Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

ENSO-related hydroclimatic variability and soybean yield anomalies in the eastern Brazilian Cerrado: A 33-year regional assessment

Patrício Gomes Leite
Federal University of Maranhão
Jordânio Inácio Marques
Federal University of Maranhão
Washington da Silva Sousa
Federal University of Maranhão
Marcos Vinícius da Silva
mv.silva@ufma.br

Federal University of Maranhão

Research Article

Keywords: El Niño–Southern Oscillation, soybean yield, hydroclimate, climatic water deficit, heat extremes, climate variability

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11381186/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

ENSO-related hydroclimatic variability and soybean yield anomalies in the eastern Brazilian Cerrado: A 33-year regional assessment

Patrício Gomes Leite [1], Jordânio Inácio Marques [1], Washington da Silva Sousa [1], Marcos Vinícius da Silva [1,*]

[1] Chapadinha Science Center, Federal University of Maranhão, Chapadinha, MA, 65500-000, Brazil. E-mail: patricio.leite@ufma.br, jordanio.marques@ufma.br, washington.sousa@ufma.br, mv.silva@ufma.br

* Corresponding author: mv.silva@ufma.br

Abstract

El Niño–Southern Oscillation (ENSO) is an important source of interannual hydroclimatic variability, yet its relationships with growing-season conditions and soybean yield variability remain insufficiently resolved in the expanding agricultural frontiers of the eastern Brazilian Cerrado. We assessed associations among ENSO, hydroclimatic conditions, and detrended soybean yield anomalies across 15 municipalities in Maranhão, Brazil, from 1993 to 2025. December–February Oceanic Niño Index (ONI) was evaluated in relation to growing-season precipitation, climatic water deficit, maximum temperature, hot-day frequency, dry-spell duration, seasonal precipitation distribution, and detrended soybean yield anomalies using correlation analyses and municipality fixed-effects models. Robustness was assessed using harvest-year cluster and temporal moving-block bootstraps, exclusion of the influential 2016 harvest, and leave-one-year-out analyses. Higher ONI was associated with lower soybean yield anomalies (Spearman's $\rho = -0.544$, $p = 0.001$) and with a drier and hotter growing-season environment, characterized by lower precipitation, greater climatic water deficit, higher mean maximum temperature, and more frequent days with maximum temperature ≥35 °C. In the baseline fixed-effects model, a 1 °C increase in ONI was associated with a 6.23-percentage-point decrease in yield anomaly (95% harvest-year cluster-bootstrap CI: −10.32 to −1.83). The ONI coefficient remained negative across model specifications and sensitivity analyses, although exclusion of the influential 2016 harvest attenuated the baseline estimate to −3.55 percentage points $°C^{-1}$ (95% CI: −5.57 to −1.28). Strong covariation among the hydroclimatic indicators limited the statistical separation of their partial associations with soybean yield anomalies. These findings identify a consistent regional association among ENSO variability, coupled drier and hotter growing-season conditions, and soybean yield anomalies in Maranhão, highlighting the value of regional-scale analyses for understanding how large-scale climate variability is expressed in agriculturally relevant conditions within the Brazilian Cerrado.

Keywords: El Niño–Southern Oscillation; soybean yield; hydroclimate; climatic water deficit; heat extremes; climate variability

1. Introduction

Soybean (*Glycine max* (L.) Merr.) production has expanded extensively across the Brazilian Cerrado, including the MATOPIBA agricultural frontier encompassing parts of Maranhão, Tocantins, Piauí, and Bahia. Predominantly rainfed production in this region is strongly exposed to seasonal rainfall variability and high atmospheric demand. In the eastern Amazonia–Cerrado transition, recent decades have been characterized by increasing temperature and atmospheric dryness, more frequent dry days, declining precipitation and humidity, and changes in wet-season characteristics (Marengo et al., 2022). These conditions are particularly relevant for soybean production, which has expanded into areas where variability in water availability and temperature can substantially influence productivity (Silva et al., 2023).

El Niño–Southern Oscillation (ENSO) is a major source of interannual climate variability affecting South American hydroclimate, but its regional expression is spatially and seasonally heterogeneous. Agricultural responses therefore depend not simply on the occurrence of El Niño or La Niña, but on how ENSO-related variability is expressed through growing-season rainfall, water availability, temperature, and weather extremes. In southern Brazil, for example, soybean responses to ENSO vary with location and sowing date because crops experience different combinations and timing of meteorological conditions and extremes (Nóia Júnior et al., 2020). Water and thermal conditions may also covary: drought and heat can occur jointly and constrain soybean production, with geographically heterogeneous impacts across Brazilian production regions (Parisoto et al., 2026). Thus, precipitation, climatic water deficit, maximum temperature, hot-day frequency, dry-spell duration, and seasonal rainfall

63 distribution provide complementary dimensions of the hydroclimatic environment relevant to
64 soybean production.

65 Nevertheless, a consistent ENSO–yield association cannot be assumed across spatial
66 scales. Da Cunha Mello et al. (2026) reported generally weak ONI–yield relationships at state
67 and national scales, alongside substantial heterogeneity among municipalities. These findings
68 indicate that ONI alone has limited value as a general predictor of Brazilian soybean yield, but
69 they also motivate geographically focused analyses of regions where ENSO variability may
70 coincide with a coherent growing-season hydroclimatic signal.

71 This question remains insufficiently resolved for Maranhão, an expanding soybean-
72 producing region in the eastern Brazilian Cerrado. Long-term evidence is lacking on whether
73 interannual ONI variability is associated with a coherent regional hydroclimatic pattern and
74 with soybean yield anomalies after long-term productivity trends are removed. Addressing this
75 question also requires accounting for temporal dependence, influential extreme harvests, and
76 covariation among hydroclimatic indicators, which can complicate interpretation of their partial
77 associations with crop yield.

78 Accordingly, this study assessed ENSO-related variability in growing-season
79 hydroclimate and soybean yield anomalies across 15 municipalities in Maranhão, Brazil, during
80 1993–2025. We evaluated whether December–February (DJF) Oceanic Niño Index (ONI)
81 variability was associated with growing-season hydroclimatic conditions, including the
82 seasonal distribution of precipitation; whether detrended soybean yield anomalies varied with
83 ONI and ENSO phase; and whether the ONI–yield association persisted across municipality
84 fixed-effects specifications, hydroclimatic covariate adjustments, temporal-dependence
85 sensitivity analyses, and exclusion of influential harvest years. We hypothesized that higher
86 ONI values would be associated with a coupled drier and hotter growing-season environment
87 and with more negative soybean yield anomalies.

2. Materials and methods

2.1. Study area and municipality selection

The study was conducted in Maranhão State, northeastern Brazil, within an agricultural frontier predominantly associated with the Cerrado biome and the MATOPIBA region. The analysis covered 15 soybean-producing municipalities distributed across four production regions: southern Maranhão (10 municipalities), central Maranhão (2), eastern Maranhão (2), and western Maranhão (1). Together, these municipalities produced approximately 2.94 million tonnes of soybean in 2025, accounting for 67.0% of the state's production.

Municipalities were selected using two successive criteria. First, all municipalities contributing at least 1% of Maranhão's soybean production in 2025 were identified from the Brazilian Municipal Agricultural Production database (PAM/IBGE). This criterion yielded 22 municipalities, collectively representing 81.3% of statewide production. Second, municipalities were required to have at least 15 harvest years with available records of soybean harvested area and production between 1993 and 2025. Seven municipalities with shorter production records were excluded, resulting in the final set of 15 municipalities and an unbalanced panel of 439 municipality–harvest-year observations.

The study region is characterized predominantly by a tropical savanna climate with a dry winter (Aw, Köppen classification). Rainfall is strongly seasonal, with the wet season generally extending from November to April. Soybean sowing typically coincides with the onset of the rainy season, mainly in November and December, whereas harvesting generally occurs between March and May. Accordingly, November–April was adopted as the reference growing-season window for calculating agroclimatic indicators. This fixed window provides a consistent basis for interannual comparisons, although actual sowing dates and crop developmental stages may vary among municipalities and years.

The geographical distribution of the selected municipalities and their production regions is shown in Fig. 1. Regional soybean production in 2025 and the availability of municipality–harvest-year observations are summarized in Table 1.

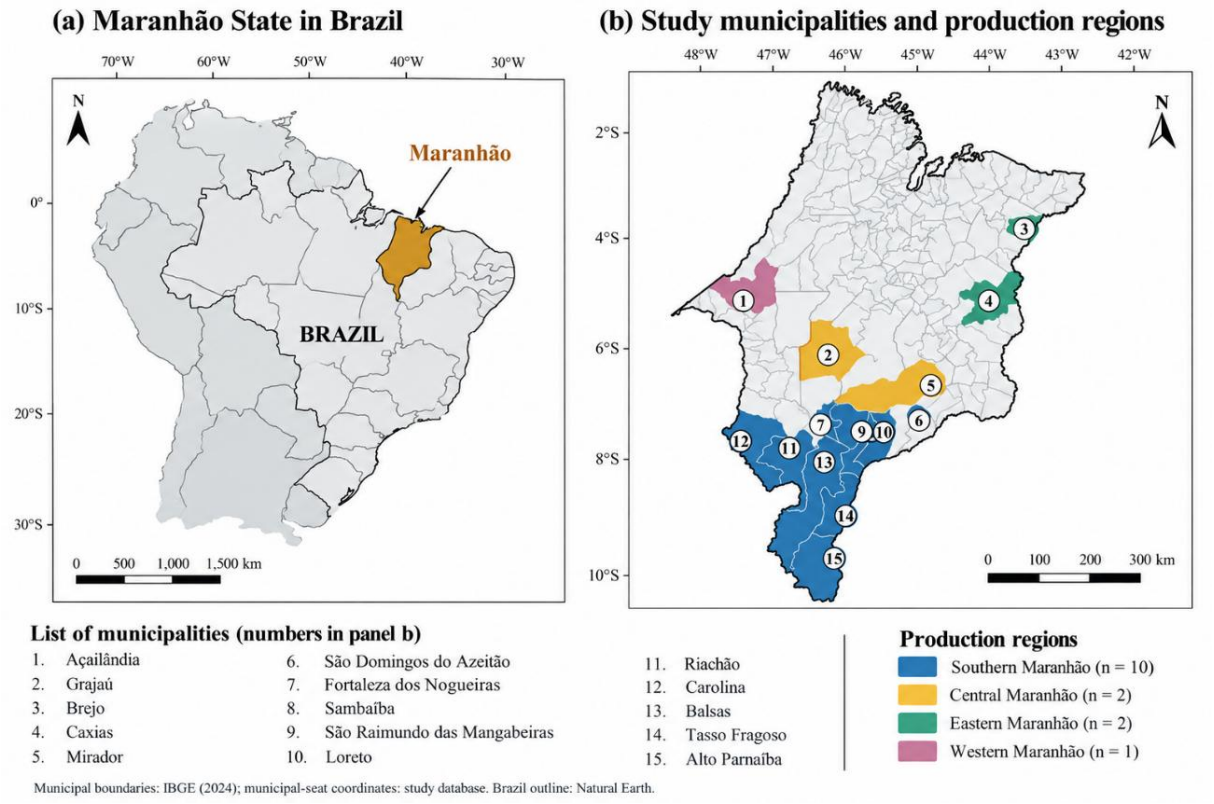


Fig. 1. Location of Maranhão State in Brazil (a) and geographical distribution of the 15 soybean-producing municipalities included in the study (b). Municipalities are identified by numbers and grouped into four production regions. Municipal boundaries were obtained from IBGE (2025).

Table 1. Soybean production and data availability across the four study regions in Maranhão, Brazil.

Production region	Municipalities (n)	Soybean production in 2025 (t)	Share of study-area production (%)	Municipality-year observations
Southern Maranhão	10	2,264,995	77.02	324
Central Maranhão	2	272,339	9.26	54
Eastern Maranhão	2	111,215	3.78	45
Western Maranhão	1	292,209	9.94	16
Total	15	2,940,758	100.00	439

2.2. Soybean production data and analytical period

Annual municipal soybean production data were obtained from the Municipal Agricultural Production survey (PAM), available through the Brazilian Institute of Geography and Statistics (IBGE) SIDRA database, Table 1612 (IBGE, 2026). The dataset covered 33 harvest years, from 1993 to 2025, and included harvested area (ha), production (tonnes), and reported yield (kg ha^{-1}).

Soybean yield was independently calculated from harvested area and production:

$$Y_{i,t} = \frac{1000 P_{i,t}}{A_{i,t}} \quad (1)$$

where $Y_{i,t}$ is soybean yield (kg ha^{-1}) in municipality i and harvest year t, $P_{i,t}$ is production (tonnes), and $A_{i,t}$ is harvested area (ha).

Recalculated yields were compared with the corresponding values reported by IBGE to assess internal consistency. Municipality–year records without valid harvested-area or production data were treated as missing and were neither interpolated nor replaced. The resulting dataset contained 439 observations across the 15 municipalities and 33 harvest years.

Harvest years were defined according to the year in which the crop was harvested. Thus, the 2016 harvest corresponds to the growing season beginning in November 2015 and extending into 2016. This convention was applied consistently when matching agricultural production records with meteorological variables and the El Niño–Southern Oscillation (ENSO) index.

2.3. Meteorological data

Daily meteorological data were obtained from the National Aeronautics and Space Administration (NASA) Langley Research Center Prediction Of Worldwide Energy Resources (POWER) project for the period from 1 January 1992 to 31 December 2025 (NASA POWER, 2026). Data were retrieved for the geographical coordinates of each municipality's administrative center. The series began in 1992 to include the complete November 1992–April 1993 growing season corresponding to the first harvest year.

The meteorological variables comprised corrected daily precipitation (PRECTOTCORR), mean air temperature (T2M), maximum air temperature (T2M_MAX), minimum air temperature (T2M_MIN), relative humidity (RH2M), incoming shortwave solar radiation (ALLSKY_SFC_SW_DWN), and wind speed at 2 m (WS2M). These variables were used to characterize rainfall distribution, thermal conditions, atmospheric evaporative demand, and climatic water availability during the soybean growing season.

Table 2. Daily meteorological variables.

Variable	NASA POWER parameter	Unit
Precipitation	PRECTOTCORR	mm day^{-1}
Mean air temperature	T2M	°C
Maximum air temperature	T2M_MAX	°C
Minimum air temperature	T2M_MIN	°C
Relative humidity	RH2M	%
Incoming shortwave radiation	ALLSKY_SFC_SW_DWN	MJ m^{-2} day^{-1}
Wind speed at 2 m	WS2M	m s^{-1}

The meteorological database comprised 186,285 daily municipality-level records, corresponding to 12,419 days for each of the 15 municipalities. Data screening included checks for missing values, invalid codes, and physical consistency among minimum, mean, and maximum temperatures.

Because NASA POWER provides gridded meteorological estimates, geographically close municipalities may share the same meteorological grid cell. This occurred for Sambaíba and São Raimundo das Mangabeiras, which had identical meteorological series but distinct soybean production records. Both municipalities were retained in the agricultural panel; however, regional climate analyses were conducted at the harvest-year level to avoid interpreting duplicated spatial information as independent evidence.

As an additional data-quality assessment, annual precipitation totals derived from NASA POWER were compared with available records from nearby weather stations for 1993–2023. Stations were assigned to municipalities according to geographic proximity and data availability. Pearson correlations indicated generally moderate to strong agreement, although

performance varied spatially. Correlations were approximately 0.72–0.77 for several municipality–station comparisons, while weaker agreement was observed in some locations, including Alto Parnaíba and Carolina. These comparisons supported the use of NASA POWER as a spatially consistent source of long-term meteorological information while indicating that gridded estimates may not fully capture local rainfall variability.

2.4. ENSO characterization

The El Niño–Southern Oscillation (ENSO) was characterized using the Oceanic Niño Index (ONI), obtained from the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC; NOAA/CPC, 2026). The ONI represents the three-month running mean of sea surface temperature anomalies in the Niño 3.4 region (5°N–5°S, 170°W–120°W), based on the Extended Reconstructed Sea Surface Temperature version 5 (ERSSTv5; Huang et al., 2017) and centered 30-year base periods updated every five years.

Each soybean harvest year was assigned the ONI value for the December–January–February (DJF) season preceding harvest. This trimester was selected because it overlaps the main rainy season in Maranhão and captures ENSO conditions during an important portion of the regional soybean growing season. For example, the 2016 harvest was associated with the DJF 2015–2016 ONI value.

The ONI assigned to each harvest year was expressed as:

$$ONI_t = \frac{1}{3}\left(SSTA_{Dec,t-1} + SSTA_{Jan,t} + SSTA_{Feb,t}\right) \quad (2)$$

where ONI_t is the December–February *ONI* associated with harvest year *t*, and *SSTA* denotes the monthly sea surface temperature anomaly in the Niño 3.4 region. Equation (2) expresses the temporal alignment adopted in this study; the index values themselves were obtained directly from *NOAA* rather than recalculated from sea surface temperature observations.

Following the historical NOAA/CPC classification, harvest years were classified as El Niño when the ONI reached or exceeded +0.5 °C for at least five consecutive overlapping three-month periods, and as La Niña when it reached or fell below −0.5 °C for the same minimum duration. Years not meeting either criterion were classified as ENSO-neutral. The resulting classification comprised 10 El Niño, 10 neutral, and 13 La Niña harvest years.

Table 3. Classification of soybean harvest years according to ENSO phase during 1993–2025.

ENSO phase	Harvest years	n
El Niño	1995, 1998, 2003, 2005, 2007, 2010, 2015, 2016, 2019, 2024	10
Neutral	1993, 1994, 1997, 2002, 2004, 2013, 2014, 2017, 2020, 2025	10
La Niña	1996, 1999, 2000, 2001, 2006, 2008, 2009, 2011, 2012, 2018, 2021, 2022, 2023	13

Source: NOAA Climate Prediction Center, ONI (ERSSTv5), according to the classification adopted in the study.

The DJF ONI was also retained as a continuous variable in the correlation and regression analyses, allowing the magnitude and direction of ENSO-related sea surface temperature anomalies to be examined without reducing the information to three categorical phases. Because each harvest year was assigned a single ONI value shared by all municipalities, ENSO-related statistical inference accounted for the common temporal exposure rather than treating municipality-level observations within the same year as independent ENSO events.

2.5. Agroclimatic indicators and climatic water balance

Daily meteorological records were used to derive 13 agroclimatic indicators for each municipality and harvest year. The indicators characterized precipitation amount and seasonal

distribution, dry-spell duration, thermal conditions, atmospheric evaporative demand, and climatic water availability. All growing-season indicators were calculated for the period from 1 November of the year preceding harvest to 30 April of the harvest year.

Precipitation was evaluated both over the complete growing-season window and over three consecutive two-month periods: November–December, January–February, and March–April. This subdivision was adopted to investigate whether ENSO-related precipitation variability differed within the agricultural season. These periods represent fixed calendar windows rather than observed phenological stages, as municipality-specific sowing and developmental records were not available. The agroclimatic indicators and their definitions are summarized in Table 4.

Table 4. Agroclimatic indicators calculated for each municipality and soybean harvest year.

Group	Indicator	Definition	Unit
Precipitation	Growing-season rainfall	Total rainfall from November to April	mm
Precipitation	Nov–Dec rainfall	Rainfall accumulated during November and December	mm
Precipitation	Jan–Feb rainfall	Rainfall accumulated during January and February	mm
Precipitation	Mar–Apr rainfall	Rainfall accumulated during March and April	mm
Precipitation	Rainy days	Number of days with rainfall ≥ 1 mm	days
Dry spells	Longest dry spell	Longest sequence of days with rainfall below 1 mm during December–March	days
Dry spells	Prolonged dry spells	Number of dry spells lasting at least 10 consecutive days during December–March	count

Temperature	Mean maximum temperature	Average daily maximum temperature during November–April	°C
Temperature	Mean air temperature	Average daily mean temperature during November–April	°C
Temperature	Hot days	Number of days with maximum temperature ≥ 35 °C	days
Evaporative demand	Reference evapotranspiration	Cumulative daily ETo during November–April	mm
Water balance	Rainfall minus ETo	Difference between cumulative rainfall and ETo	mm
Water balance	Climatic water deficit	Cumulative daily water deficit estimated using a sequential soil-water balance	mm

Source: Authors' calculations based on daily NASA POWER meteorological data.

2.6. Reference evapotranspiration

Daily reference evapotranspiration (ETo, mm day^{-1}) was estimated using the FAO-56 Penman–Monteith equation (Allen et al., 1998), as expressed in Eq. (3):

$$ET_o = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T+273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (3)$$

where R_n is net radiation (MJ m^{-2} day^{-1}), G is soil heat flux (MJ m^{-2} day^{-1}), T is mean daily air temperature (°C), u_2 is wind speed at 2 m (m s^{-1}), e_s and e_a are saturation and actual vapor pressures (kPa), respectively, Δ is the slope of the saturation vapor pressure curve (kPa °C^{-1}), and γ is the psychrometric constant (kPa °C^{-1}).

Soil heat flux was assumed to be negligible at the daily timescale ($G = 0$). Mean air temperature for the evapotranspiration calculation was obtained from the arithmetic mean of

daily maximum and minimum temperatures. Saturation vapor pressure was calculated from the corresponding saturation vapor pressures at maximum and minimum temperatures, whereas actual vapor pressure was estimated using daily relative humidity.

Net radiation was calculated as the difference between net shortwave radiation and net longwave radiation, according to Eq. (4):

$$R_n = R_{ns} - R_{nl} \quad (4)$$

Net shortwave radiation was estimated using an albedo of 0.23, as shown in Eq. (5):

$$R_{ns} = (1 - 0.23)R_s = 0.77R_s \quad (5)$$

where R_s is incoming solar radiation (MJ m^{-2} day^{-1}). Net longwave radiation was estimated using the FAO-56 formulation, accounting for air temperature, atmospheric vapor pressure, and the ratio between measured and clear-sky solar radiation. Extraterrestrial radiation and clear-sky radiation were calculated as functions of latitude and day of the year. Atmospheric pressure was estimated assuming an elevation of 300 m, resulting in a psychrometric constant of approximately 0.065 kPa $°C^{-1}$.

Daily ETo values were summed over the November–April growing-season window to quantify cumulative atmospheric evaporative demand.

2.7. Sequential climatic water balance

Climatic water availability was characterized using a sequential daily soil-water balance based on the Thornthwaite–Mather approach (Thornthwaite and Mather, 1955). A soil available

water capacity (AWC) of 100 mm was assumed for all municipalities, providing a standardized reference for comparisons across locations and harvest years.

The water balance was initialized on 1 January 1992 with soil-water storage at field capacity and was calculated continuously through 31 December 2025. This continuous implementation allowed antecedent soil-water conditions to influence subsequent growing seasons.

For days when precipitation equaled or exceeded reference evapotranspiration ($P_t \geq ET_{o,t}$), soil-water storage was updated according to Eq. (6):

$$S_t = min\left[AWC, S_{t-1} + P_t - ET_{o,t}\right] \quad (6)$$

For days when precipitation was lower than reference evapotranspiration ($P_t < ET_{o,t}$), soil-water storage was calculated using Eq. (7):

$$S_t = S_{t-1} exp\left(\frac{P_t - ET_{o,t}}{AWC}\right) \quad (7)$$

where S_t and S_{t-1} are soil-water storage on the current and previous days (mm), respectively, P_t is daily precipitation (mm), and AWC is the assumed available water capacity (100 mm).

Actual evapotranspiration within this reference water-balance framework was estimated using Eq. (8):

$$ET_{a,t} = \begin{cases} ET_{o,t}\,, & P_t \geq ET_{o,t} \\ P_t + (S_{t-1} - S_t), & P_t < ET_{o,t} \end{cases} \quad (8)$$

The daily climatic water deficit was then calculated as the difference between reference and actual evapotranspiration, according to Eq. (9):

$$D_t = ET_{o,t} - ET_{a,t} \quad (9)$$

Finally, cumulative growing-season water deficit was obtained by summing the daily deficits from November to April, as expressed in Eq. (10):

$$D_{GS,t} = \sum_{d=Nov\ 1}^{Apr\ 30} D_d \quad (10)$$

The resulting indicator represents climatic water limitation under a standardized reference evapotranspiration and soil-water storage framework. It does not account explicitly for cultivar-specific crop coefficients, rooting depth, phenological development, or spatial differences in soil-water holding capacity. Accordingly, it was interpreted as a comparative agroclimatic indicator rather than a direct measurement of soybean water stress.

2.8. Soybean yield detrending and anomaly calculation

Long-term changes in soybean yield may reflect technological advances, improvements in crop management, changes in cultivar selection, and the expansion of cultivation into areas with different production potential. To reduce the influence of these long-term changes on the analysis of interannual variability, soybean yield was detrended separately for each municipality using an ordinary least-squares linear regression of observed yield against harvest year, as expressed in Eq. (11):

$$Y_{i,t} = \alpha_i + \beta_i t + \varepsilon_{i,t} \quad (11)$$

where $Y_{i,t}$ is the observed soybean yield (kg ha^{-1}) in municipality i and harvest year t, α is the municipality-specific intercept, β_i represents the annual linear trend in yield (kg ha^{-1} year^{-1}), and $\varepsilon_{i,t}$ is the residual variation not explained by the linear trend.

Each regression was fitted using only the harvest years with valid yield records for the corresponding municipality. Missing observations were not interpolated. The fitted values from Eq. (11) were used to estimate the yield expected under the municipality-specific long-term trend, as shown in Eq. (12):

$$\hat{Y}_{i,t} = \hat{\alpha}_i + \hat{\beta}_i t \quad (12)$$

Relative yield anomalies were subsequently calculated as the percentage deviation of observed yield from the corresponding trend estimate, according to Eq. (13):

$$YA_{i,t} = 100\left(\frac{Y_{i,t}-\hat{Y}_{i,t}}{\hat{Y}_{i,t}}\right) \quad (13)$$

where $YA_{i,t}$ is the relative soybean yield anomaly (%). Negative values indicate yields below the municipality-specific trend, whereas positive values indicate yields above the trend.

Expressing anomalies as percentages allowed yield variability to be compared among municipalities with different average productivity levels. The detrending procedure was applied independently to all 15 municipalities, resulting in an unbalanced panel of 439 municipality–harvest-year observations for the subsequent statistical analyses.

To account for differences in baseline climatic conditions among municipalities, climatic water deficit and mean maximum temperature were standardized separately within

each municipality using the observations available in the integrated agricultural–climatic panel. Standardization was performed according to Eq. (14):

$$z_{i,t} = \frac{X_{i,t} - \bar{X}_i}{s_i} \quad (14)$$

where $X_{i,t}$ represents the climatic indicator for municipality i in harvest year t, $\bar{X}_i$ is the mean of that indicator across the municipality's available observations in the analytical panel, and s_i is the corresponding sample standard deviation. Standardization was performed before excluding 2016 for sensitivity analyses, ensuring that the full-sample and restricted-sample models used the same climatic reference distributions. ONI was retained in °C, whereas soybean yield anomalies were expressed as percentages.

This transformation expresses each observation relative to the historical climatic conditions of its municipality. Consequently, regression coefficients for standardized climatic predictors represent changes in soybean yield anomaly, in percentage points, associated with a one-standard-deviation increase in the corresponding indicator.

The ONI was retained in its original units (°C), as it represents a common large-scale climatic index shared by all municipalities within each harvest year. Yield anomalies were also retained in percentage units to preserve their direct agronomic interpretation.

2.9. Statistical analysis

Statistical analyses were designed to investigate three complementary aspects of ENSO-related hydroclimatic variability: (i) associations between ENSO and growing-season hydroclimatic conditions, including the seasonal distribution of rainfall; (ii) associations between ENSO and detrended soybean yield anomalies; and (iii) the sensitivity of these associations to correlated climatic predictors and influential harvest years.

The analysis combined harvest-year-level correlations with municipality fixed-effects regression models. Because all municipalities shared the same ONI value within a given harvest year, uncertainty estimation accounted for the clustering of municipality-level observations by harvest year, while additional moving-block bootstrap analyses were used to assess sensitivity to short-term temporal dependence across consecutive harvest years.

2.10. Regional aggregation and ENSO-related hydroclimatic variability

For each harvest year, regional hydroclimatic indicators were calculated as the arithmetic mean of the corresponding values across the 15 municipalities. This produced 33 annual observations for each indicator over the 1993–2025 period.

Associations between DJF ONI and the regional hydroclimatic indicators were evaluated using Spearman's rank correlation coefficient ρ. The analysis focused on growing-season precipitation, the longest dry spell, cumulative climatic water deficit, mean maximum temperature, and the number of days with maximum temperature ≥ 35 °C.

To investigate the seasonal distribution of ENSO-related rainfall variability, separate correlations were calculated between ONI and precipitation accumulated during November–December, January–February, and March–April. Because the three periods are consecutive and non-overlapping, their accumulated precipitation equals total growing-season precipitation.

For each family of simultaneously tested indicators, the resulting p-values were adjusted using the Benjamini–Hochberg false discovery rate procedure (Benjamini and Hochberg, 1995). Associations were also recalculated after excluding the 2016 harvest to assess their sensitivity to this exceptionally dry and hot growing season.

Temporal uncertainty was further assessed using a non-circular moving-block bootstrap (Künsch, 1989) with a block length of three consecutive harvest years and 2,000 resamples. Blocks were restricted to consecutive calendar years; therefore, no block was allowed to cross

394 the gap created by excluding 2016. Percentile-based 95% confidence intervals were obtained
395 from the bootstrap distributions.
396
397 *2.11. ENSO phases and regional soybean yield anomalies*
398 Descriptive comparisons among El Niño, ENSO-neutral, and La Niña harvest years
399 were based on regional annual means for the hydroclimatic indicators and on municipality–
400 harvest-year observations for soybean yield anomalies.
401 For each harvest year, the regional yield anomaly was calculated as the arithmetic mean
402 of the available municipal yield anomalies, according to Eq. (15):
403
404 $$\overline{YA}_t = \frac{1}{n_t}\sum_{i=1}^{n_t} YA_{i,t} \quad (15)$$
405 where $\overline{YA}_t$ is the regional mean soybean yield anomaly in harvest year t, $YA_{i,t}$ is the detrended
406 yield anomaly for municipality i, and n_t is the number of municipalities with valid yield
407 observations in that year.
408
409 Regional annual anomalies were associated with the corresponding DJF ONI values
410 using Spearman's rank correlation coefficient. This analysis was repeated after excluding 2016
411 to evaluate the sensitivity of the relationship to that harvest year.
412 For descriptive comparisons among ENSO phases, regional annual hydroclimatic
413 observations and municipality–harvest-year soybean yield anomalies were grouped according
414 to the classification presented in Table 3. Means and standard deviations were calculated
415 separately for El Niño, neutral, and La Niña conditions at their respective analytical levels.
416 Because the municipal yield panel was unbalanced, the number of municipalities
417 contributing to each annual regional yield anomaly varied over time. Accordingly, these annual

means were interpreted as summaries of the available study municipalities rather than estimates derived from a constant spatial sample.

2.12. Municipality fixed-effects regression models

The association between ENSO and soybean yield anomalies was investigated using linear regression models with municipality fixed effects. This specification accounts for time-invariant differences among municipalities while estimating the association between interannual ENSO variability and detrended yield anomalies.

The baseline model included DJF ONI as the principal explanatory variable, as expressed in Eq. (16):

$$YA_{i,t} = \alpha_i + \beta_1 ONI_t + \varepsilon_{i,t} \quad (16)$$

where α_i is the municipality-specific intercept, β_1 is the coefficient associated with ONI, and $\varepsilon_{i,t}$ is the residual error.

Additional specifications included standardized climatic water deficit and mean maximum temperature, separately and jointly. The complete model was expressed as Eq. (17):

$$YA_{i,t} = \alpha_i + \beta_1 ONI_t + \beta_2 zD_{i,t} + \beta_3 zT_{max,i,t} + \varepsilon_{i,t} \quad (17)$$

where $zD_{i,t}$ and $zT_{max,i,t}$ represent municipality-standardized climatic water deficit and mean maximum temperature, respectively.

The four regression specifications are summarized in Table 5.

Table 5. Municipality fixed-effects regression models used to assess associations between ENSO, hydroclimatic conditions, and soybean yield anomalies.

Model	Predictors	Purpose
F1	ONI	Baseline ENSO–yield association
F2	ONI + standardized water deficit	Association after accounting for climatic water deficit
F3	ONI + standardized maximum temperature	Association after accounting for thermal conditions
F4	ONI + standardized water deficit + standardized maximum temperature	Joint specification including ENSO and both climatic indicators

All models include municipality fixed intercepts. The response variable is detrended soybean yield anomaly (%). ONI is expressed in °C, whereas the climatic predictors are standardized within each municipality.

Model fit was summarized using the adjusted coefficient of determination (R_{adj}^2). Comparisons among model specifications were used to evaluate differences in descriptive fit and coefficient stability rather than to establish causal pathways.

2.13. Multicollinearity and uncertainty estimation

Multicollinearity among ONI, standardized climatic water deficit, and standardized maximum temperature was assessed using variance inflation factors (VIFs), accounting for the municipality fixed effects included in the regression models. The VIF for each predictor was calculated according to Eq. (18):

$$VIF_j = \frac{1}{1-R_j^2} \quad (18)$$

458 where R_j^2 is the coefficient of determination obtained by regressing predictor *j* on the remaining
459 predictors and the municipality fixed effects.
460
461 Because ONI varies among harvest years but not among municipalities within a given
462 year, conventional observation-level standard errors could overstate the precision of its
463 estimated association with yield. Therefore, coefficient uncertainty was assessed using a
464 harvest-year cluster bootstrap.
465 In each bootstrap iteration, harvest years were sampled with replacement, and all
466 available municipal observations belonging to each selected year were retained together. The
467 regression models were then refitted to each resampled dataset. A total of 1,000 bootstrap
468 iterations were used to obtain percentile-based 95% confidence intervals for the regression
469 coefficients. The procedure was applied to all four model specifications, both with the complete
470 dataset and after excluding the 2016 harvest.
471
472 *2.14. Temporal dependence and sensitivity analyses*
473 To evaluate whether the regression results were sensitive to temporal dependence among
474 harvest years, an additional moving-block bootstrap was implemented using blocks of three
475 consecutive years and 2,000 resamples. Unlike the harvest-year cluster bootstrap, which
476 resamples individual years, the moving-block procedure retains short sequences of consecutive
477 harvest years. Only valid consecutive-year blocks were admitted, and the resampling procedure
478 excluded datasets with rank-deficient regression design matrices.
479 The moving-block analysis focused on the baseline model (F1) and the complete
480 climatic specification (F4). Both models were evaluated using the full dataset and the dataset
481 excluding 2016. The sensitivity of the estimated coefficients to individual harvest years was
482 additionally assessed using leave-one-year-out (LOYO) analysis. Each harvest year was

483 excluded in turn, and models F1 and F4 were refitted to the remaining observations. For each
484 excluded year *t*, the change in an estimated coefficient was calculated as Eq. (19):
485
486 $$\Delta\hat{\beta}_{j,t} = \hat{\beta}_{j,(-t)} - \hat{\beta}_{j,full} \quad (19)$$
487 where $\hat{\beta}_{j,(-t)}$ is the coefficient estimated after excluding harvest year *t*, and $\hat{\beta}_{j,full}$ is the
488 corresponding estimate obtained using all harvest years.
489
490 Residuals and Cook's distances (Cook, 1977) were also examined to identify
491 observations and harvest years with comparatively large contributions to model influence. For
492 each harvest year, individual Cook's distances were summed as a descriptive diagnostic; these
493 sums were not interpreted as equivalent to a Cook's distance calculated from the joint deletion
494 of an entire year.
495 Particular attention was given to 2016 because preliminary diagnostics identified it as
496 an influential harvest year. Its regional hydroclimatic conditions were compared with the 1993–
497 2025 distributions, and the corresponding municipal yield anomalies were examined to
498 characterize the spatial extent of the observed production losses.
499
500 *2.15. Statistical interpretation*
501 All statistical analyses were interpreted as observational associations. Municipality
502 fixed effects account for time-invariant differences among municipalities but do not eliminate
503 potentially confounding factors that vary among harvest years.
504 Furthermore, climatic water deficit and maximum temperature are physically related
505 and statistically correlated. Consequently, changes in the ONI coefficient after adding these
506 predictors were interpreted as sensitivity to model specification, rather than evidence of causal
507 mediation.

The 2016 exclusion analyses were used to quantify the influence of an exceptional harvest year, not to justify removing it from the principal dataset. The complete 1993–2025 dataset was retained as the primary analytical sample.

2.16. *Computational tools and reproducibility*

All data processing, agroclimatic calculations, and statistical analyses were performed using MATLAB R2026a (MathWorks, Natick, MA, USA). The analytical workflow comprised meteorological data quality assessment, calculation of agroclimatic indicators, integration of agricultural and climatic datasets, municipality-specific soybean yield detrending, and statistical analyses of ENSO-related variability.

Daily meteorological observations were organized by municipality and date and used to calculate reference evapotranspiration, sequential climatic water balance, and growing-season agroclimatic indicators. These indicators were aggregated by municipality and harvest year and integrated with agricultural production records and DJF ONI values using municipality and harvest year as matching variables. Data integrity checks included duplicate detection, chronological consistency, valid agricultural production values, and verification of the expected numbers of observations. The sources, temporal coverage, and dimensions of the resulting datasets are summarized in Table 6.

Table 6. Data sources, temporal coverage, and dimensions of the analytical datasets.

Dataset	Source	Temporal coverage	Number of records
Daily meteorological data	NASA POWER	1992–2025	186,285
ENSO index	NOAA/CPC	1993–2025	33 harvest years

Agroclimatic indicators	Authors' calculations	1993–2025	495
Soybean production and yield	IBGE/PAM	1993–2025	439
Integrated analytical panel	Authors' calculations	1993–2025	439

Note: The meteorological dataset contains 12,419 daily records for each of the 15 municipalities. The agroclimatic panel contains 15 municipalities over 33 harvest years. The agricultural and integrated panels are unbalanced because valid soybean production records are unavailable for some municipality–year combinations.

The statistical workflow was implemented using MATLAB scripts for Spearman rank correlations, Benjamini–Hochberg false discovery rate adjustment, municipality fixed-effects regression, multicollinearity diagnostics, and sensitivity analyses. The four fixed-effects specifications were fitted using ordinary least squares with municipality-specific intercepts. Climatic water deficit and maximum temperature were standardized within municipalities using the observations available in the integrated analytical panel, whereas ONI and relative yield anomalies were retained in their original units.

Uncertainty was assessed using complementary resampling procedures. For the fixed-effects regression models, a harvest-year cluster bootstrap was performed with 1,000 valid resamples, retaining all available municipal observations associated with each sampled harvest year. For the temporal robustness analyses, a non-circular moving-block bootstrap was performed using blocks of three chronologically consecutive harvest years and 2,000 valid resamples. Only admissible consecutive-year blocks were sampled, and regression resamples with rank-deficient design matrices or unsuccessful model fits were rejected. When 2016 was excluded, blocks were not permitted to cross the resulting temporal gap.

Temporal moving-block bootstrap procedures with 2,000 resamples were also applied to the regional hydroclimatic and seasonal precipitation correlations. All stochastic analyses were initialized using the Mersenne Twister random number generator with a fixed seed of 2026 (rng(2026,'twister')).

Model sensitivity was further assessed by excluding the 2016 harvest and by conducting leave-one-year-out analyses. Regression residuals and Cook's distances were examined to characterize the influence of individual observations and harvest years. Intermediate datasets and analytical outputs were exported in tabular formats, including the integrated analytical panel, regional annual summaries, regression coefficients, bootstrap confidence intervals, collinearity diagnostics, and harvest-year influence measures.

3. Results

3.1. ENSO-related variability in growing-season hydroclimate

ENSO variability was associated with substantial changes in growing-season hydroclimatic conditions across the 15 municipalities during 1993–2025. At the regional scale, higher DJF Oceanic Niño Index (ONI) values were associated with lower November–April precipitation (*Spearman's* ρ = -0.518, p = 0.0020, FDR-adjusted q = 0.0051) and greater cumulative climatic water deficit (ρ =0.488, p=0.0039, q = 0.0066). Positive associations were also observed between ONI and mean maximum temperature (ρ = 0.544, p = 0.0011, q = 0.0051), the number of days with maximum temperature at or above 35 °C (ρ = 0.448, p = 0.0089, q = 0.0112), and the longest dry spell (ρ = 0.358, p = 0.0405, q = 0.0405). All five associations remained significant after adjustment for multiple testing in the complete 33-year series (Table 7).

Table 7. Growing-season hydroclimatic conditions by ENSO phase and their associations with DJF ONI in Maranhão, Brazil, 1993–2025.

Indicator	El Niño (n = 10)	Neutral (n = 10)	La Niña (n = 13)	Spearman's ρ	p	FDR q
Precipitation (mm)	861.2 ± 155.8	994.1 ± 183.8	1,108.8 ± 194.3	−0.518	0.0020	0.0051
Longest dry spell (days)	8.09 ± 2.74	6.42 ± 1.50	6.04 ± 1.40	0.358	0.0405	0.0405
Climatic water deficit (mm)	207.3 ± 87.1	145.5 ± 57.6	113.2 ± 63.0	0.488	0.0039	0.0066
Mean maximum temperature (°C)	31.67 ± 1.04	30.93 ± 1.06	30.38 ± 0.75	0.544	0.0011	0.0051
Days with Tmax ≥ 35 °C	28.04 ± 16.33	20.07 ± 14.70	13.23 ± 9.65	0.448	0.0089	0.0112

Correlations were calculated using one regional mean per harvest year (n = 33). FDR-adjusted q-values were calculated using the Benjamini–Hochberg procedure across the five hydroclimatic indicators. ENSO phases were assigned according to sustained NOAA/CPC episodes.

Descriptive comparisons among ENSO phases showed consistent differences in the regional hydroclimatic indicators. El Niño harvest years had a mean growing-season precipitation of 861.2 mm, compared with 994.1 mm during neutral years and 1,108.8 mm during La Niña years. Mean climatic water deficit was 207.3, 145.5, and 113.2 mm, respectively. El Niño years also exhibited higher mean maximum temperatures, more days with maximum temperature at or above 35 °C, and longer maximum dry spells (Table 7). These phase comparisons are descriptive; the correlation analyses use ONI as a continuous variable.

The associations were not solely attributable to the extreme 2016 harvest. After excluding 2016, ONI remained negatively correlated with growing-season precipitation (ρ = -0.471, q =0.0163) and positively correlated with climatic water deficit (ρ = 0.439, q = 0.0200), mean maximum temperature (ρ = 0.500, q =0.0163), and the number of days with maximum temperature at or above 35 °C (ρ = 0.395, q = 0.0318). The association with the longest dry

spell weakened (ρ = 0.298, p = 0.0979, q = 0.0979) and no longer met the 5% significance threshold.

Temporal moving-block bootstrap analyses provided additional evidence of the persistence of the principal associations. In the complete series, the 95% bootstrap confidence intervals excluded zero for precipitation (−0.774 to −0.274), climatic water deficit (0.163 to 0.786), mean maximum temperature (0.287 to 0.797), days with maximum temperature at or above 35 °C (0.038 to 0.763), and the longest dry spell (0.093 to 0.636). After excluding 2016, the confidence intervals continued to exclude zero for precipitation (−0.734 to −0.220), climatic water deficit (0.120 to 0.750), mean maximum temperature (0.270 to 0.767), and the longest dry spell (0.051 to 0.586), whereas the interval for hot days included zero (−0.015 to 0.717). Thus, the bootstrap analysis supported persistent associations for precipitation, climatic water deficit, and maximum temperature across both datasets, while the evidence for dry-spell duration and hot-day frequency was less consistent across inferential approaches.

The relationships between DJF ONI and the three principal hydroclimatic indicators are illustrated in Figure 2. Higher ONI values were generally associated with lower growing-season precipitation, greater cumulative climatic water deficit, and higher mean maximum temperature. The 2016 harvest is highlighted because it represented an extreme hydroclimatic year and was examined separately in the sensitivity analyses.

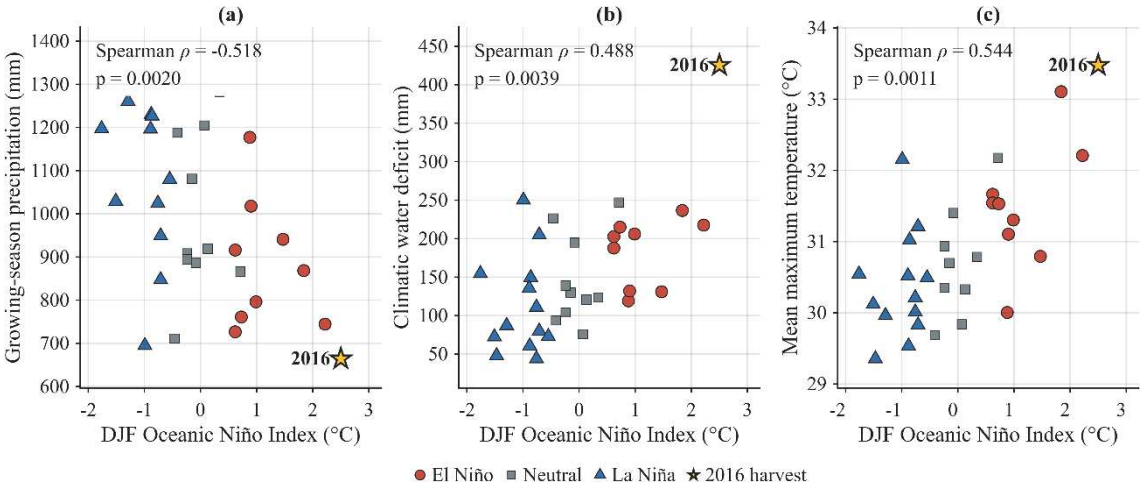


Figure 2. Relationships between the December–February (DJF) Oceanic Niño Index (ONI) and regional growing-season (a) precipitation, (b) cumulative climatic water deficit, and (c) mean maximum temperature across 15 municipalities in Maranhão, Brazil, during 1993–2025. Each point represents the regional mean for one harvest year (n = 33). Colors and symbols distinguish ENSO phases, and the 2016 harvest is highlighted by a star. Spearman's correlation coefficients and associated p-values are displayed in each panel.

3.2. Seasonal distribution of precipitation

The association between ENSO and growing-season precipitation varied across the three two-month periods considered in this study. DJF ONI was negatively correlated with regional precipitation during November–December (ρ =-0.419, p = 0.0153) and March–April (ρ =-0.416, p = 0.0160). Both associations remained significant after Benjamini–Hochberg adjustment (q = 0.0240). In contrast, the correlation for January–February was weaker and not statistically significant (ρ = -0.212, p = 0.2371). These results indicate that the association between DJF ONI and precipitation was more evident at the beginning and end of the November–April growing-season window than during its middle months (Table 8).

Table 8. Seasonal precipitation by ENSO phase and Spearman correlations with DJF ONI across 15 municipalities in Maranhão, Brazil, 1993–2025.

Period	l Niño (n = 10)	Neutral (n = 10)	La Niña (n = 13)	Spearman's ρ	p	FDR q
November–December	212.9 ± 68.3	252.9 ± 66.3	318.0 ± 96.5	−0.419	0.0153	0.0240
January–February	339.2 ± 47.7	407.2 ± 127.0	384.7 ± 82.5	−0.212	0.2371	0.2371
March–April	309.1 ± 74.4	334.0 ± 97.7	406.1 ± 115.6	−0.416	0.0160	0.0240

Note: Precipitation values are regional annual means ± standard deviations (mm). Correlations were calculated using one regional mean per harvest year (n = 33). FDR-adjusted q-values were calculated across the three seasonal precipitation tests. Phase comparisons are descriptive.

Descriptive comparisons among ENSO phases showed a similar seasonal pattern. Mean precipitation during November–December and March–April was lowest in El Niño years and highest in La Niña years, whereas January–February showed a different ordering, with the highest mean precipitation during neutral years (Table 8). These phase differences are descriptive and were not subjected to inferential comparisons.

The sensitivity analysis excluding 2016 yielded negative correlations for November–December (ρ = −0.362, p = 0.0415) and March–April (ρ = −0.367, p = 0.0391). However, neither association remained significant after adjustment for multiple testing (q = 0.0622 for both periods), while the January–February association remained weak (ρ = −0.222, p = 0.2227). Temporal moving-block bootstrap confidence intervals provided complementary evidence of uncertainty. In the complete series, the 95% intervals excluded zero for November–December (−0.688 to −0.037) and March–April (−0.715 to −0.035). After excluding 2016, the interval for November–December included zero (−0.646 to 0.037), whereas that for March–April remained entirely negative (−0.690 to −0.017).

Taken together, these results indicate that ONI-related precipitation variability was not distributed uniformly throughout the growing-season window. Negative associations were observed during November–December and March–April, whereas little evidence of an association was found for January–February. The November–December association was sensitive to the exclusion of 2016, while the March–April association remained negative in the temporal moving-block bootstrap after excluding that harvest. Nevertheless, neither association met the FDR-adjusted significance threshold in the restricted series, indicating that the seasonal patterns should be interpreted cautiously.

3.3. ENSO-related soybean yield anomalies

Soybean yield anomalies were negatively associated with DJF ONI over the 1993–2025 period. Based on one regional mean per harvest year (\(n = 33\)), the Spearman correlation between ONI and detrended soybean yield anomalies was $\rho = -0.544$ ($p = 0.0011$), indicating that higher ONI values tended to coincide with more negative regional yield anomalies. Because yield anomalies were calculated separately for each municipality relative to its municipality-specific linear trend before regional aggregation, this association reflects interannual departures from local yield trajectories rather than differences in absolute yield levels among municipalities.

Descriptive summaries of municipality–harvest-year observations showed a corresponding pattern across ENSO phases. Mean yield anomalies were −6.15 ± 18.52% under El Niño conditions, −0.24 ± 10.54% under neutral conditions, and +4.75 ± 11.35% under La Niña conditions. These statistics describe the 439 municipality–harvest-year observations and are presented as descriptive summaries rather than inferential comparisons among ENSO phases.

The 2016 harvest was the most prominent negative yield event in the study period. The regional mean yield anomaly was approximately −42.7%, and 14 of the 15 municipalities exhibited negative anomalies. This harvest also coincided with particularly adverse regional hydroclimatic conditions: growing-season precipitation averaged 664.3 mm, compared with 999.0 mm across the complete study period, while climatic water deficit reached 425.5 mm compared with a long-term mean of 151.5 mm. Mean maximum temperature was 33.48 °C, the number of days with maximum temperature ≥35 °C reached 63.9, and the longest dry spell reached 11.6 days, compared with corresponding long-term means of 30.94 °C, 19.8 days, and 6.8 days, respectively. Thus, 2016 combined the largest negative regional yield anomaly with markedly drier and hotter growing-season conditions than the long-term averages.

Despite the magnitude of the 2016 event, the regional ONI–yield association was not dependent solely on this harvest. After excluding 2016, the Spearman correlation remained negative and statistically significant ($\rho = -0.499$, $p = 0.0036$, \(n = 32\)), compared with $\rho = -0.544$ in the complete series. Thus, exclusion of 2016 attenuated the correlation without altering its direction or statistical significance, indicating that the negative regional association between ONI and soybean yield anomalies extended beyond this single harvest.

The relationship between DJF ONI and regional soybean yield anomalies is illustrated in Figure 3. Panel (a) shows the negative ONI–yield association across harvest years, whereas panel (b) shows the interannual variation in regional yield anomalies. The 2016 harvest is highlighted because of its exceptionally large negative yield anomaly and its relevance to the sensitivity analyses.

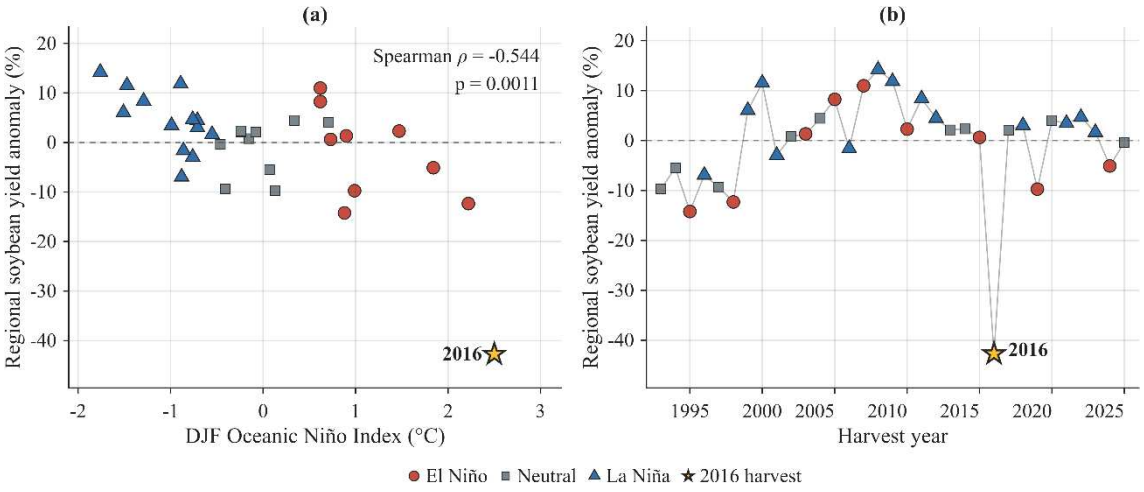


Figure 3. ENSO variability and regional soybean yield anomalies across 15 municipalities in Maranhão, Brazil, during 1993–2025. **(a)** Relationship between DJF ONI and regional mean detrended soybean yield anomalies (\(n = 33\)), with Spearman's ρ and p-value. **(b)** Interannual variation in regional mean yield anomalies. Colors and symbols indicate ENSO phases, and the 2016 harvest is highlighted by a star in both panels.

3.4. Fixed-effects models of soybean yield anomalies

The negative association between ONI and soybean yield anomalies was also evident in the municipality fixed-effects models (Table 9). In the baseline model (F1), a 1 °C increase in DJF ONI was associated with a 6.23-percentage-point decrease in soybean yield anomaly (β = -6.23\) percentage points $°C^{-1}$). The 95% harvest-year cluster-bootstrap confidence interval ranged from −10.32 to −1.83 percentage points $°C^{-1}$, excluding zero. The model had an adjusted R^2 of 0.182.

Table 9. Municipality fixed-effects models relating soybean yield anomalies to DJF ONI and growing-season hydroclimatic indicators across 15 municipalities in Maranhão, Brazil, 1993–2025.

Model	Predictor	Coefficient	95% cluster-bootstrap CI

F1	ONI	−6.23	−10.32 to −1.83
F2	ONI	−4.79	−7.51 to −1.97
	Climatic water deficit	−2.83	−6.50 to 2.67
F3	ONI	−5.85	−9.68 to −2.11
	Mean maximum temperature	−0.70	−4.23 to 3.68
F4	ONI	−5.50	−8.81 to −2.09
	Climatic water deficit	−7.31	−14.04 to 1.99
	Mean maximum temperature	+5.54	−1.28 to 13.30

Note: All models included municipality fixed effects. Soybean yield anomalies are expressed as percentages. ONI was retained in °C; therefore, its coefficients represent the change in yield anomaly, in percentage points, associated with a 1 °C increase in DJF ONI. Climatic water deficit and mean maximum temperature were standardized within municipality; their coefficients represent the change in yield anomaly, in percentage points, associated with a one-standard-deviation increase in the corresponding predictor. Confidence intervals are percentile 95% intervals obtained from 1,000 harvest-year cluster-bootstrap resamples, with all available municipality observations from each sampled harvest year retained together.

Adding growing-season climatic water deficit (F2) reduced the magnitude of the ONI coefficient from −6.23 to −4.79 percentage points $°C^{-1}$, while the coefficient for standardized climatic water deficit was −2.83 percentage points per standard deviation. However, its 95% cluster-bootstrap confidence interval included zero (−6.50 to 2.67). When mean maximum temperature was included instead (F3), the ONI coefficient remained negative (−5.85 percentage points $°C^{-1}$), whereas the coefficient for standardized maximum temperature was −0.70 percentage points per standard deviation, with a confidence interval that also included zero (−4.23 to 3.68).

The model containing both hydroclimatic indicators (F4) had the highest adjusted R^2 (0.235). The ONI coefficient remained negative (−5.50 percentage points $°C^{-1}$; 95% cluster-

bootstrap CI: −8.81 to −2.09), whereas the coefficients for standardized climatic water deficit (−7.31 percentage points per standard deviation; 95% CI: −14.04 to 1.99) and mean maximum temperature (+5.54 percentage points per standard deviation; 95% CI: −1.28 to 13.30) were individually uncertain. The change in sign and magnitude of the climatic coefficients when both predictors were included is consistent with their strong covariation and warrants caution when interpreting their partial regression coefficients.

Overall, ONI retained a negative coefficient across all four model specifications, and its harvest-year cluster-bootstrap confidence intervals consistently excluded zero. In contrast, the coefficients for climatic water deficit and mean maximum temperature were less stable across specifications, and their confidence intervals included zero. Thus, the fixed-effects models consistently supported a negative association between ONI and detrended soybean yield anomalies in the full panel, but provided limited evidence for statistically separable partial associations between yield anomalies and the hydroclimatic predictors included in the multivariable models.

3.5. Temporal robustness and influence of the 2016 harvest

Sensitivity analyses showed that the negative association between DJF ONI and soybean yield anomalies persisted after exclusion of the 2016 harvest, although its estimated magnitude was attenuated (Table 10). In the ONI-only model (F1), the coefficient changed from −6.23 percentage points $°C^{-1}$ in the complete panel to −3.55 percentage points $°C^{-1}$ after excluding 2016, while the harvest-year cluster-bootstrap 95% confidence interval remained entirely negative (−5.57 to −1.28). A similar pattern was observed in F4, in which the ONI coefficient changed from −5.50 to −3.97 percentage points $°C^{-1}$, with a cluster-bootstrap confidence interval of −6.82 to −1.55 after excluding 2016.

Table 10. Sensitivity of municipality fixed-effects estimates to exclusion of the 2016 harvest and alternative bootstrap procedures accounting for harvest-year clustering and temporal dependence.

Model	Predictor	Full panel coefficient	Cluster-bootstrap 95% CI	Excluding 2016 coefficient	Cluster-bootstrap 95% CI	3-year block-bootstrap 95% CI, full	3-year block-bootstrap 95% CI, excluding 2016
F1	ONI	−6.23	−10.32 to −1.83	−3.55	−5.57 to −1.28	−10.49 to −2.16	−5.02 to −1.59
F4	ONI	−5.50	−8.81 to −2.09	−3.97	−6.82 to −1.55	−8.81 to −2.18	−6.66 to −1.49
	Climatic water deficit	−7.31	−14.04 to 1.99	−0.82	−5.37 to 2.99	−15.61 to 1.42	−7.25 to 2.54
	Mean maximum temperature	+5.54	−1.28 to 13.30	+1.53	−2.13 to 7.65	−1.90 to 15.33	−3.30 to 9.98

Note: F1 includes DJF ONI and municipality fixed effects; F4 additionally includes within-municipality standardized growing-season climatic water deficit and mean maximum temperature. ONI coefficients are expressed as percentage points per °C, whereas coefficients for the standardized hydroclimatic predictors are expressed as percentage points per standard deviation. Cluster-bootstrap intervals are percentile 95% confidence intervals obtained from 1,000 harvest-year resamples, with all available municipality observations from each sampled harvest year retained together. Temporal intervals are percentile 95% confidence intervals obtained from 2,000 valid non-circular moving-block bootstrap resamples using blocks of three chronologically consecutive harvest years. In analyses excluding 2016, blocks were not allowed to cross the resulting temporal gap.

765 Accounting for short-term temporal dependence using the 3-year non-circular moving-
766 block bootstrap produced a similar pattern. For F1, the 95% confidence interval for the ONI
767 coefficient was −10.49 to −2.16 percentage points °C^{-1} in the complete panel and −5.02 to −1.59
768 after excluding 2016. For F4, the corresponding intervals were −8.81 to −2.18 and −6.66 to
769 −1.49 percentage points °C^{-1}, respectively. Thus, the ONI confidence intervals remained
770 entirely negative under both harvest-year cluster resampling and temporally constrained
771 moving-block resampling, with and without the 2016 harvest.
772 In contrast, the partial coefficients of the hydroclimatic predictors in F4 were sensitive
773 to the exclusion of 2016. The coefficient for standardized climatic water deficit changed from
774 −7.31 to −0.82 percentage points per standard deviation, whereas that for standardized mean
775 maximum temperature changed from +5.54 to +1.53 percentage points per standard deviation.
776 Confidence intervals for both predictors included zero under both bootstrap procedures, with
777 and without the 2016 harvest. These changes, together with the strong covariation among the
778 hydroclimatic indicators, indicate limited statistical separation of their partial associations with
779 yield anomalies and should not be interpreted as evidence that water or thermal conditions were
780 agronomically unimportant.
781 The influence of 2016 was also evident in the model diagnostics. Excluding this harvest
782 shifted the ONI coefficient by +2.68 percentage points °C^{-1} in F1 and by +1.53 percentage
783 points °C^{-1} in F4, corresponding to less negative estimates in both models. In F4, the mean
784 residual across municipalities in 2016 was −19.35 percentage points, with a root mean square
785 residual of 24.74 percentage points. As descriptive influence diagnostics, the sums of the
786 individual municipality-level Cook's distances for the 2016 observations were 0.290 in F1 and
787 0.236 in F4. These values summarize the influence of the individual 2016 observations and
788 should not be interpreted as Cook's distances obtained from joint deletion of the entire harvest
789 year.

Leave-one-year-out analysis further assessed the sensitivity of the estimated ONI association to individual harvest years. The ONI coefficient remained negative across all sequential year omissions, indicating that its direction was not dependent on any single harvest year. Excluding 2016 produced one of the largest attenuations in coefficient magnitude. Together with the harvest-year cluster-bootstrap and moving-block bootstrap results, the LOYO analysis indicates that 2016 contributed substantially to the magnitude of the estimated ONI–yield association but did not determine its negative direction.

Overall, the sensitivity analyses showed that the negative ONI coefficient persisted after exclusion of the 2016 harvest, across leave-one-year-out analyses, and under bootstrap resampling that accounted for harvest-year clustering and short-term temporal dependence. However, its magnitude was attenuated when 2016 was excluded, indicating that this extreme harvest contributed appreciably to the strength of the estimated association. In contrast, the partial coefficients for climatic water deficit and mean maximum temperature were more sensitive to model specification and to the inclusion of 2016, with confidence intervals that consistently included zero.

3.6. Covariation among hydroclimatic indicators

The growing-season hydroclimatic indicators exhibited substantial interannual covariation at the regional scale. Climatic water deficit was strongly and positively correlated with mean maximum temperature (ρ = 0.934) and with the number of days with maximum temperature ≥35 °C (ρ = 0.949), while mean maximum temperature was also strongly correlated with the frequency of days ≥35 °C (ρ = 0.895). Thus, harvest years characterized by greater climatic water deficit generally also exhibited higher mean maximum temperatures and more frequent days with maximum temperature ≥35 °C.

This covariation was also reflected in the fixed-effects regression diagnostics. In the complete analytical panel, variance inflation factors were 1.53 for ONI, 4.63 for standardized climatic water deficit, and 4.88 for standardized mean maximum temperature. After excluding the 2016 harvest, the corresponding values were 1.32, 4.16, and 4.55, respectively. The higher VIFs for the two hydroclimatic predictors are consistent with the substantial shared variation between climatic water deficit and mean maximum temperature, whereas ONI exhibited considerably less collinearity with the other predictors. Excluding 2016 did not materially alter this dependence structure: the regional correlation between climatic water deficit and mean maximum temperature remained high (ρ = 0.928), as did that between climatic water deficit and the number of days with maximum temperature ≥35 °C (ρ = 0.944).

The combined pattern of coefficient stability and hydroclimatic covariation is summarized in Figure 4. Across the fixed-effects model specifications, ONI retained a negative coefficient, with harvest-year cluster-bootstrap confidence intervals entirely below zero in both the complete panel and the analyses excluding 2016 (Figure 4a). Exclusion of 2016 shifted the ONI estimates toward zero, indicating that this harvest contributed appreciably to the magnitude of the estimated association without determining its negative direction. In contrast, the partial coefficients for standardized climatic water deficit and mean maximum temperature were less stable across model specifications, and their confidence intervals included zero. The strong correlations among the hydroclimatic indicators (Figure 4b) provide important statistical context for this instability, as water deficit, maximum temperature, and the frequency of days ≥35 °C varied closely together across harvest years.

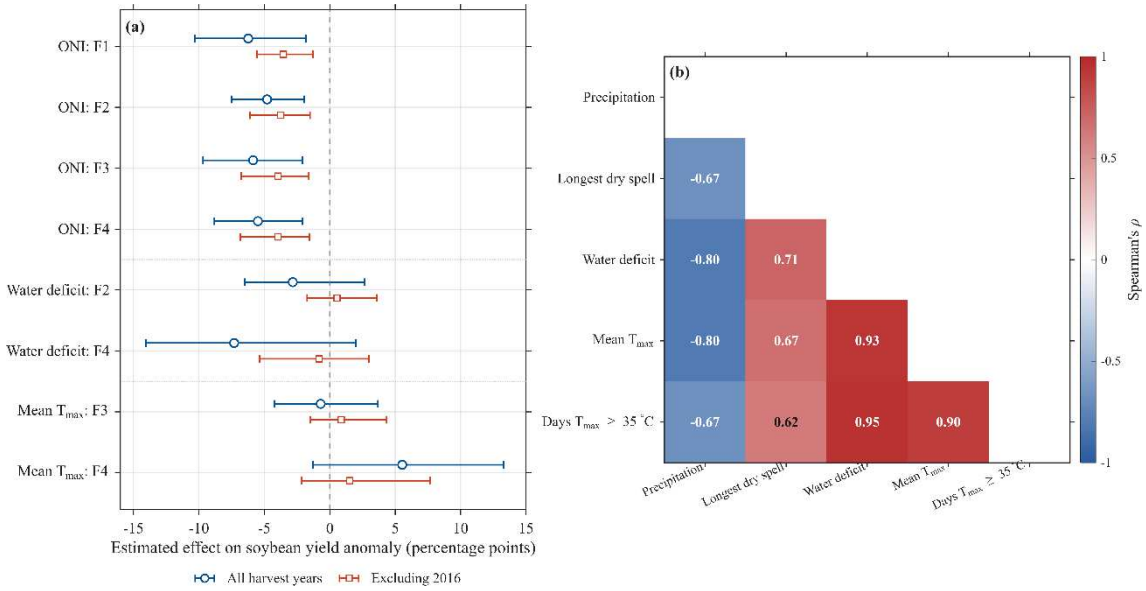


Figure 4. Robustness of fixed-effects estimates and covariation among growing-season hydroclimatic indicators. **(a)** Fixed-effects coefficients and 95% harvest-year cluster-bootstrap confidence intervals for the complete panel and after excluding 2016. **(b)** Spearman correlations among regional hydroclimatic indicators across harvest years (n = 33). The dashed line in (a) denotes a coefficient of zero.

The contrast between the comparatively stable ONI coefficient and the less stable partial coefficients for climatic water deficit and mean maximum temperature indicates that the multivariable models more consistently characterize the overall ONI–yield association than statistically separate the hydroclimatic components associated with it. Accordingly, the changes in sign and magnitude of the hydroclimatic coefficients when climatic water deficit and mean maximum temperature were included simultaneously should not be interpreted as evidence of opposing physiological responses. Rather, they reflect the limited ability of the available observational data to distinguish the partial associations of strongly correlated components of growing-season hydroclimatic variability.

4. Discussion

The present study identified a consistent regional association between ENSO variability, growing-season hydroclimate, and soybean yield anomalies in the eastern Brazilian Cerrado. Across the 33-year record, higher DJF ONI values were associated with lower regional soybean yield anomalies and with a combination of lower precipitation, greater climatic water deficit, higher maximum temperatures, and more frequent days with maximum temperature ≥35 °C. El Niño harvest years were also characterized descriptively by negative mean yield anomalies, whereas La Niña years showed positive mean anomalies. The negative ONI–yield association persisted across fixed-effects model specifications and multiple sensitivity analyses, although its magnitude was influenced by the extreme 2016 harvest. Taken together, these findings indicate that interannual ENSO variability provides a large-scale climatic context for coupled hydroclimatic conditions associated with soybean yield variability in this part of the Brazilian Cerrado.

This regional signal is noteworthy because ENSO–soybean yield relationships are spatially heterogeneous across Brazil. Nóia Júnior et al. (2020) showed that soybean responses to ENSO in southern Brazil varied with location and sowing date, reflecting differences in crop exposure to weather conditions and extremes. More recently, da Cunha Mello et al. (2026) reported generally weak ONI–yield relationships at state and national scales, alongside substantial heterogeneity among municipalities. The consistent negative association observed in Maranhão should therefore be interpreted as a regional feature rather than generalized to Brazilian soybean production. More broadly, this contrast illustrates how geographically focused analyses can identify regional climate–yield associations that may be attenuated when spatially heterogeneous responses are aggregated over larger areas.

The hydroclimatic results provide a plausible environmental context for the observed ONI–yield association. Higher ONI values were associated with lower growing-season precipitation and greater climatic water deficit, as well as higher mean maximum temperatures

879 and more frequent days with maximum temperature ≥35 °C. Descriptively, El Niño harvest
880 years were generally drier and warmer than La Niña harvest years. The co-occurrence of water
881 limitation and elevated temperature is agronomically relevant because drought and heat can
882 jointly constrain soybean production, including in expanding production regions of Brazil
883 (Parisoto et al., 2026). The seasonal analysis further showed that the negative association
884 between ONI and precipitation was concentrated in November–December and March–April,
885 with little evidence of an association during January–February. This temporal heterogeneity
886 may be important because soybean exposure to adverse weather conditions depends on sowing
887 date and the developmental period in which those conditions occur (Nóia Júnior et al., 2020).
888 However, because annual municipality-specific sowing dates and phenological information
889 were unavailable, the seasonal precipitation patterns identified here cannot be assigned to
890 specific soybean developmental stages.

891 The strong covariation among hydroclimatic indicators is critical for interpreting the
892 multivariable models. Harvest years with greater climatic water deficit also tended to have
893 higher maximum temperatures and more frequent days ≥35 °C, while precipitation was
894 negatively correlated with these indicators. This coupled hydroclimatic structure is consistent
895 with the instability of the partial coefficients for climatic water deficit and mean maximum
896 temperature when both predictors were included in the same model. Their higher VIFs relative
897 to ONI, together with bootstrap confidence intervals that included zero, further indicate limited
898 statistical separation of their partial associations with yield anomalies. Consequently, changes
899 in the sign or magnitude of these coefficients should not be interpreted as evidence of opposing
900 physiological responses. Likewise, persistence of the ONI coefficient after adjustment for the
901 hydroclimatic predictors does not demonstrate an ENSO association independent of
902 hydroclimate, nor does the attenuation of that coefficient constitute evidence of mediation.
903 Finally, the climatic water deficit calculated here should be interpreted as a climatic water-

balance indicator rather than a direct measure of crop water stress, which also depends on soil water storage and properties, rooting characteristics, cultivar, phenology, and management.

The 2016 harvest illustrates both the importance of compound hydroclimatic extremes and the need for caution when attributing individual events to a single large-scale climate driver. It exhibited the largest negative regional soybean yield anomaly in the study period and coincided with exceptionally low precipitation, high climatic water deficit, elevated maximum temperature, frequent days ≥35 °C, and a prolonged dry spell. The broader 2015–2016 drought was also exceptional across tropical South America. Erfanian et al. (2017) showed that anomalously warm sea-surface temperatures in both the tropical Pacific and Atlantic contributed substantially to the drought, but could not fully account for the severity of the observed rainfall deficits across parts of Amazonia and the Nordeste region. The strong 2015–2016 El Niño therefore provides an important large-scale climatic context for the conditions observed in Maranhão, but should not be interpreted as the sole cause of the regional yield loss. Consistent with this caution, excluding 2016 attenuated the magnitude of the ONI coefficient without changing its negative direction, and the corresponding cluster- and moving-block-bootstrap confidence intervals remained entirely below zero. Thus, 2016 contributed substantially to the magnitude of the estimated ONI–yield association but did not determine its negative direction.

These findings complement recent evidence that ENSO alone has limited value as a universal predictor of Brazilian soybean yield (da Cunha Mello et al., 2026). The substantial variability among municipalities and harvest years within ENSO phases also argues against a deterministic interpretation of El Niño and La Niña categories. In Maranhão, ONI may instead provide large-scale climatic context for periods of elevated regional hydroclimatic risk, although the present retrospective analysis did not evaluate prospective forecasting skill. For operational applications, the potential value of ENSO information should therefore be tested in

combination with regional precipitation and temperature forecasts, atmospheric and soil water availability, sowing dates, and crop-development information. Such integrated approaches could evaluate whether the lead time associated with large-scale climate information provides useful predictive skill once the local processes linking climate variability to crop production are explicitly represented.

The main limitations of this study arise from the spatial and temporal resolution of the observational data. Municipality-level yield statistics integrate heterogeneity in soils, cultivars, sowing dates, and management practices, while detrending cannot remove all time-varying non-climatic influences on productivity. Regional averaging may also mask localized responses, and the 33-year record encompasses a finite number of ENSO episodes and harvest years, limiting the resolution with which temporal heterogeneity can be characterized. In addition, the absence of annual municipality-specific sowing dates, phenological information, soil water conditions, and management data restricts interpretation of the specific processes linking hydroclimatic variability to yield anomalies. These constraints preclude causal attribution and limit the separation of individual hydroclimatic mechanisms. Nevertheless, the persistence of the negative ONI–yield association across complementary model specifications and sensitivity analyses supports its interpretation as a regional statistical relationship rather than one determined solely by the 2016 extreme harvest. More broadly, the results show that higher ONI values tended to coincide with a coupled drier and hotter growing-season environment and more negative soybean yield anomalies in the studied municipalities, highlighting the importance of regional-scale analyses for understanding how large-scale climate variability is expressed in agriculturally relevant conditions within the Brazilian Cerrado.

5. Conclusions

This 33-year regional assessment showed a consistent association between ENSO variability, growing-season hydroclimate, and soybean yield anomalies across the studied municipalities in Maranhão, within the eastern Brazilian Cerrado. Higher DJF ONI values were associated with lower soybean yield anomalies and with a drier and hotter growing-season environment, characterized by lower precipitation, greater climatic water deficit, higher mean maximum temperatures, and more frequent days with maximum temperature ≥35 °C. The negative ONI–yield association persisted across fixed-effects model specifications, harvest-year cluster and temporal moving-block bootstrap procedures, exclusion of the influential 2016 harvest, and leave-one-year-out analyses, although its estimated magnitude was attenuated when 2016 was excluded.

Strong covariation among precipitation, climatic water deficit, and thermal indicators showed that the hydroclimatic conditions associated with ENSO variability formed a coupled regional signal whose individual components could not be reliably separated in the multivariable models. These findings support the use of ENSO information as a large-scale climatic context for assessing interannual hydroclimatic and soybean yield variability in Maranhão, rather than as a deterministic predictor of crop performance. Future work should evaluate whether combining ENSO information with regional precipitation and temperature forecasts, soil water conditions, sowing dates, and crop-development information provides useful prospective skill for soybean climate-risk assessment.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

6. References

978 Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration: Guidelines for
979 computing crop water requirements. FAO Irrigation and Drainage Paper No. 56. Food and
980 Agriculture Organization of the United Nations, Rome.
981 Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: A practical and
982 powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57, 289–300.
983 https://doi.org/10.1111/j.2517-6161.1995.tb02031.x
984 Cook, R.D., 1977. Detection of influential observation in linear regression. Technometrics 19,
985 15–18. https://doi.org/10.1080/00401706.1977.10489493
986 da Cunha Mello, F.D., Kumar, P., Nascimento, E.G.S., 2026. Weak and heterogeneous ENSO
987 teleconnections to Brazilian soybean yields: a municipal-to-national assessment. Theor.
988 Appl. Climatol. 157, 322. https://doi.org/10.1007/s00704-026-06266-z
989 Erfanian, A., Wang, G., Fomenko, L., 2017. Unprecedented drought over tropical South
990 America in 2016: significantly under-predicted by tropical SST. Sci. Rep. 7, 5811.
991 https://doi.org/10.1038/s41598-017-05373-2
992 Huang, B., Thorne, P.W., Banzon, V.F., Boyer, T., Chepurin, G., Lawrimore, J.H., Menne, M.J.,
993 Smith, T.M., Vose, R.S., Zhang, H.-M., 2017. Extended Reconstructed Sea Surface
994 Temperature, Version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. J. Clim.
995 30, 8179–8205. https://doi.org/10.1175/JCLI-D-16-0836.1
996 IBGE, 2025. Malha Municipal Digital 2024. Instituto Brasileiro de Geografia e Estatística,
997 Diretoria de Geociências, Rio de Janeiro. Accessed 29 September 2026.
998 IBGE, 2026. Produção Agrícola Municipal (PAM). Instituto Brasileiro de Geografia e
999 Estatística, Rio de Janeiro. Sistema IBGE de Recuperação Automática (SIDRA). Accessed
1000 29 September 2026.
1001 Künsch, H.R., 1989. The jackknife and the bootstrap for general stationary observations. Ann.
1002 Stat. 17, 1217–1241. https://doi.org/10.1214/aos/1176347265

1003 Marengo, J.A., Jimenez, J.C., Espinoza, J.-C., Cunha, A.P., Aragão, L.E.O., 2022. Increased
1004 climate pressure on the agricultural frontier in the Eastern Amazonia–Cerrado transition
1005 zone. Sci. Rep. 12, 457. https://doi.org/10.1038/s41598-021-04241-4
1006 NASA POWER, 2026. Prediction Of Worldwide Energy Resources (POWER). NASA Langley
1007 Research Center, Hampton, VA, USA. Accessed 29 September 2026.
1008 NOAA/CPC, 2026. Oceanic Niño Index (ONI): Historical El Niño/La Niña episodes (1950–
1009 present), ERSSTv5. National Oceanic and Atmospheric Administration, Climate Prediction
1010 Center. Accessed 29 September 2026.
1011 Nóia Júnior, R.S., Fraisse, C.W., Karrei, M.A.Z., Cerbaro, V.A., Perondi, D., 2020. Effects of
1012 the El Niño Southern Oscillation phenomenon and sowing dates on soybean yield and on the
1013 occurrence of extreme weather events in southern Brazil. Agric. For. Meteorol. 290, 108038.
1014 https://doi.org/10.1016/j.agrformet.2020.108038
1015 Parisoto, G.J., Muñoz-Arriola, F., Pilau, F.G., 2026. Interannual and intraseasonal effects of
1016 drought and heatwaves on expanding soybean production regions in Brazil. Atmosphere 17,
1017 367. https://doi.org/10.3390/atmos17040367
1018 Silva, D.S., Arima, E.Y., dos Reis, T.N.P., Rattis, L., 2023. Temperature effect on Brazilian
1019 soybean yields, and farmers' responses. Int. J. Agric. Sustain. 21, 2173370.
1020 https://doi.org/10.1080/14735903.2023.2173370
1021 Thornthwaite, C.W., Mather, J.R., 1955. The water balance. Publications in Climatology 8 (1),
1022 1–104.