the forecast included in the January report, we assume it incorporates information up until the last day of December of the previous year.

Table 3 shows some of the results of the Giacomini-With tests. The test results allow us to compare all the forecasts coming from our panel Spatial Error Models (SEM) and Time series models from groups 1, 2 and 3 with published USDA forecasts. While we ran pairwise comparisons for all the available models, in this table we show the ones that outperform USDA forecasts[5].

Table 3: Model comparison (GW results) on models that perform better than USDA forecasts

Model 1	Model 2	Data Availability	Coeff.	Std. Error	t-Statistic	Prob.
SEM_DEC	USDA_Jan_report	DECEMBER	-106,87	56,91	-1,87	0,1095
TS_G2_01_DEC	USDA_Jan_report	DECEMBER	-115,53	65,51	-1,76	0,1283
TS_G3_03_JAN	USDA_Feb_report	JANUARY	-150,24	80,26	1,87	0,1104
TS_G2_06_JAN	USDA_Feb_report	JANUARY	-137,35	72,93	-1,88	0,1087
SEM_FEB	USDA_Mar_report	FEBRUARY	-77,918	43,01	-1,81	0,1201

The first two columns show the models being compared. "Data Availability" refers to the moment when the data becomes available for input into the model. For example, in the first row, the

[5] Results can be made available upon request to the authors.

SEM_DEC, -the spatial error model- and the USDA_Jan_report forecast take into account data available in December. The "Coeff." column shows the coefficient obtained from regressing a constant on the difference in RMSE between models, which is the basis of the Giacomini-White test. The last three columns present statistics from the test.

The first row of the table compares the forecast obtained from a version of the SEM panel shown in Table 2 with the USDA's January forecast, denoted as "USDA_Jan_report". The only difference between the model used in the first row of the Table 3 for the forecasting comparison and that shown in Table 2, is the way in which we incorporate the variable "El Niño temperature anomaly (-1)", in the model.[6]

Both forecasts use data available in December[7]. The test results in a negative coefficient, indicating that, on average, the panel model systematically exhibits a lower RMSE than the USDA forecast based on the same period. This suggests that we can obtain more accurate forecasts with the spatial error panel model than those in the USDA report. Moreover, the coefficient is statistically significantly different from zero (coef. < 0; p-value < 15%; = 10.95), reinforcing the advantage of the panel over the USDA's December forecast.

Similarly, the second row of Table 3 presents the results of comparing one of the Group 2 time-series models for December— TS_G2_01_DEC—with the USDA forecast for the same period. As shown, this version also outperforms the USDA forecast. This model incorporates variables from the entire country, and we attempted to include climate characteristics at the beginning of the campaign, some trend-related aspects during the campaign, and climate volatility that could affect yields. In this way, we incorporate the ENSO annual index observed in the calendar year prior to the campaign as an indicator of climatic conditions ahead of the main part of the growing season. The results show a positive effect, suggesting that higher temperatures are likely associated with increased precipitation and improved soybean yields. In the aim of incorporating tendential aspect, we include a trend, a squared trend (that result in negative and positive coefficients respectively) and the intensity of use of transgenic Soybean. These aspects show a growth in the soybean yields along time. Finally, we incorporate volatility through the deviation from the average of accumulated precipitation and accumulated hours of heat over 30ºC; as we previously expect more precipitations results in better yield, while more extreme temperatures affect yields in a negative way.

---

[6] For forecasting purposes we modify the variable "El Niño temperature anomaly (-1)" because we want to capture different patterns inside the year. For more detail, see Appendix 4

[7] The results of SEM models based on information available for different months previous to the harvest season are shown in Appendix 5.

The forecast obtained using the time series model TS_G3_03_JAN corresponds to a MIDAS model with information from the core zone. We find that it further improves upon the USDA forecast for January information (USDA_Feb_report). The model incorporates precipitation information for two months at the beginning of germination period, specifically the deviation from the historical average of precipitation during December and January. Both coefficients in the regression show positive values, which is an indicator that precipitation over the historical average observed in the core zone improves the soybean yield in the total country.

$$\log(rinde)_t = c + \beta_1(Dev\_hist\_precip\_Dic(t-1)) + \beta_3(Dev\_hist\_precip\_jan(t)) + \sum_{\tau=0}^{7} Temp.Ocean.quarter(-6)^{H}_{\frac{t-\tau}{S}}\left(\sum_{J=0}^{4}\tau^{J}\theta_j\right) + e_t \quad (3)$$

Equation (3) shows the formula for the MIDAS times series model TS_G3_03_JAN. Since ENSO is a quarterly indicator, it is included in the model within the polynomial of variables with a higher frequency. The interesting aspect of this result is that our yield variable, for example for the campaign that begins in 2020, is considered to end in 2021 in our model and is assigned 2021 as the calendar year. Therefore, the yield quantified for 2021 coexists with ENSO data from six previous quarters. These six lags in the calendar period correspond to the third quarter of 2020, meaning that the model identifies the relevant ENSO values as those observed in the quarter preceding the campaign.

As shown in Table 3, the times series model TS_G2_06_JAN outperforms the February USDA report. This model incorporates a tendential component, represented by the intensity of transgenic soybean use over time, and climate volatility, captured through differences between precipitation and its historical average. Additionally, the model places greater emphasis on initial climatic conditions. To account for this, we include the El Niño temperature anomaly at the beginning of the campaign (measured as the annual ENSO index lagged by one period) and the average temperature during the March–June period of the previous campaign. All the coefficients have the expected signs: greater utilization of transgenic soybeans is associated with higher yields; an increase in precipitation above the historical average also leads to higher yields; and a higher probability of El Niño correlates with increased yield. Finally, the increase in average temperature at the end of the previous campaign can be interpreted as reflecting the overall trend rather than extreme weather events.

The spatial error model SEM_FEB is composed of the same variables as the panel data model shown in Appendix 5 and the SEM_DEC model; the only difference with the latter is that the panel model has been estimated using information available up to February. The forecast is then estimated, to be compared with the USDA results.

Another aspect to consider, which was already mentioned at the beginning of this section, is the forecasting (or anticipatory) ability of the models. In this context, a model should be able to outperform the USDA in terms of predictive capacity for a specific month, but this advantage should not be reversed the following month. That is, the model should maintain its performance relative to the USDA even when using less recent input information than what the USDA has available.

Table 4 shows the predictive performance results of the panel spatial error model -SEM_DEC- compared to the USDA's January report, which was reported also in Table 3. The subsequent USDA reports, which include updated information, are not able to provide statistically superior forecasts compared to those generated by the panel.

In this sense, we could say that the panel forecasts using information available in December yield results similar to those obtained by the USDA in the two following months, and therefore, we can say that the panel anticipates the USDA's results.

Table 4: GW results of Spatial error model (SEM) December panel vs USDA report forecasts

Model 1	Data Avail.	Model 2	Data Avail.	Coeff.	Std. Error	t-Statistic	Prob.
SEM_DEC	DEC	USDA_Jan_report	DEC	-106,87	56,91	-1,88	0,1095
SEM_DEC	DEC	USDA_Feb_report	JAN	-21,31	36,47	-0,58	0,5804
SEM_DEC	DEC	USDA_Mar_report	FEB	65,82	43,77	1,50	0,1833

The same occurs with TS_G2_01_DEC, which anticipates up to March, and with TS_G3_03_JAN and TS_G2_06_JAN, which anticipate up to April.

5. Conclusion

This paper studies the impact of climate variability on soybean yields in Argentina using a novel delegation-level panel dataset matched with high-frequency weather data aligned to the crop's phenological calendar. Employing a two-stage empirical strategy, we first estimate a spatial error panel model that captures the effects of weather shocks, economic incentives, and technological adoption on yields. We then use these insights to construct time-series forecasting models—both at the national and regional levels—and benchmark their performance against USDA forecasts.

Our findings confirm that extreme heat significantly depresses yields, while moderate precipitation and warm ENSO phases (El Niño) support them. Technological advances and favorable price ratios also contribute positively to yield growth. The significance of the spatial error term highlights the role of unobserved, regionally correlated factors.

On the forecasting front, we show that climate-based models, even those using only early-season data, can match or outperform USDA projections. The Giacomini–White test provides formal evidence of this predictive edge, particularly during years of heightened climate variability

These results underscore the value of integrating high-resolution climate data into agricultural forecasting models. Improved early-season forecasts can enhance the anticipation of foreign exchange pressures and support more responsive macroeconomic policy. In addition, our framework offers a foundation for designing climate stress scenarios, helping policymakers assess the vulnerabilities of Argentina's agricultural sector in the face of growing climate risks.

References

Ahumada, H., & Cornejo, M. (2021). Are Soybean Yields Getting a Free Ride from Climate Change? Evidence from Argentine Time Series Data. *Econometrics*, *9*(2), 24. https://doi.org/10.3390/econometrics9020024

Baskerville, G. L., & Emin, P. (1969). Rapid Estimation of Heat Accumulation from Maximum and Minimum Temperatures. *Ecology*, *50*(3), 514–517. https://doi.org/10.2307/1933912

Bert, F., de Estrada, M., Naumann, G., Negri, R., Podestá, G., de los Milagros Skansi, M., ... & Quesada, M. (2021). The 2017-18 drought in the Argentine Pampas–Impacts on Agriculture. *United Nations Office for Disaster Risk Reduction (2021). GAR Special Report on Drought*.

Bisang, R., & Sztulwark, S. (2007). Tramas productivas de alta tecnología y ocupación. El caso de la soja transgénica en la Argentina. In *Estructura productiva y empleo: Un enfoque transversal* (2007th ed.). Ministerio de Trabajo, Empleo y Seguridad Social.

Burke and Emerick (2016) Adaptation to climate change: Evidence from US agriculture. American Economic Journal: Economic Policy 8; 106-40

Burke, M., Hsiang, S. & Miguel, E. Global non-linear effect of temperature on economic production. Nature 527, 235–239 (2015).

Cadenazzi, G. (2009). La historia de la soja en Argentina. De los inicios al boom de los 90. In *Ponencia presentada en el XXVII Congreso de la Asociación Latinoamericana de Sociología, VIII Jornadas de Sociología de la Universidad de Buenos Aires, Asociación Latinoamericana de Sociología, Buenos Aires, Argentina* (Vol. 31, pp. 000-062).

Chen, S., Chen, X., & Xu, J. (2013). *Impacts of Climate Change on Corn and Soybean Yields in China*. AAEA & CAES Joint Annual Meeting. Washington DC, August 4-6.

Cornejo, M. (2021). Forecasting crop yields through climate variables using mixed frequency data. The case of Argentine soybeans. *Económica*, *67*, 022. https://doi.org/10.24215/18521649e022

Deschênes, O., & Greenstone, M. (2007). The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather. *American economic review*, *97*(1), 354-385.

Elhorst, J.P. (2017) Spatial Panel Data Analysis. In: Shekhar S., Xiong H., Zhou X. (Eds.) *Encyclopedia of GIS*, 2nd edition, pp. 2050-2058. Springer International Publishing, Cham, Switzerland.

Fisher, A. C., Hanemann, W. M., Roberts, M. J., & Schlenker, W. (2012). The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather: Comment. *American Economic Review*, *102*(7), 3749–3760. https://doi.org/10.1257/aer.102.7.3749

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131(1-2), 59-95.

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*(6), 1545-1578.

Hsiang, S. (2016). Climate econometrics. *Annual Review of Resource Economics*, *8*(1), 43-75..

INDEC Instituto Nacional de Estadística y Censos Informes técnicos. Vol. 7, nº 39 ISSN 2545-6636 Comercio exterior. Vol. 7, nº 4 Complejos exportadores Año 2022 ISSN 2545-6644

Instituto Geográfico Nacional de la República Argentina.IGN. Base de Datos Geoespacial Institucional.Capas SIG

Lenssen, N. J., Goddard, L., & Mason, S. (2020). Seasonal forecast skill of ENSO teleconnection maps. *Weather and Forecasting*, 35(6), 2387-2406.

Lobell, D. B., & Burke, M. B. (2008). Why are agricultural impacts of climate change so uncertain? The importance of temperature relative to precipitation. *Environmental Research Letters*, *3*(3), 034007.

Lobell, D. B., & Field, C. B. (2007). Global scale climate–crop yield relationships and the impacts of recent warming. *Environmental research letters*, *2*(1), 014002.

Lobell, D. B., Schlenker, W., & Costa-Roberts, J. (2011). Climate trends and global crop production since 1980. *Science*, *333*(6042), 616-620.

McCarl, B. A., Villavicencio, X., & Wu, X. (2008). Climate Change and Future Analysis: Is Stationarity Dying? *American Journal of Agricultural Economics*, *90*(5), 1241–1247. https://doi.org/10.1111/j.1467-8276.2008.01211.x

Miao, R., Khanna, M., & Huang, H. (2016). Responsiveness of crop yield and acreage to prices and climate. *American Journal of Agricultural Economics*, *98*(1), 191-211.

Naumann, G., Vargas, W. M., Barbosa, P., Blauhut, V., Spinoni, J., & Vogt, J. V. (2019). Dynamics of Socioeconomic Exposure, Vulnerability and Impacts of Recent Droughts in Argentina. *Geosciences*, *9*(1), 39. https://doi.org/10.3390/geosciences9010039

Satorre, E. H. (2005). Cambios tecnológicos en la agricultura argentina actual. *Ciencia hoy*, *15*(87), 24-31.

Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences*, *106*(37), 15594–15598. https://doi.org/10.1073/pnas.0906865106

Schlenker, W., Hanemann, W. M., & Fisher, A. C. (2006). The impact of global warming on US agriculture: an econometric analysis of optimal growing conditions. *Review of Economics and statistics*, *88*(1), 113-125.

Secretaría de Agricultura Ganadería y Pesca (SAGyP). Dataset Estimaciones Agrícolas

Strada, J., & Vila, I. A. (2022). La producción de soja en Argentina: Causas e impactos de su expansión. *Estudios de Economía Política y Sistema Mundial*, *23*.

Thomasz, E. O., Eriz, M., Rondinone, G., & Vilker, A. (2017). *The economic impact of the extreme climatic events in Argentina* (Doctoral dissertation, Universidad de Buenos Aires. Facultad de Ciencias Económicas.).

U.S. Department of Agriculture. (2016–2024). *Oilseeds: World markets and trade* [Monthly reports, December–April issues]. Foreign Agricultural Service. https://www.fas.usda.gov/data/oilseeds-world-markets-and-trade

World Bank. 2024. Towards a More Competitive, Inclusive and Resilient Agrifood Sector in Argentina. Washington, DC.”

Appendices

Appendix 1: Causality tests

When analyzing two lags, we find that the null hypothesis that prices do not cause production cannot be accepted; however, Argentine production has no effect on prices.

Table A1: DLPROD vs. DLPRECIO

Pairwise Granger Causality Tests (1980 to 2024; 2 lags)

Null Hypothesis:	Obs	F-Statistic	Prob.
DLPROD does not Granger Cause DLPRECIO	31	0.57428	0.5701
DLPRECIO does not Granger Cause DLPROD		3.57279	0.0426

The null hypothesis of no Granger causality from planted area to yields cannot be accepted; whereas, the hypothesis that yields do not Granger-cause planted area cannot be rejected.

On the other hand, the null hypothesis of no Granger causality between planted area and ENSO cannot be rejected in either direction, indicating that the effect of ENSO on yields is isolated from any potential changes in planted area induced by ENSO.

Table A2: L_SUP, ENSO and LOG_RINDE

Pairwise Granger Causality Tests (1989 to 2022; 2 lags)

Null Hypothesis:	Obs	F-Statistic	Prob.
L_SUP does not Granger Cause ENSO	32	0.3726	0.6924
ENSO does not Granger Cause L_SUP		1.0931	0.3495
LOG_RINDE does not Granger Cause ENSO	32	2.6924	0.0859
ENSO does not Granger Cause LOG_RINDE		9.4367	0.0008
LOG_RINDE does not Granger Cause L_SUP	32	0.2634	0.7704
L_SUP does not Granger Cause LOG_RINDE		5.4377	0.0104

There is no evidence of bidirectional Granger causality between the international price and the ENSO index, indicating that they are statistically independent in this context.

Table A3: DLPRECIO vs. ENSO

Pairwise Granger Causality Tests (1989 to 2022; 2 lags)

Null Hypothesis:	Obs	F-Statistic	Prob.
ENSO does not Granger Cause DLPRECIO	32	0.3471	0.7100
DLPRECIO does not Granger Cause ENSO		0.5429	0.5875

Appendix 2: Hausman test for Fixed Effects and Random Effects panels

Since we have data from all relevant delegations (i.e., the sample covers the entire population of regions), panel fixed effects estimations are preferred because they capture the individual heterogeneity of each region without assuming that these differences are random. A random effects estimation would be more appropriate if the data were a random sample from a larger set of regions.

To test the cross-section random effects, we conducted the Hausman Test. The results yielded a Chi-square statistic of 45.666 with 10 degrees of freedom, and a p-value of 1.648e-06.

These results imply that fixed effects should be used. This suggests that the unobserved differences between regions are correlated with the explanatory variables, invalidating the key assumption of random effects.

Appendix 3: Moran and LM tests of spatial error correlation significance

The spatial weight matrix used in the regression and for this test is a spatial contiguity matrix. This spatial contiguity matrix includes pairwise comparisons of all delegations, where the (i, j) element of W is unity if delegations i and j share a common boundary, and 0 otherwise. The matrix is normalized so that the sum of the elements in each row is equal to one.

The Moran test is conducted over the cross-section of the delegations' soybean yields for each year.

Table A4: Moran test results

Year	Moran I	p-value	Year	Moran I	p-value	Year	Moran I	p-value
1989	4,197	2,70E-05	2002	2,639	8,32E-03	2015	3,629	2,85E-04
1990	1,504	1,33E-01	2003	3,100	1,94E-03	2016	4,603	4,15E-06
1991	2,024	4,30E-02	2004	3,549	3,87E-04	2017	0,917	3,59E-01
1992	4,149	3,34E-05	2005	4,519	6,23E-06	2018	3,197	1,39E-03
1993	0,325	7,45E-01	2006	2,271	2,31E-02	2019	2,932	3,37E-03
1994	2,857	4,28E-03	2007	0,898	3,69E-01	2020	3,031	2,44E-03
1995	2,002	4,53E-02	2008	0,925	3,55E-01	2021	2,602	9,26E-03
1996	3,326	8,80E-04	2009	3,061	2,20E-03	2022	2,843	4,47E-03
1997	1,739	8,20E-02	2010	0,262	7,93E-01			
1998	3,894	9,87E-05	2011	3,148	1,64E-03			
1999	2,950	3,18E-03	2012	6,473	9,62E-11			
2000	3,612	3,03E-04	2013	4,096	4,21E-05			
2001	0,232	8,17E-01	2014	0,659	5,10E-01			

Appendix 4: SEM Panel models for forecasting

We develop an alternative variable for all SEM forecasting models (SEM_DEC, SEM_JAN, SEM_FEB and SEM_MAR) that aims to capture within-year temperature patterns more effectively, which may provide valuable insights for yield prediction.

To show the procedure applied, the following table describes the SEM_MAR model. The main difference between the Spatial Error Model (SEM) in Table 2 and the SEM_ MAR model lies in how they incorporate the frequency of the El Niño temperature anomaly.

Specifically, when we refer to the El Niño temperature anomaly (-1) in the SEM Table 2 model, we are referring to the average temperature during the calendar year that starts just before the campaign. In this context, the anomaly (-1) includes temperature information both prior to and during the campaign. While this annual average anomaly provides useful information about the overall impact of temperature on yield, it may also obscure important intra-annual patterns that are relevant for forecasting.

Table A5: SEM March

	Coefficient	Std. Error	z	P>|z|
Trend	-2,748	0,951	-2,890	0,004
Quadratic trend	0,001	0,000	2,900	0,004
El Niño temperature anomaly (Dec(-1))	0,178	0,044	4,070	-
El Niño temperature anomaly (Mar(-1))	-0,180	0,060	-2,990	0,003
El Niño temperature anomaly (June(-1))	0,057	0,038	1,500	0,134
Transgenic Soybean	0,312	0,095	3,290	0,001
Ratio of soybean price/fertilizer price (-1)	0,173	0,084	2,060	0,040
LUC indicator	0,084	0,051	1,660	0,097
GDD	-0,000	0,000	-2,620	0,009
GDD over 30ºC	-0,000	0,000	-2,260	0,024
Precipitation	0,002	0,000	7,560	-
Precipitation squared	-0,000	0,000	-5,570	-
Spatial correlation	0,595	0,029	20,580	0,000
Number of groups = 32				
Panel length = 35				

In the table above we show the quarterly pattern, where we incorporate the quarterly variables that end in December, March and June, previous to the start of the campaign. The difference in signs between temperature in different periods indicates that different combinations of quarterly temperature could give different results of the soybean yields even if the average is the same.

Appendix 5: Estimation results: SEM with spatial fixed effects

	SEM December	SEM January	SEM February	SEM March
Trend	-2,060	-2,192*	-3,084***	-2,748***
	(1,257)	(1,152)	(1,017)	(0,951)
Quadratic trend	0,001*	0,001*	0,001***	0,001***
	(0,000)	(0,000)	(0,000)	(0,000)
El Niño temperature anomaly (-3Q)	0,219***	0,220***	0,202***	0,178***
	(0,059)	(0,054)	(0,047)	(0,044)
El Niño temperature anomaly (-2Q)	-0,261***	-0,252***	-0,213***	-0,180***
	(0,080)	(0,073)	(0,065)	(0,060)
El Niño temperature anomaly (-1Q)	0,139***	0,113**	0,068	0,057
	(0,051)	(0,047)	(0,042)	(0,038)
Transgenic Soybean	0,347***	0,339***	0,372***	0,312***
	(0,130)	(0,118)	(0,102)	(0,095)
Ratio of soybean price/fertilizer price (-1)	0,257**	0,194*	0,196**	0,173**
	(0,112)	(0,104)	(0,090)	(0,084)
LUC indicator	0,077	0,076	0,079	0,084*
	(0,052)	(0,052)	(0,051)	(0,051)
Growing degree days	-0,0004*	-0,0003**	-0,0003***	-0,0002***
	(0,000)	(0,000)	(0,000)	(0,000)
GDD over 30ºC	-0,000058	-0,000007	-0,000139	-0,000260**
	(0,000)	(0,000)	(0,000)	(0,000)
Precipitation	0,001***	0,002***	0,002***	0,002***
	(0,000)	(0,000)	(0,000)	(0,000)
Precipitation squared	-0,000001***	-0,000001***	-0,000001***	-0,000001***
	(0,000)	(0,000)	(0,000)	(0,000)
Spatial correlation	0,694***	0,668***	0,624***	0,595***
	(0,023)	(0,025)	(0,028)	(0,029)
Number of groups = 32				
Panel length = 35				

Standard errors are reported in parentheses.

*, **, *** indicates significance at the 90%, 95%, and 99% level, respectively

Previous volumes in this series

1277 July 2025	Firm-level CO2 emissions and production networks: evidence from administrative data in Chile	Pablo Acevedo, Elías Albagli, Gonzalo García-Trujillo and María Antonia Yung
1276 July 2025	Economic activity, inflation, and monetary policy after extreme weather events: ENSO and its economic Impact in the Peruvian economy	John Aguirre, Alan Ledesma, Fernando Perez and Youel Rojas
1275 July 2025	Decoding climate-related risks in sovereign bond pricing: a global perspective	Sofia Anyfantaki, Marianna Blix Grimaldi, Carlos Madeira, Simona Malovana and Georgios Papadopoulos
1274 July 2025	Incorporating physical climate risks into banks' credit risk models	Vasily Pozdyshev, Alexey Lobanov and Kirill Ilinsky
1273 June 2025	Global portfolio investments and FX derivatives	Tsvetelina Nenova, Andreas Schrimpf and Hyun Song Shin
1272 June 2025	Financial conditions and the macroeconomy: a two-factor view	Marco Jacopo Lombardi, Cristina Manea and Andreas Schrimpf
1271 June 2025	Monetary policy and earnings inequality: inflation dependencies	Jaanika Meriküll and Matthias Rottner
1270 May 2025	Stablecoins and safe asset prices	Rashad Ahmed and Iñaki Aldasoro
1269 May 2025	Expecting job replacement by GenAI: Effects on workers' economic outlook and behavior	Yusuke Aoki, Joon Suk Park, Yuya Takada and Koji Takahashi
1268 May 2025	Towards verifiability of total value locked (TVL) in decentralized finance	Pietro Saggese, Michael Fröwis, Stefan Kitzler, Bernhard Haslhofer and Raphael Auer
1267 May 2025	Collateralized lending in private credit	Iñaki Aldasoro and Sebastian Doerr
1266 May 2025	Exchange Rate Effects on Firm Performance: A NICER Approach	Nuwat Nookhwun, Jettawat Pattararangrong and Phurichai Rungcharoenkitkul

All volumes are available on our website www.bis.org.