A Unified Computational Pipeline for Characterizing Crop Sensitivity in Northern Italy

Esther Helga
au668255@uni.au.dk

Anders Lydig Kristensen
au606349@uni.au.dk

Bioinformatics Research Center (BiRC)
Aarhus University

Vocational Project in collaboration with

Teknologisk Institut
(Danish Technological Institute)

Abstract

As the Mediterranean Basin warms at rates exceeding global averages, the agricultural stability of Northern Italy's Po Valley faces increasing thermal and hydrological risks. This project evaluates the capacity of a unified computational pipeline to characterize the divergent climate vulnerabilities of five major crops: Winter Wheat, Spring Wheat, Maize, Rice, and Soybean. By harmonizing high-resolution ERA5-Land reanalysis data with the Global Dataset of Historical Yields (1982–2016), we implemented a two-stage modeling workflow utilizing Elastic Net regularization and Gamma Generalized Linear Models (GLMs). The analysis reveals distinct biological "signals" of vulnerability. We identify a "Wheat Paradox", where early spring warmth acts as a catalyst for Winter Wheat (Pseudo-$R^2 = 0.91$) but a stressor for Spring Wheat (Pseudo-$R^2 = 0.57$), highlighting a significant predictability gap driven by management heterogeneity. Furthermore, summer crops exhibit "Hydrological Dominance", where yield is governed by soil water retention—a proxy for irrigation resilience—rather than immediate precipitation. Additionally, Rice yields show a counter-intuitive negative response to June solar radiation, indicative of floodwater overheating. These findings validate the pipeline as a diagnostic tool for mapping the boundary between climate-driven biological responses and human-managed agricultural outcomes.

Keywords: Climate Change, Po Valley, Agricultural Resilience, Crop Yield Modeling, Phenology, ERA5-Land, Elastic Net Regularization.

January 2026

Contents

List of Abbreviations

AIC Akaike Information Criterion

ERA5 ECMWF Reanalysis v5 (Climate Dataset)

GDHY Global Dataset of Historical Yields

GLM Generalized Linear Model

GxExM Genotype × Environment × Management

NDVI Normalized Difference Vegetation Index

PE / PEV Potential Evaporation

SSR / SR Surface Net Solar Radiation

SW / SWVL1 Volumetric Soil Water Layer 1

TP / P Total Precipitation

T2M / T Temperature measured at 2 meters above surface

1 Introduction

1.1 Global Food Security and the Mediterranean Hotspot

Climate change has emerged as the primary threat to global food security, with increasing yield volatility observed across major breadbasket regions (MedECC, 2020). Within this global context, the Mediterranean Basin is recognized as a prominent "climate change hot-spot", where warming trends significantly exceed global averages (Giorgi, 2006). This intensification of the hydrological cycle and thermal regimes places a disproportionate burden on Mediterranean agricultural systems, which must maintain productivity under increasingly frequent weather extremes (Drobinski et al., 2025).

1.2 The Po Valley: A Multi-Crop Breadbasket

Northern Italy's Po Valley serves as a critical agricultural hub, accounting for a significant portion of the European Union's cereal and rice production (Hristov et al., 2020). The region is characterized by high agricultural diversity where both C3 (Wheat, Rice, Soybean) and C4 (Maize) systems coexist. These systems possess fundamentally different photosynthetic pathways, with C4 species generally exhibiting higher water-use efficiency and thermal tolerance than their C3 counterparts (Sage, 2004). However, the region's reliance on a highly managed hydrological system—fed by Alpine snowmelt and the Po River—makes it uniquely vulnerable to shifting climate patterns (Musolino et al., 2018).

For this study, five dominant crops were selected to evaluate divergent regional vulnerabilities:

- **Winter Wheat:** Autumn-sown (November–June); highly dependent on winter thermal accumulation (Mereu et al., 2021; Poggi et al., 2022).
- **Spring Wheat:** Spring-sown (March–July); vulnerable to early-season desiccation (Monteleone et al., 2023).
- **Maize, Rice, and Soybean:** Summer-sown (May–September); these crops occupy the peak heat window and are historically buffered by intensive irrigation (Monteleone and Borzí, 2024; Mereu et al., 2021).

1.3 Phenology and Stage-Specific Vulnerability

The impact of a climate stressor is not uniform across time; it is dictated by the crop's phenological stage. For instance, heat stress during the anthesis (flowering) window can cause irreversible yield loss due to pollen sterility, while moisture stress during grain-filling leads to reduced kernel weight (Monteleone et al., 2023; Shi et al., 2016). In irrigated rice systems, extreme solar radiation in early summer can lead to floodwater overheating, creating a thermal bottleneck for young plants (Bregaglio et al., 2017). Characterizing these stage-specific "signals" requires a modeling framework that can isolate monthly variations rather than relying on growing-season averages.

1.4 Modeling with Gridded Data: From Local to Regional

Traditionally, crop modeling relied on localized weather stations, which lack the spatial continuity required for regional assessment. The development of high-resolution reanalysis products like ERA5-Land has transformed this field by providing a continuous, spatially-explicit record of land-surface variables (Muñoz-Sabater et al., 2021). When integrated with the Global Dataset of Historical Yields

(GDHY) (Iizumi and Sakai, 2020), these products allow for the characterization of climate–yield relationships across entire regional landscapes, enabling a more holistic view of regional agricultural resilience (Elliott et al., 2015; Ringeval et al., 2021).

1.5 Motivation and Objectives

Despite the abundance of data, a significant gap remains in the comparative assessment of multi-crop systems. Most research focuses on single-crop responses, masking the divergent sensitivities that exist between crops sharing the same landscape. There is a need for unified computational pipelines to identify the specific thermal and hydrological bottlenecks across diverse cropping systems (Knapen et al., 2025). By applying a constant modeling logic to these five crops, this study aims to reveal the underlying biological "signals" that define the Po Valley's resilience. The primary goal of this project is to evaluate the capacity of such a unified pipeline to characterize and compare the climate vulnerabilities of five major crops in Northern Italy. To achieve this, the project addresses the following research questions:

1. To what extent can a unified computational pipeline characterize the divergent climate sensitivities of major crops in Northern Italy?
2. Which climate variables (derived from ERA5-Land reanalysis) emerge as the strongest predictors of yield across the crop portfolio?
3. How do climate sensitivities differ between cereal crops (e.g., wheat and maize) and other major regional crops?
4. Does a unified pipeline approach perform with equal efficacy across all crop types, or are there inherent biological/management-driven limits to predictability?

Specific Objectives

- **Data Integration:** Integrating high-resolution gridded yield data (Iizumi and Sakai, 2020) with climate reanalysis data (ERA5-Land) into a spatially-explicit modeling framework for a representative region.
- **Pipeline Implementation:** Develop and validate a standardized modeling workflow utilizing Elastic Net regularization for variable selection and Gamma Generalized Linear Models (GLMs) for yield characterization.
- **Comparative Characterization:** Identify and interpret the standardized coefficients of the "champion" models to map the specific thermal and hydrological bottlenecks for Maize, Rice, Soybean, Winter Wheat, and Spring Wheat.
- **Predictability Evaluation:** Compare the goodness-of-fit (CS Pseudo-R^2) across crops to identify the boundaries between climate-driven biological responses and human-managed agricultural variance.

2 Methods

The code for this project can be found at: https://github.com/Lydig/climarisc

The methodology follows a structured data science pipeline designed to characterize the divergent climate sensitivities of five distinct cropping systems in Northern Italy. The workflow consists of five

primary phases: (1) Data Acquisition, (2) Study Area Scouting and Selection, (3) Spatial-Temporal Harmonization, (4) Exploratory Collinearity and Distribution Diagnosis, and (5) Regularized Modeling and Refinement.

2.1 Data Acquisition

To quantify the relationship between climate stressors and crop yields, we integrated two distinct global gridded datasets spanning the period from 1981 to 2016.

2.1.1 Crop Yield Data

Historical crop yield estimates were obtained from the Global Dataset of Historical Yields (GDHY v1.2+1.3) (Iizumi and Sakai, 2020). This dataset provides annual yield values (measured in tonnes per hectare, t/ha) at a spatial resolution of $0.5° \times 0.5°$ (50 km). For this analysis, we acquired data for five independent cropping systems: Maize, Rice, Soybean, Spring Wheat, and Winter Wheat.

2.1.2 Climate Stressor Data

Climatological variables were sourced from the ERA5-Land reanalysis dataset (Muñoz-Sabater et al., 2021). We utilized the monthly averaged reanalysis product at a resolution of $0.1° \times 0.1°$ ($\sim$9 km) to match the seasonal scale of agricultural production. We retrieved five variables representing water supply/demand and energy inputs: 2m Temperature (measured at 2 meters above the surface; T_{2m}, hereafter T), Total Precipitation (TP, hereafter P), Volumetric Soil Water Layer 1 ($SWVL1$, hereafter SW), Surface Net Solar Radiation (SSR, hereafter SR), and Potential Evaporation (PEV, hereafter PE).

2.2 Study Area Identification and Scouting

2.2.1 Selection of Northern Italy

A global scouting process (Appendix 1, Figure 10) identified Northern Italy (approx. 43°N–47°N, 6°E–13°E)s the optimal study region. It was selected based on the intersection of three criteria:

- **Agricultural Diversity:** It represents a rare high-latitude region where C3 (Wheat, Soybean, Rice) and C4 (Maize) systems coexist in an intensive management environment.
- **Climate Vulnerability:** Northern Italy is an area known to be increasingly affected by climate-driven extremes, providing a strong signal for vulnerability modeling (Mereu et al., 2021; Monteleone and Borzí, 2024).
- **Regional Relevance:** Selecting a European system situated the project within a familiar management context, aligning with the regional food security interests of this study.

We enforced a balanced panel design, retaining grid cells with uninterrupted yield records from 1982 to 2016. The year 1981 was excluded due to data sparsity in the GDHY record. This resulted in a core study area of 41 grid cells for Rice, Soybean, and Wheat varieties, and 42 cells for Maize.

2.3 Data Processing and Harmonization

2.3.1 Spatial Alignment (Regridding)

To reconcile the spatial mismatch between datasets, ERA5-Land data was upscaled to the 0.5° GDHY grid using a block-averaging technique:

$$V_{0.5^\circ}(lat, lon) = \frac{1}{N} \sum_{i=1}^{N} V_{0.1^\circ}(i) \tag{1}$$

where $V_{0.1^\circ}(i)$ represents the values of the $N = 25$ climate grid cells (0.1°) falling within the footprint of a single 0.5° yield cell.

2.3.2 Temporal Aggregation and Crop-Specific Seasons

Monthly climate signals were aggregated into seasonal metrics based on the specific growing windows for each system:

- **Maize, Rice, and Soybean:** May to September (MJJAS).
- **Spring Wheat:** March to July (MAMJJ).
- **Winter Wheat:** November to June (NDJFMAMJ).

Aggregation logic was based on variable type: State Variables (T, SW) were averaged over the season, while Flux Variables (TP, SR, PE) were calculated as cumulative sums. For Winter Wheat, a "Crop Year" logic assigned climate data from the preceding calendar year (November–December) to the harvest year.

2.4 Exploratory Analysis and Collinearity Diagnosis

Prior to modeling, we conducted a systematic visual inspection of the dataset using Scatterplot Matrices (SPLOMs). The SPLOM diagnostics identified severe multicollinearity across the monthly climate variables, particularly between sequential temperature and potential evaporation measurements. This visual confirmation of redundant information necessitated a regularized modeling approach capable of objective variable selection.

2.5 The Statistical Modeling Pipeline

2.5.1 Generalized Linear Modeling (Gamma GLM)

We employed a Gamma GLM with a log link function (McCullagh and Nelder, 1989). The choice of the Gamma distribution was driven by three theoretical considerations specific to agricultural data:

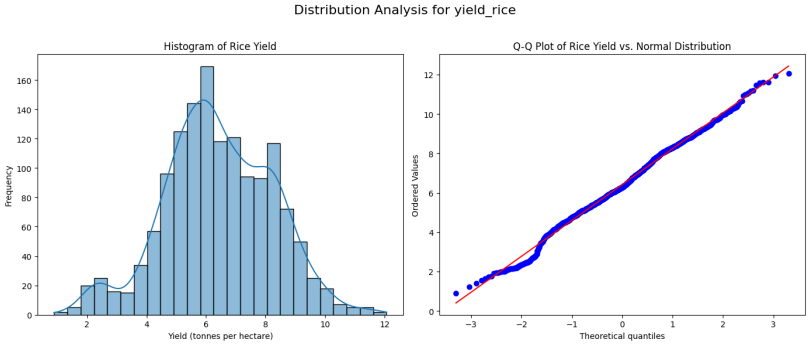
1. **Strict Positivity:** Crop yields are strictly non-negative, and the Gamma distribution naturally enforces this constraint, unlike Gaussian models which can predict negative yields.
2. **Mean-Variance Relationship:** Biological data typically exhibits heteroscedasticity, where variance increases with the mean.
3. **Multiplicative Effects:** The log link function $ln(\mu)$ implies that climate stressors act multiplicatively (proportional yield reductions) rather than additively, a standard assumption in agricultural vulnerability research (McCullagh and Nelder, 1989).

While visual inspection of histograms and Q-Q plots (Figure 1) indicated varying degrees of skewness across the crops, with Rice approaching normality while Winter Wheat showed distinct tails, the Gamma GLM framework was still selected based on the domain standard for crop yield data.

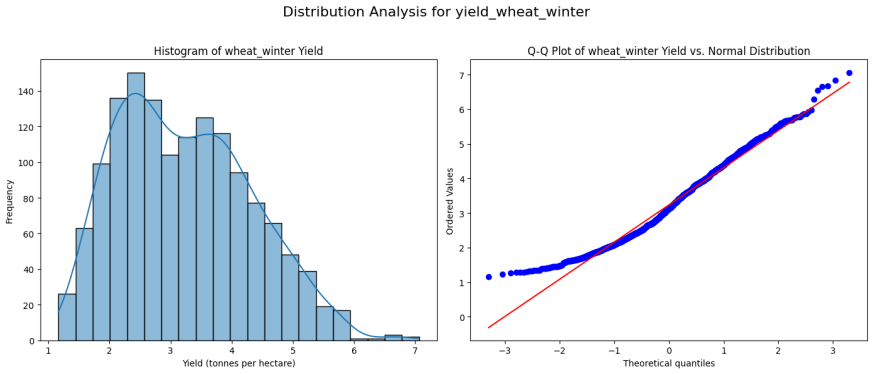
2.5.2 Spatial Modeling and Technological Trends

To isolate the climate signal, we included two control co-factors:

1. **Technological Trend:** A linear *Year* variable was included as a covariate to account for long-term improvements in management and genetics (internal detrending).
2. **Spatial Modeling via B-splines:** To account for spatial autocorrelation and unobserved geographical drivers—such as soil characteristics and terrain—we modeled spatial heterogeneity using B-spline basis functions on Latitude and Longitude (Hastie et al., 2009).



(a) Rice Yield Distribution



(b) Winter Wheat Yield Distribution

Figure 1: Distributional analysis (Histograms and Q-Q plots) for two representative cropping systems. The observed skewness and departure from the normal reference line in the Q-Q plots justify the use of a Gamma Generalized Linear Model (GLM).

Unlike treating grid cells as discrete categorical entities (fixed effects), B-splines represent spatial variation as a continuous, linear combination of basis functions. This approach recognizes that baseline agricultural productivity varies smoothly across the region. We specified four degrees of freedom for the spline bases, a standard parameterization that captures regional-scale gradients without over-fitting localized noise.

2.5.3 Variable Selection via Elastic Net Regularization

To handle the multicollinearity identified in the SPLOMs, we utilized Elastic Net Regularization (Zou and Hastie, 2005). All predictors were Z-score standardized for this phase to ensure a fair penalty application. This is a critical requirement for penalized regression; because L1 and L2 penalties are scale-sensitive, standardization ensures that the regularization is applied uniformly across all features regardless of their original units (Hastie et al., 2009). The mixing parameter was set to 0.5 to balance Lasso (selection) and Ridge (grouping) properties. Variables that maintained non-zero coefficients at the optimal logarithmic `alpha` path were retained for final explanatory modeling.

2.5.4 Standardization Strategy and Numerical Stability

To satisfy the distinct mathematical requirements of each pipeline stage, a three-step standardization strategy was applied:

1. **Feature Selection (Elastic Net):** All predictors were Z-score standardized. This was strictly necessary to ensure the L1/L2 penalties were applied fairly across variables with different units (e.g., mm vs °C).
2. **Model Fitting (GLM):** For the final explanatory models, variables were reverted to their original physical units to ensure agronomic interpretability (e.g., coefficients representing yield change per °C). **Exception:** Solar Radiation (SR) and its squared terms (SR^2) were retained in Z-score format. The magnitude of raw solar radiation (Joules/m^2) is exponentially larger than other variables, which caused convergence failures in the non-linear optimization algorithms.
3. **Comparative Analysis:** Post-hoc standardized coefficients (β_{std}) were calculated (see Section 2.6.2) to allow for the direct ranking of effect sizes across different crops and climate drivers.

2.6 Model Refinement and Evaluation

Final models were refitted using the selected variables in their original physical units (except where standardization was required for stability) to produce agronomically interpretable vulnerability curves. We performed a hypothesis-driven refinement using the Akaike Information Criterion (AIC) to test for:

1. **Non-Linearity:** Testing for optimal thresholds by adding quadratic terms (x^2).
2. **Interaction Discovery:** Testing agronomically plausible interactions, such as Resource Buffering (Heat × Water) and Temporal Compounding (Early-season × Late-season Temperature).

Terms were only retained if they significantly reduced the AIC and adhered to the Principle of Marginality.

2.6.1 Performance Metric: Cox-Snell Pseudo-R^2

Final model performance was assessed using the Cox-Snell Pseudo-R^2. Unlike the standard R^2 used in OLS, Pseudo-R^2 is derived from likelihood ratios and is a robust choice for non-Gaussian models like the Gamma GLM. This value represents the improvement in predictive power provided by the climate stressors and spatial controls relative to a null (intercept-only) model. A higher

value indicates that the model successfully captures a larger proportion of the year-to-year yield variability driven by the environment (Cox and Snell, 1989).

2.6.2 Coefficient Standardization for Comparative Analysis

To facilitate a direct comparison of the relative impact of disparate climate stressors (measured in varying units such as °C, mm, and J/m^2) and to compare effects across different cropping systems, a post-hoc standardization of the final model coefficients was performed. Standardized coefficients (β_{std}) were calculated by adjusting the raw model estimates (β_{raw}) by the ratio of the standard deviation of the predictor (SD_x) to the standard deviation of the yield (SD_y):

$$\beta_{std} = \beta_{raw} \times \frac{SD(Predictor)}{SD(Yield)}$$

This transformation yields a dimensionless metric, representing the change in the outcome variable (in standard deviation units) for every one standard deviation increase in the climate stressor, holding all other covariates constant. This approach is essential for the comparative nature of this study, as it accounts for the differing variances in yield across the five crop types and allows for the objective ranking of "vulnerability signals" presented in the final analysis (Table 4).

All computational analyses were implemented in Python (v3.10.18) using the statsmodels library for GLM estimation (Seabold and Perktold, 2010) and scikit-learn for the Elastic Net regularization.

3 Results

3.1 Exploratory Data Analysis and Pipeline Initiation

The first stage of the computational pipeline focused on characterizing the statistical properties of the integrated datasets and narrowing the high-dimensional feature space of ERA5-Land climate stressors. Given the inherent complexity of gridded agricultural data, this phase was important for ensuring that the following modeling choices were both statistically robust and biologically grounded.

3.1.1 Distributional Assumptions and Model Justification

A fundamental requirement for reliable yield modeling is the selection of an appropriate error distribution. Initial EDA revealed that crop yields across the Northern Italian study area exhibited significant departures from normality. As illustrated in the distributional analysis of two representative crops (Winter Wheat and Rice; see Figure 1 in Methods), the yield data presented some skewness and non-canonical tails in the Q-Q plots. These observations confirmed that a standard Ordinary Least Squares (OLS) regression, which assumes normally distributed residuals, would be inappropriate. Consequently, the pipeline was initiated using a Generalized Linear Model (GLM) framework with a Gamma distribution and a log link function. This choice is mathematically optimized for strictly positive, continuous data with a variance that increases with the mean, a common characteristic of biological yield data.

3.1.2 Multicollinearity and Feature Space Reduction

A primary challenge in agro-climate modeling is the high degree of multicollinearity between environmental variables. As demonstrated in the Pairwise Correlation Matrix for Rice (Figure 2), several stressors exhibit significant multicollinearity that justifies a penalized regression approach. Specifically, the matrix reveals a strong temporal priming effect; solar radiation in the early season (May and June) shows a moderate-to-high positive correlation ($r \approx 0.25$ to 0.70) with temperatures across all subsequent months. Interestingly, this relationship weakens for late-season radiation, suggesting that early-season energy inputs are a primary driver of seasonal warming in Northern Italy. Furthermore, a distinct negative feedback loop is visible between soil water and monthly temperatures; lower soil moisture levels are consistently associated with higher temperatures, likely due to the reduction in latent heat cooling. These redundant and intertwined features, including the expected negative correlation between precipitation and solar radiation (cloud cover), create a feature space that can lead to model instability in standard regression.

To address this, the pipeline utilized Elastic Net Regularization to identify the most parsimonious set of predictors. The Regularization Path (Figure 3) visualize this process, showing how the model coefficients are penalized as the alpha (penalty strength) increases. We used the visuals from this to choose our alpha for the final Elastic Net. This data-driven approach allowed the pipeline to "shred" the noise of redundant climate signals, successfully reducing the initial pool of monthly stressors down to the 7–12 core variables that exert the most significant influence on yield for each specific crop.

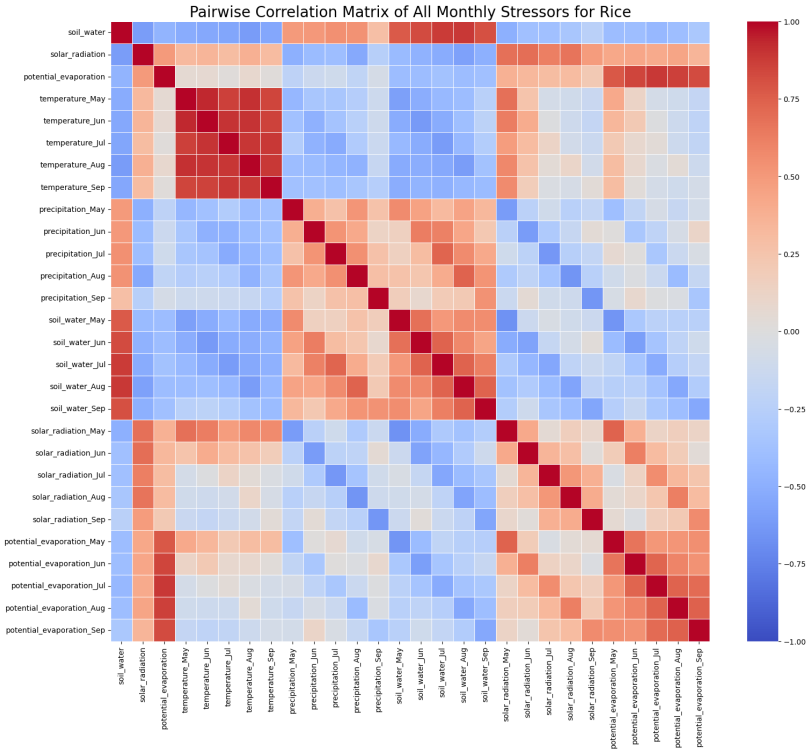


Figure 2: Pairwise Correlation Matrix of all monthly stressors for Rice. The heatmap illustrates strong intra-seasonal and inter-variable correlations among climate variables, such as Temperature, Potential Evaporation, and Soil Water, highlighting the high degree of multicollinearity in the initial feature space.

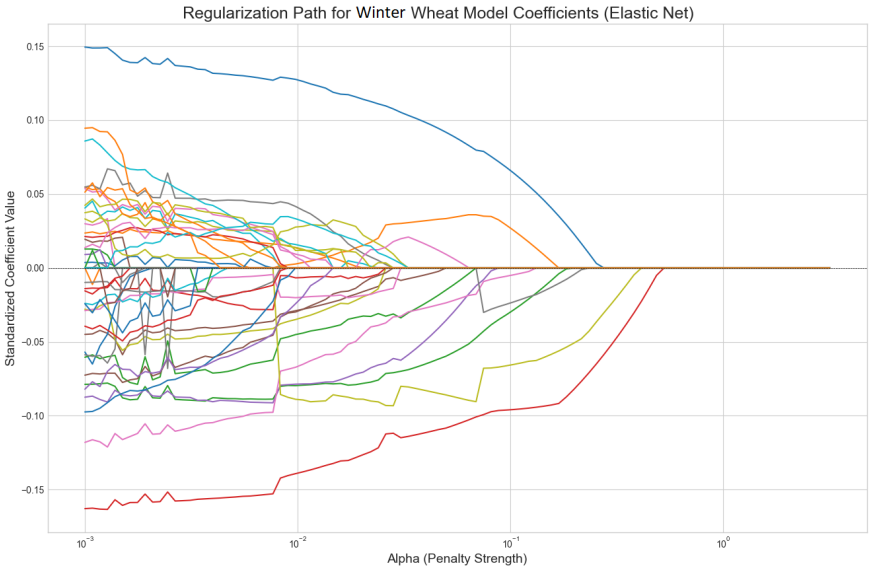


Figure 3: Regularization path for Winter Wheat model coefficients (Elastic Net). The plot illustrates the shrinkage of climate coefficients toward zero as the penalty strength (α) increases. This data-driven approach allows the pipeline to "shred" redundant noise and select only the most parsimonious predictors for the final GLM.

3.2 Iterative Model Optimization: The Rice Case Study

To evaluate the effectiveness of the pipeline's iterative refinement process, Rice (Oryza sativa) was selected as a representative case study. Rice cultivation in Northern Italy is uniquely dependent on complex water management and high energy inputs during specific phenological windows, making it an ideal candidate for testing non-linear and interaction-based model structures.

3.2.1 Evolution of Model Parsimony and Explanatory Power

As detailed in Table 1, the model underwent a four-step optimization process. The initial baseline model (step 1), constructed using linear predictors selected via Elastic Net, provided a strong foundation with a Pseudo-R^2 of 0.699. However, the subsequent introduction of non-linear and interaction terms resulted in a consistent reduction in the Akaike Information Criterion (AIC), indicating that the increased model complexity was statistically justified.

Table 1: Iterative Development and Optimization of the Rice Yield Model

Step	**Model Description**	**AIC**	**ΔAIC**	**Pseudo-R^2**
1	**Baseline:** Linear terms selected via Elastic Net and manually	4841.18	–	0.699
2	**Quadratic:** Addition of non-linear term PE^2_{May}	4815.85	-25.33	0.713
3	**Interaction:** Addition of interaction term $PE_{May} \times SW_{Aug}$	4806.65	-9.20	0.717
4	**Refined Stressor:** Growing Season $SR \rightarrow SR_{June}$	**4800.07**	**-6.58**	**0.719**

Note: PE: Potential Evaporation; *SW*: Soil Water; *SR*: Solar Radiation. ΔAIC is calculated relative to the immediate previous step. Step 4 represents the final champion model for Rice, achieving the lowest AIC and highest explanatory power.

The most significant single improvement in model fit occurred in Step 2 with the addition of a quadratic term for Potential Evaporation in May (PE^2_{May}). This addition reduced the AIC by 25.33 units, suggesting that the relationship between early-season evaporative demand and yield is intrinsically non-linear. As visualized in (Figure 4), yield potential accelerates at higher evaporation levels, identifying a thermal 'sweet spot' for early-season establishment.

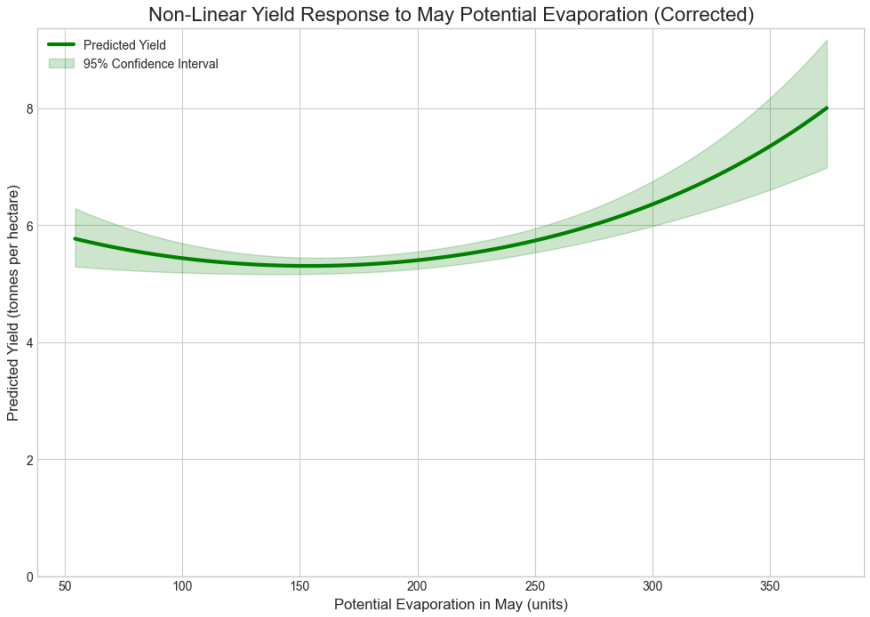


Figure 4: Non-linear yield response of Rice to Potential Evaporation (PE) in May. The convex (upward-curving) relationship suggests a thermal "sweet spot" (> 300 units) for early-season establishment, where high evaporative demand serves as a proxy for the spring warmth required to escape cold-stress windows.

In the context of the Po Valley, May is a critical period for sowing and early tillering. While moderate evaporation is necessary for healthy transpiration, the "sweet spot" for yield is found at the higher end of the PE_{May} range (> 300 units). This suggests that high early-season energy inputs, representing warmer, sunnier springs, are vital for robust plant establishment. In this region, where April and May can often be too cold for optimal rice germination, higher potential evaporation serves as a proxy for the thermal accumulation required to escape the "cold-stress" window, thereby setting a higher trajectory for final yield Bregaglio et al. (2017).

3.2.2 Synergies and Monthly Refinements

Following the confirmation of non-linearity, Step 3 introduced an interaction term between May Potential Evaporation and August Soil Water ($PE_{\text{May}} \times SW_{\text{Aug}}$). This was a key refinement based on the hypothesis that early-season energy demand influences the crop's resilience to water availability during the critical grain-filling stage in August. This hypothesis is supported by physiological evidence suggesting that early thermal accumulation determines the crop's total leaf area and biomass, which subsequently dictates the transpiration demand and hydrological sensitivity during the reproductive phase (Bregaglio et al., 2017; Fahad et al., 2018). The interaction term lowered the AIC by an additional 9.20 units. It should be noted that while this interaction is statistically significant ($p < 0.05$), the confidence intervals in the interaction plot (Figure 5) show some overlap.

This suggests that while the synergistic effect is real, it plays a secondary role compared to the dominant main effects of hydrological availability.

As shown in the Interaction Effect Plot (Figure 5), the yield benefit derived from August soil water is dependent on the energy levels experienced in May. When early-season energy is high (High May PE; Red Line), the crop achieves a higher overall yield baseline and appears less sensitive to fluctuations in late-summer soil moisture. Conversely, when May is cool and energy-limited (Low May PE; Orange Line), the crop becomes more reliant on high August soil water to recover yield potential. This suggests that a strong start in May "primes" the rice plant, making it more robust against the hydrological stresses of the grain-filling stage in August.

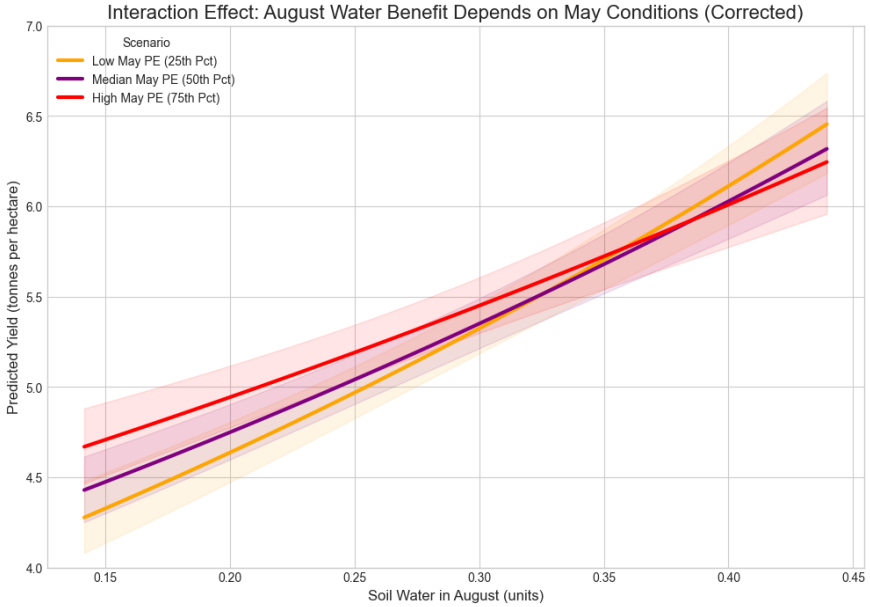


Figure 5: Interaction effect between May Potential Evaporation and August Soil Water for Rice. The divergent slopes demonstrate that the benefit of late-season soil moisture is magnified by high energy inputs during May, suggesting that early thermal accumulation "primes" the plant's hydrological sensitivity.

The final refinement (Step 4) involved a pivot from the growing-season average solar radiation to a month-specific stressor: June Solar Radiation (SR_{June}). The full final model (See equation (3) in Appendix 2) achieved the lowest overall AIC (4800.07) and the highest explanatory power (Pseudo-$R^2 = 0.719$) (See Table 2).

Table 2: Overview of Model Specifications and Performance Results

Crop Name	Grid Cells	Time Span	Total Obs.[a]	EN Sel.[b]	Final Number of Predictors (Breakdown)[c]	CS Pseudo R^2	AIC
Maize	42	1982–2016	1,470	8	8 (7+1NL)	0.8755	5,309.83
Rice	41	1982–2016	1,435	7	8 (6+1NL+1I)	0.7194	4,800.07
Soybean	41	1982–2016	1,435	9	11 (7+2NL+2I)	0.8513	3,447.76
Winter Wheat	41	1982–2016	1,435	9	12 (8+2NL+2I)	0.9101	2,476.85
Spring Wheat	41	1982–2016	1,435	8	12 (8+3NL+1I)	0.5762	3,467.84

[a] Total Observations = years (35) × grid cells.
[b] Number of predictors selected by Elastic Net.
[c] Breakdown: (Linear terms + NL: Non-linear terms + I: Interaction terms).

Unlike the beneficial energy signals found in May, the Yield Response to June Solar Radiation (Figure 6) reveals a clear, negative linear trend. In Northern Italy, June represents the critical rooting and early reproductive phases (Bregaglio et al., 2017; Poggi et al., 2022). While solar radiation is necessary for photosynthesis, extreme levels in June often act as a proxy for the severe Mediterranean heatwaves and droughts that have become more frequent in the Po Valley. These conditions can lead to excessive water temperatures in paddies and reduced discharge from the Po River, causing thermal stress during the plant's sensitive development stages. By isolating June as a standalone stressor, the pipeline successfully separated the beneficial "establishment" energy of May from the detrimental "heat-stress" energy of June.

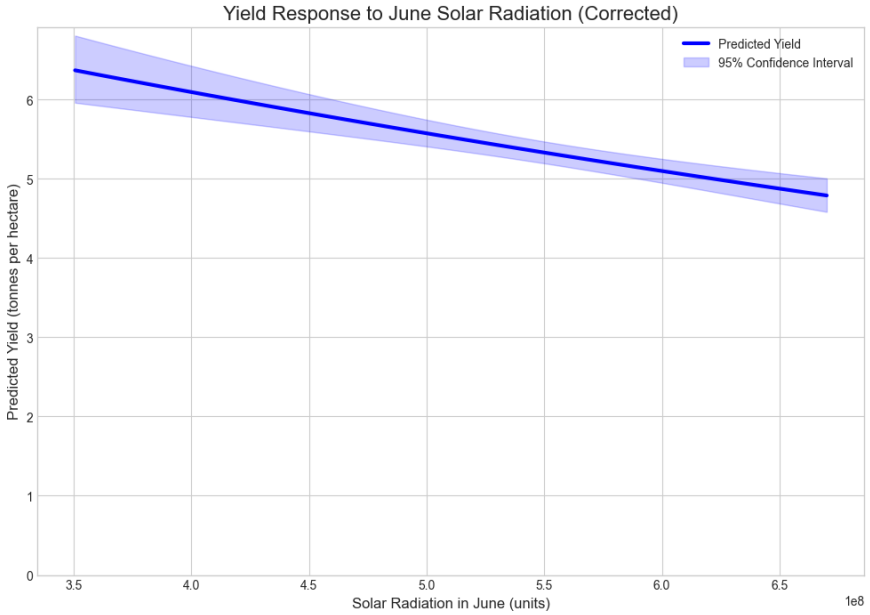


Figure 6: Yield response of Rice to June Solar Radiation (SR_{June}). In contrast to beneficial May energy, the negative linear trend identifies a regional bottleneck where extreme June radiation acts as a proxy for floodwater overheating and "thermal scalding" during the sensitive rooting phase.

3.2.3 Summary of the "Champion" Model

The refinement process confirms that for Rice in Northern Italy, the specific timing of energy inputs (June) and the non-linear interaction between evaporation and soil moisture are essential for characterizing yield volatility. By allowing the pipeline to move beyond simple linear averages, we successfully captured the specific environmental "bottlenecks" that define Rice's vulnerability in this region.

The finalized standardized coefficients for this optimized Rice model, which represent the relative strength of each climate signal, are integrated into the broader comparative overview of all study crops in Table 4. This final model serves as the basis for the characterization of divergent climate sensitivities discussed in Section 3.4 and further interpreted in Section 4.2.

3.3 Comparative Performance Evaluation

While the iterative refinement process proved successful for the Rice case study, applying the unified pipeline across the full portfolio of Northern Italian crops reveals a distinct spectrum of predictability. This section evaluates how well the computational framework captured the climate-yield relationships for all five crops, as summarized in Table 2 (Performance Overview) and Table 3 (Baseline vs. Final Comparison).

Table 3: Model Performance Comparison: Baseline vs. Final Refined Models

Crop	Baseline Pseudo R^2	Final Pseudo R^2	% Improv.	Baseline AIC	Final AIC	ΔAIC
Maize	0.8722	0.8755	0.38%	5327.32	5309.83	-17.49
Rice	0.6988	0.7194	2.95%	4841.18	4800.07	-41.11
Soybean	0.8294	0.8513	2.64%	3516.86	3447.76	-69.10
Winter Wheat	0.8555	0.9101	6.38%	2681.90	2476.85	-205.05
Spring Wheat	0.5531	0.5762	4.18%	3518.87	3467.84	-51.03

Note: Baseline results represent the initial linear predictors selected by the computational pipeline. Final models incorporate non-linear (quadratic) terms and/or climate interaction effects. Improvement is calculated as the relative increase in Pseudo-R^2 from the baseline to the final model.

3.3.1 The High-Predictability Group: Winter Wheat, Maize, and Soybean

The pipeline achieved really high explanatory power for three of the five crops, with Pseudo-R^2 values exceeding 0.85. Winter Wheat emerged as the "champion" model of the study, achieving a Final Pseudo-R^2 of 0.9101. The refinement process was particularly impactful for this crop; the transition from the linear baseline to the final interaction model resulted in a massive AIC reduction of 205.05 units (Table 3), the largest improvement across the study.

Similarly, Maize (0.8755) and Soybean (0.8513) exhibited high sensitivity to the climate stressors identified by the pipeline. Because these crops are cultivated during Northern Italies peak summer window (May–September), their growth is largely governed by the interplay between high evaporative demand and the managed water supply. The high Pseudo-R^2 values suggest that for these heavily irrigated systems, the gridded ERA5-Land variables (specifically Soil Water and July Temperature) serve as excellent proxies for the regional climate signals that drive yield volatility. Essentially, the standardized management of these crops 'clarifies' the climate signal, making them more predictable for the pipeline than the more opportunistically managed Spring Wheat.

3.3.2 The Modeling Challenge: Spring Wheat and the "Predictability Gap"

In contrast to the high performance seen in Winter Wheat, Spring Wheat presented the greatest challenge for the unified pipeline, achieving a Final Pseudo-R^2 of 0.5762. While the refinement process did improve the model (reducing AIC by 51.03 units from the baseline), a significant "predictability gap" remains compared to the other crops.

This "predictability gap" likely stems from the high sensitivity of spring-sown cereals to human intervention, effectively introducing "Management Noise" into the signal. While Winter Wheat follows a standardized phenological calendar synchronized with the climate, Spring Wheat in Northern Italy is often grown opportunistically (e.g., as a catch crop or following winter crop failure). Consequently, yield variability is driven less by the regional climate signal and more by field-level decisions—such as variable sowing dates and irrigation intensity—that are obscured by the 55km spatial resolution of the reanalysis data.

3.3.3 Unified Pipeline Validation

Despite the variation in absolute Pseudo-R^2 values, the methodological logic of the pipeline was validated across all five cropping systems. As shown in Table 3, the iterative refinement process (adding non-linear and interaction terms) resulted in a lower AIC and a higher Pseudo-R^2 in 100% of the cases.

The improvement was not uniform: Winter Wheat saw a 6.38% relative increase in goodness-of-fit, while Maize saw only 0.38%. This indicates that while the pipeline is universally applicable, the complexity of the climate-yield relationship is crop-specific. Some crops (like Maize) are governed by a few dominant linear stressors, whereas others (like Winter Wheat and Rice) require the inclusion of complex interactions and non-linear "sweet spots" to truly characterize their vulnerability.

3.4 Characterization of Climate Vulnerabilities (The "Signals")

3.4.1 Thermal Vulnerabilities: The Wheat Divergence

The unified pipeline revealed a clear phenological divide between Winter and Spring Wheat, particularly in their sensitivity to the transition from winter to spring. While these varieties share a similar botanical structure, their divergent "thermal memories" create opposite yield responses to March and May temperatures.

The "March Paradox": Growth vs. Stress As detailed in Table 4, the model identified a significant contradiction in March temperature sensitivity. Winter Wheat exhibits a strong positive coefficient (+0.3051), while Spring Wheat shows a negative signal (−0.0731). This divergence could be rooted in the distinct developmental stages of the two crops during this window:

- **Winter Wheat:** Having over-wintered, the crop is established. Warmth in March facilitates rapid spring tillering and biomass accumulation (Huzsvai et al., 2022; Poggi et al., 2022).
- **Spring Wheat:** Sown in March, the crop is in its most fragile germination phase. High temperatures increase soil evaporation, causing "seedling thirst" and harming early root establishment (Mereu et al., 2021; Monteleone et al., 2023).

Table 4: Final Model Standardized Coefficients and Interaction Effects

Terms	**Maize**	**Rice**	**Soybean**	**W. Wheat**	**S. Wheat**
Temperature (March)	–	–	–	***0.3051	***-0.0731
Temperature (May)	–	–	–	-0.1221	***0.3191
Temperature (July)	0.0114	–	–	–	**-0.1985
Temperature (November)	–	–	–	***-0.0615	–
Soil Water	***0.0331	–	***0.1138	–	–
Soil Water (August)	–	***0.0914	–	–	–
Precipitation (April)	–	–	–	–	***0.0639
Precipitation (May)	–	–	***0.0706	–	–
Precipitation (Sept)	***0.0098	–	–	–	–
Pot. Evap. (Jan)	–	–	–	***-0.1638	–
Pot. Evap. (Apr)	–	–	–	***0.0287	**0.0207
Pot. Evap. (May)	***0.0800	-0.0302	***0.2109	–	–
Solar Rad. (June)	–	***-0.0209	–	–	–
Solar Rad. (July)	–	–	0.0308	–	–
Non-Linear Terms					
Temperature (March)^2	–	–	–	***0.0162	–
Temperature (May)^2	–	–	–	0.0067	**0.0101
Temperature (July)^2	***-0.0021	–	–	–	**0.0124
Precipitation (April)^2	–	–	–	–	**-0.00001
Precipitation (May)^2	–	–	***-0.00001	–	–
Pot. Evap. (May)^2	–	***0.0009	–	–	–
Solar Rad. (July)^2	–	–	***-0.0255	–	–
Interaction Terms					
Temp (Mar) × Temp (May)	–	–	–	***-0.1641	–
Temp (May) × Temp (July)	–	–	–	–	***-0.1274
Pot. Evap. (Jan) × Temp (May)	–	–	–	***0.0564	–
Pot. Evap. (May) × Soil Water (Aug)	–	***-0.0498	–	–	–
Pot. Evap. (May) × Solar Rad. (July)	–	–	***-0.0904	–	–
Solar Rad. (July) × Soil Water	–	–	**-0.0221	–	–

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.
Pot. Evap. = Potential Evaporation; Solar Rad. = Solar Radiation; W. Wheat = Winter Wheat; S. Wheat = Spring Wheat.
Values represent standardized coefficients. Dash (–) indicates terms not selected for the specific crop model.
Principle of Marginality: Main effects appearing without significance stars (e.g., PE_{May} for Rice) are retained because they are required mathematical components of significant interaction or quadratic terms.

Compounding Vulnerabilities: The Interaction of Early and Late Heat The "signals" of vulnerability are not isolated to single months; they are compounding. This is demonstrated by the highly significant interaction terms in the final models.

For Winter Wheat (Figure 7): The model identifies a strong negative interaction between March and May temperatures (-0.1641). As seen in the Interaction Plot (Figure 7), while a warm March is generally beneficial, the vulnerability to May heat is compounded by a warm March. If the season starts warm in March, the wheat plant accelerates its development. If this is followed by a hot May, the plant is "caught" in the sensitive flowering (anthesis) phase earlier and more intensely. The steeper decline in the response for a warm March compared to a cold March suggests that an early spring "primes" the plant for greater damage when summer heatwaves arrive prematurely.

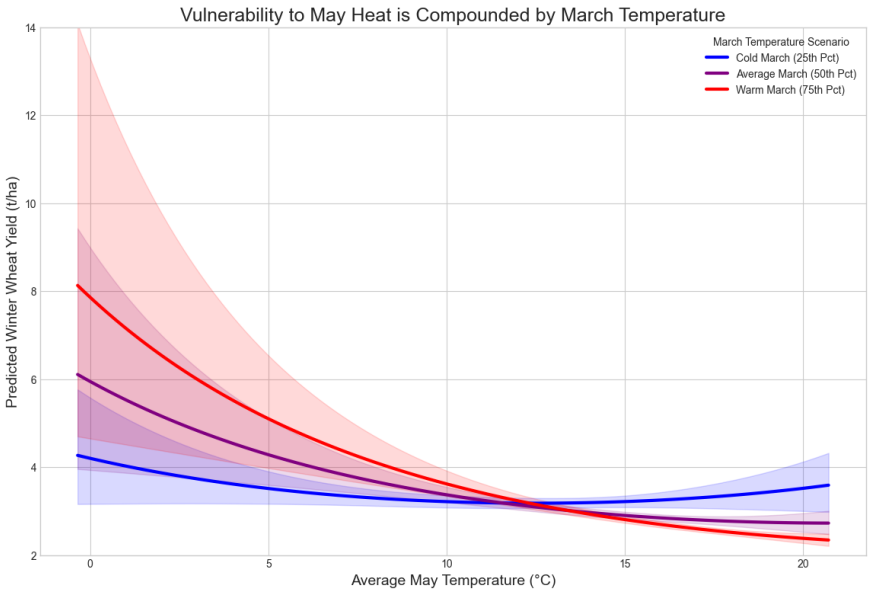


Figure 7: Interaction plot for Winter Wheat. The steeper decline in the "Warm March" scenario (red) relative to a "Cold March" (blue) demonstrates how early-season phenological acceleration compounds the crop's vulnerability to late-spring heatwaves during anthesis.

For Spring Wheat (Figure 8): A similar "double-hit" logic applies later in the season. Table 4 shows a positive coefficient for May Temperature (+0.3191) but a negative one for July (−0.1985). For Spring Wheat, May is still a "growth" phase where warmth aids biomass. However, July is the critical grain-filling period (Monteleone et al., 2023; Poggi et al., 2022). The negative interaction between May and July (−0.1274) shows that high May temperatures exacerbate the yield loss from July heatwaves. A plant that was forced to grow rapidly in May has a larger leaf surface area and higher transpiration demand, making it much more susceptible to "crashing" when July temperatures spike.

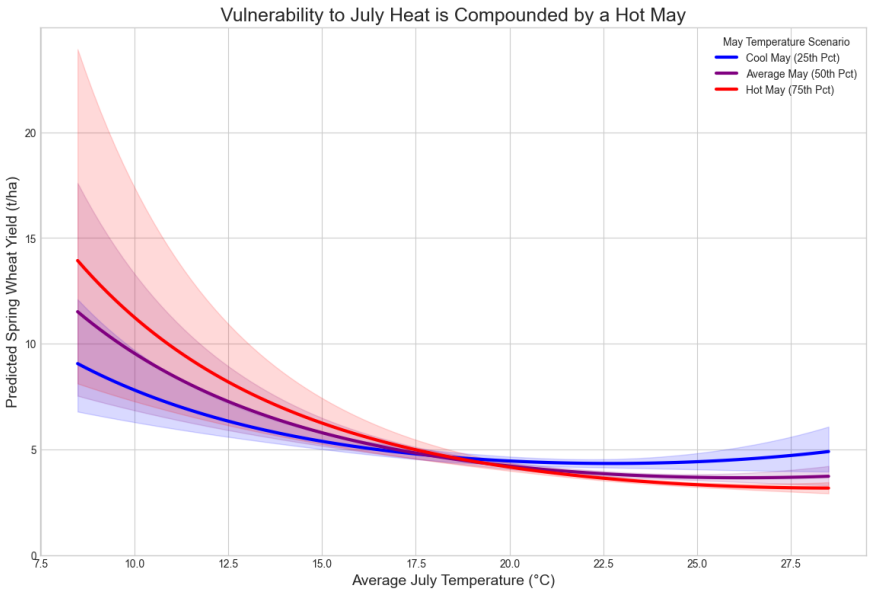


Figure 8: Interaction plot for Spring Wheat. High temperatures during the May growth phase (red) exacerbate yield losses from July heatwaves. This "double-hit" logic suggests that rapid early biomass development increases transpiration demand, making the crop more susceptible to "crashing" during grain-filling.

Summary of Divergent Sensitivity The pipeline successfully characterized that for Winter Wheat, the "pivot point" for vulnerability is the March–May transition, whereas for Spring Wheat, it is the May–July window. This finding proves that a "one-size-fits-all" climate adaptation strategy is insufficient; farmers must manage these crops based on their specific thermal windows of vulnerability.

3.4.2 Hydrological Dominance

While thermal stressors define the "Wheat Paradox" in spring, the summer-sown crops (Maize and Soybean) are governed by a different regime: Hydrological Dominance. For these crops, which occupy the peak heat window of May to September, water availability is the primary determinant of yield stability.

Soil Water as a Proxy for Irrigation Resilience As detailed in Table 4, Soil Water emerged as the strongest and most consistently positive predictor for the summer crops. Specifically, Soybean (0.1138^{***}) and Maize (0.0331^{***}) show high sensitivity to moisture levels.

The statistical preference for Soil Water over raw Precipitation (P) is a key finding. It suggests that Soil Water serves as a superior proxy for the actual moisture available to the plant, effectively integrating both natural rainfall and the extensive human-managed irrigation network characteristic of the Po Valley. The high significance of Soil Water over raw Precipitation is a key finding of the pipeline; it suggests that Soil Water serves as a better proxy for the actual moisture available to the plant, integrating both natural rainfall and human-managed irrigation. As confirmed by the yield response in (Figure 9), these systems are essentially "water-limited" in their current climatic context.

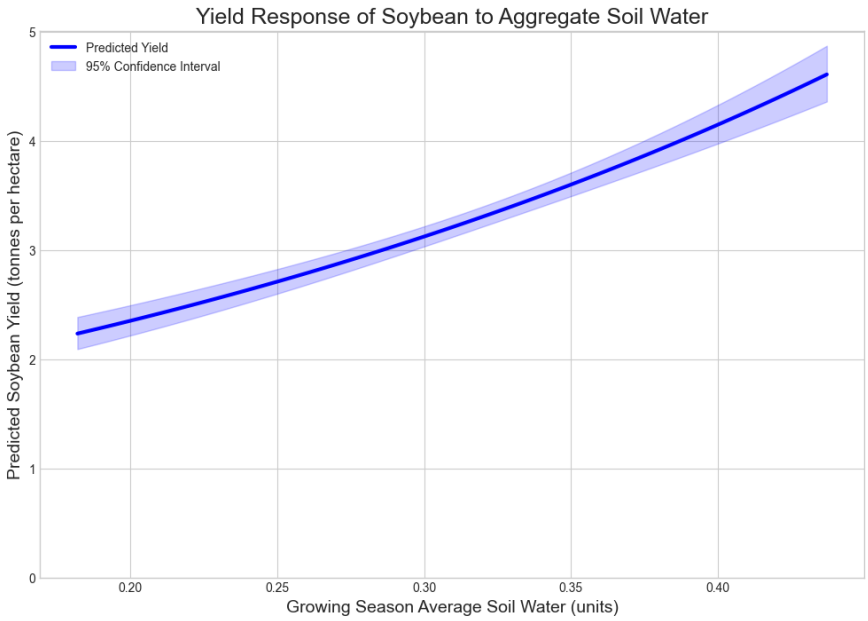


Figure 9: Standardized yield response of Soybean to aggregate Soil Water. The strong, positive linear relationship confirms the "hydrological dominance" of summer-sown crops in the Po Valley, where yield stability is governed by the managed water buffer rather than raw precipitation.

The "Thirst" of Maize and Soybean The vulnerability profiles for these two crops, however, show slight divergences in timing:

- **Soybean:** In addition to Soil Water, Soybean shows a positive response to Precipitation in May (0.0706^{***}). This suggests that early-season moisture is good for soybean establishment, likely to ensure a deep root system (Ordóñez et al., 2020) before the onset of July heatwaves
- **Maize:** Maize shows a secondary significant signal for Precipitation in September (0.0098^{***}). While lower in magnitude, this signal captures the vulnerability of the grain-filling stage. Maize that lacks moisture during the final starch-accumulation phase in September suffers from shriveled kernels and reduced test weight (Çakir, 2004).

Potential Evaporation as a Proxy for Thermal Accumulation A notable finding in Table 4 is the strong positive influence of Potential Evaporation in May (PE_{May}) for both Maize (0.0800^{***}) and Soybean (0.2109^{***}). While PE is typically viewed as a stressor representing water loss, its positive coefficient during the early growing season reflects its role as a proxy for thermal accumulation and soil warming.

In the Po Valley, May temperatures frequently fluctuate near the lower biological threshold for summer crop germination. High PE_{May} values are indicative of clear-sky conditions and high vapor pressure deficits, which accelerate the warming of the topsoil and irrigation water (Bregaglio et al., 2017; Mereu et al., 2021). This early-season "energy boost" is critical for robust plant establishment and root architecture development. The model successfully captures that in Northern Italy, the early-season yield trajectory is limited by energy (warmth), whereas the mid-to-late season trajectory (July–August) is limited by water (Soil Water). This creates a dual-phase vulnerability profile: crops require high energy to establish in May but rely on hydrological buffering to survive the resulting atmospheric demand in July.

3.4.3 Energy Dynamics: The Rice and Solar Radiation Mystery

In most agronomic models, solar radiation is treated as a positive driver of yield due to its role in fueling photosynthesis. However, the unified pipeline identified a significant and counter-intuitive negative coefficient for June Solar Radiation (SR_{June}) in the Rice model (-0.0209^{***}, Table 4). This negative "signal" reveals a specific regional vulnerability inherent to the irrigated paddy systems of Northern Italy.

The "Overheating" Signal As visualized in the Yield Response Plot (Figure 6), there is a clear downward linear trajectory: as solar radiation in June increases, the predicted yield consistently declines. While the magnitude of the coefficient is small, its high statistical significance suggests that SR_{June} acts as a proxy for an environmental stressor that overrides the benefits of increased light availability.

3.4.4 Non-Linearity and Synergies: Capturing Complex Vulnerabilities

A core strength of the unified computational pipeline is its ability to move beyond simple linear assumptions. In biological systems, the relationship between climate and yield is rarely additive; instead, it is defined by thresholds, "tipping points," and compounding stresses. The final models (Table 4 & Table 2) confirm that non-linear and interaction terms were critical for maximizing the explanatory power across all five crops. Consequently, certain linear main effects lacking statistical significance were retained in the final specification to satisfy the Principle of Marginality, providing the necessary baseline for these higher-order signals.

Tipping Points and Quadratic Thresholds As detailed in the Non-Linear Terms section of Table 4, several crops exhibit significant quadratic responses ($Term^2$). These terms identify biological thresholds where a variable that is beneficial at low levels becomes detrimental at high levels.

- **Spring Wheat (T^2_{Jul}):** The significant quadratic term for July temperature (0.0124^{**}) suggests a "curvilinear" vulnerability. While moderate July warmth might be tolerable, once a specific thermal threshold is crossed during the grain-filling stage, the yield loss accelerates exponentially.
- **Soybean (SR^2_{Jul}):** Similarly, the negative quadratic signal for July Solar Radiation (-0.0255^{***}) identifies an energy threshold. While radiation is necessary for photosynthesis, excessive radiative energy in the peak of summer likely leads to photo-inhibition or extreme transpiration demand that the plant cannot sustain, causing a non-linear "crash" in yield potential.

The "Compounding Stress" of Spring: Winter Wheat Synergy The most complex signals identified were the interaction terms, which describe how the impact of one variable depends on the state of another. This is most clearly demonstrated in the Winter Wheat model.

As visualized in the Interaction Plot (Figure 7), the relationship between May temperature and yield is significantly altered by the conditions experienced in March ($T_{Mar} \times T_{May} = -0.1641^{***}$). The biological logic here could be that a warm March "primes" the winter wheat by accelerating its phenological development. While this leads to high initial biomass, it pushes the plant into its sensitive flowering (anthesis) phase earlier. If this accelerated growth is met with a hot May, the plant is more "vulnerable" to heat stress than a plant that developed more slowly in a cold

March. This proves that climate vulnerability is not a fixed number, but a compounding effect of the season's trajectory.

Resource Synergies: Water as a Buffer The pipeline also identified critical interactions between energy and water, particularly in Soybean and Rice.

- **Soybean ($Solar \times Water$):** The negative interaction between July Solar Radiation and Soil Water (-0.0221^{**}) suggests that the detrimental effects of extreme sun and heat are significantly worse under drought conditions.
- **Rice ($PE_{\mathbf{May}} \times SW_{\mathbf{Aug}}$):** As shown in the Rice Interaction Plot (Figure 5), the benefit provided by August soil water is "magnified" by the evaporative demand experienced in May. This suggests that the irrigation requirements of Northern Italian rice are not constant, but are dictated by the "energy history" of the early growing season.

These findings confirm that a unified pipeline, capable of testing and retaining non-linearities, is essential for characterizing the divergent and often synergistic climate sensitivities of Italian agriculture.

4 Discussion

4.1 The Utility of a Unified Pipeline for Divergent Systems

The primary objective of this project was to evaluate whether a standardized computational pipeline could effectively characterize climate vulnerabilities across a diverse portfolio of crops. Methodologically, the pipeline proved robust; the consistent reduction in AIC and improvement in Pseudo-R^2 across all five crops (Table 3) validates the iterative refinement logic. However, the results also reveal that the predictability of an agricultural system is not merely a function of model complexity, but is fundamentally governed by biological and management-driven constraints. This 'intercomparison' philosophy mirrors large-scale international efforts such as the Global Gridded Crop Model Intercomparison (GGCMI) (Elliott et al., 2015), though here it is applied at a regional scale to identify divergent sensitivities within a single geographical breadbasket.

The stark contrast between the performance of the Winter Wheat model (Pseudo-$R^2 = 0.91$) and the Spring Wheat model (Pseudo-$R^2 = 0.57$) highlights a critical "predictability gap." While both varieties were subjected to the same standardized feature selection and refinement steps, their biological alignment with the Northern Italian climate differs significantly:

1. **Phenological Synchronization:** Winter Wheat is "synchronized" with the Mediterranean seasonal cycle. By over-wintering, the crop establishes a robust root system and completes its vernalization requirement well before the onset of summer heat. Its critical reproductive phase occurs in late spring, a window that is effectively captured by ERA5-Land monthly stressors. Essentially, Winter Wheat follows a predictable "thermal script," making it an ideal candidate for regional-scale modeling.
2. **Management Noise in Spring Systems:** Conversely, Spring Wheat is often utilized as an opportunistic or "catch crop" in Northern Italy. Unlike the prestige cereal crops like Maize or Winter Wheat, which follow highly standardized planting windows, Spring Wheat is frequently sown based on field availability or following the failure of an autumn crop (Monteleone et al., 2023; Poggi et al., 2022). This introduces significant "Management Noise": a term describing yield variance caused by human decisions (sowing dates, irrigation intensity, variety choice)

rather than climatic stressors. As noted by Rezaei et al. (2018), when climate variables alone fail to capture yield variance, the discrepancy is often attributable to 'management noise,' including variety selection, fertilizer application, and localized sowing dates that are not captured in gridded reanalysis data. Because a unified pipeline lacks these management-specific inputs, its ability to "see" the climate signal is obscured by the noise of human intervention.

This divergence suggests that this unified pipeline is most valuable not as a "one-size-fits-all" solution, but as a diagnostic instrument. By applying a constant methodology to disparate crops, we can identify which agricultural systems are primarily "climate-driven" (and thus easily modeled at a 55km resolution) and which are "management-intensive" (requiring higher-resolution, variety-specific data).

The high performance of the summer crops (Maize, Soybean, and Rice) further supports this insight. These crops are grown under standardized, high-input irrigation regimes in the Po Valley. By "buffering" the crops against rainfall variability, Italian farmers effectively standardize the environment, allowing the pipeline to isolate the remaining climate signals (such as heat and solar radiation) with high precision. Consequently, the unified pipeline serves to map the boundary between predictable biological responses and complex, human-managed agricultural outcomes.

4.2 Agrological Implications: The "Wheat Paradox" and "Hydrological Buffer"

The divergent signals identified by the pipeline provide a localized map of vulnerability for Northern Italian agriculture. By synthesizing the standardized coefficients across the five crops, two dominant narratives emerge: the critical importance of spring "establishment" and the precarious nature of the region's hydrological buffer.

The observed phenological divide in March temperature sensitivity reveals a fundamental establishment bottleneck. While winter-sown varieties have already bypassed the sensitive germination phase, the negative response in spring-sown cereals suggests that early-season heat serves as a primary barrier to successful stand establishment in the Po Valley.

- Winter Wheat benefits from March warmth (as shown in Section 3.4.1) because its root system is already established; March temperatures facilitate rapid biomass accumulation and the transition to the shooting phase.
- Spring Wheat, however, suffers from March warmth. This aligns with recent case studies in Northern Italy by Poggi et al. (2022) and Monteleone et al. (2023), which suggest that as Mediterranean springs shift toward earlier thermal peaks, spring-sown cereals face a significantly higher risk of establishment failure due to topsoil moisture depletion during the germination window. At this stage, Spring Wheat is in the vulnerable germination and seedling phase. In the well-trained soils of the Po Valley, high March temperatures drive soil-surface evaporation, leading to seedling desiccation and poor stand establishment.

As climate change shifts the Italian spring toward earlier, warmer, and drier conditions, this "March Paradox" suggests a significant risk for spring-sown cereals. Without the "thermal buffer" provided by over-wintering, Spring Wheat faces an increasing risk of establishment failure, potentially forcing a regional shift toward more resilient winter-sown varieties (MedECC, 2020; Hristov et al., 2020).

The statistical preference for Soil Water over raw Precipitation indicates that Northern Italian agriculture is effectively 'decoupled' from local rainfall patterns. This highlights a profound dependence on the managed hydrological buffer provided by Alpine snowmelt and the Po River—a

buffer that is increasingly threatened by the retreat of glaciers and shifting winter precipitation regimes (Musolino et al., 2018).

The Po Valley is one of Europe's most heavily irrigated regions, relying on a complex network of canals fed by Alpine snowmelt and the Po River. The model captures this "human-managed hydrology": yield is not dictated by the rain that falls (Precipitation), but by the water that remains in the soil (Soil Water), which is supplemented by irrigation during the arid July–August window. While this irrigation currently "buffers" crops like Maize and Soybean against summer drought, it creates a systemic vulnerability. As Alpine glaciers retreat and winter snowpacks diminish, the "source" of this buffer is at risk. Our results suggest that if the irrigation network cannot maintain these Soil Water levels, the "Hydrological Dominance" of summer crops will lead to rapid, widespread yield collapses. As argued by Musolino et al. (2018) and further evidenced by the record-breaking drought of 2022 (Monteleone and Borzí, 2024), the Po Valley represents a system where irrigation acts as a primary buffer against climatic variability. Our results quantify this reliance, showing that Soil Water—a proxy for both rainfall and irrigation—is the decisive factor for summer crop resilience.

The negative response of Rice to June Solar Radiation identifies a localized thermal bottleneck specific to Mediterranean paddy systems. Beyond a certain radiative threshold, the benefits of increased photosynthesis appear to be overridden by the physical overheating of floodwater, which induces root-zone stress during the crop's early vegetative development.

Excessive radiative energy in June, when the plants are still in the early vegetative stage, can lead to the overheating of the floodwater. This phenomenon, often overlooked in global models, is a known thermal bottleneck for Mediterranean rice; Bregaglio et al. (2017) and Shi et al. (2016) identify that excessive water temperatures during the vegetative phase can scald root tissues and induce oxidative stress, significantly impairing final grain-filling. This finding suggests that "paddy overheating" may become a primary limiting factor for Italian rice production, requiring new water management strategies (such as increased water flow or intermittent flooding) to prevent the "thermal scalding" of the crop.

4.3 Methodological Critique: Strengths and Trade-offs

This project utilized a Generalized Linear Model (GLM) framework with a Gamma distribution and log link, coupled with Elastic Net regularization. While this pipeline achieved high predictive performance for the majority of the study's crops, several methodological trade-offs and "lessons learned" occured.

Ultimately, the choice of a Gamma GLM framework ensured that the coefficients remained robust across the study area's diverse productivity levels, avoiding the back-transformation biases that often plague regional yield assessments.

A key technical hurdle identified during the modeling was the numerical instability caused by the massive scale of Solar Radiation (J/m^2) relative to Temperature or Soil Water for example. The decision to selectively Z-scale only the Solar Radiation column was a "surgical" intervention designed to ensure model convergence while preserving the raw units of the remaining variables for easier sanity-checking during the training phase.

Reflecting on this approach, maintaining the stressors in their original units (e.g., °C or mm) during the refinement phase provided a valuable "sanity check" for testing non-linear and interaction terms against known biological thresholds. However, modeling with universally scaled stressors even in

the final GLM phase would have ensured maximum numerical stability and removed the need for post-hoc standardization to compare effect sizes. (Table 4). For future reference, a fully standardized pipeline from selection through to final inference would be the more robust computational choice, though the "selective scaling" used here successfully balanced the need for mathematical convergence with the requirement for agrological interpretability.

The primary limitation of the current study lies in the spatial resolution of the $55\,\text{km} \times 55\,\text{km}$ grid cells. While ERA5-Land provides high-quality regional climate signals, a 55km cell in Northern Italy can contain massive heterogeneity in soil type, microclimate, and, most importantly, human management.

This "spatial averaging" likely explains the lower Pseudo-R^2 observed in Spring Wheat. In a single 55km cell, individual farmers may plant spring wheat at different times or apply different irrigation intensities. This creates "Management Noise"—variance in the yield data that the climate-only model cannot explain. The unified pipeline successfully characterized the "average" regional signal, but it is limited by its inability to "see" the field-level decisions that dictate the success or failure of more sensitive, opportunistically managed crops.

4.4 Future Directions: Toward a GxExM Framework

While the current pipeline successfully characterized historical climate vulnerabilities, it focused primarily on the Environment (E) component of the agricultural system. To move from regional characterization toward predictive precision, future work should aim to integrate the full GxExM axis (Genotype $\times$ Environment $\times$ Management).

The inclusion of Management (M) data represents the most immediate opportunity to refine the model. As identified in the Spring Wheat analysis, "management noise" regarding specific sowing dates and irrigation volumes currently limits the model's explanatory power. Incorporating localized agricultural records would allow the pipeline to distinguish between climatic stressors and human-driven interventions. Furthermore, moving from broad species classifications to Genotype-specific (G) traits—such as the heat-sum requirements or drought-tolerance coefficients of specific hybrids—would allow for a more nuanced understanding of how different biological varieties respond to the same environmental stressors. As argued by Rezaei et al. (2018), accounting for cultivar-specific phenology and management adaptations is crucial for distinguishing between climatic shifts and technological progress

On the environmental side, the integration of Remote Sensing data offers a pathway to higher spatial resolution. Future iterations of the pipeline could incorporate satellite-derived indices such as the Normalized Difference Vegetation Index (NDVI) or Synthetic Aperture Radar (SAR) to monitor real-time biomass accumulation and phenological progress at a 1 km or field-level scale. Similar approaches have been successfully integrated into the European Union's MARS project (Rembold et al., 2013), demonstrating that satellite-derived biomass indices can effectively bridge the gap between regional reanalysis and field-level yield variability. Integrating these high-dimensional data streams would enable the model to validate "vulnerability signals" as they occur on the ground. Ultimately, evolving this framework from a historical diagnostic tool into a predictive system would allow for the simulation of crop yields under various 21st-century climate change scenarios, providing a vital computational foundation for agricultural adaptation in the Po Valley (Mereu et al., 2021; Monteleone et al., 2023).

5 Conclusion

This project has demonstrated that a unified computational pipeline, centered on penalized regression and Gamma GLMs, is a robust diagnostic tool for characterizing the divergent climate vulnerabilities of Northern Italian agriculture. By applying a constant methodology to five distinct cropping systems, the study successfully mapped the specific environmental "signals" that define the region's agricultural stability.

The findings reveal a stark contrast in crop-specific sensitivities. The "Wheat Paradox" identifies that while Winter Wheat is optimized for early spring thermal accumulation, Spring Wheat remains critically vulnerable to March heat during its establishment phase. Furthermore, the results highlight a systemic reliance on the "hydrological buffer" provided by soil water in the Po Valley; summer-sown crops like Maize and Soybean are governed more by managed soil moisture than by raw precipitation, indicating a high sensitivity to potential shifts in irrigation availability.

Methodologically, the project identified a clear "predictability gap," where the pipeline achieved high explanatory power for climate-synchronized crops like Winter Wheat ($R^2 = 0.91$) but encountered "management noise" in more opportunistically grown varieties like Spring Wheat ($R^2 = 0.57$). Ultimately, this framework provides a vital baseline for regional agro-climatic assessment. By identifying which crops are primarily climate-driven and which are limited by thermal bottlenecks or hydrological thirst, this study offers a foundation for data-driven adaptation strategies in one of Europe's most productive yet vulnerable agricultural landscapes.

References

Bregaglio, S., Hossard, L., Cappelli, G., Resmond, R., Bocchi, S., Barbier, J.-M., Ruget, F., and Delmotte, S. (2017). Identifying trends and associated uncertainties in potential rice production under climate change in mediterranean areas. *Agricultural and Forest Meteorology*, 237:219–232.

Cox, D. R. and Snell, E. J. (1989). *Analysis of Binary Data*. Chapman & Hall/CRC, 2 edition.

Drobinski, P., Rivera Ferre, M. G., Monem, M. A., et al. (2025). Nexus approach to enhance water-energy-food security and ecosystems resilience under climate change in the mediterranean. *npj Climate Action*, 4:115.

Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J., Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., et al. (2015). The global gridded crop model intercomparison: data and modeling protocols for phase 1 (v1. 0). *Geoscientific Model Development*, 8:261–277.

Fahad, S., Adnan, M., Hassan, S., Hussain, S., Wu, C., Hakeem, K., Alharby, H., and Huang, J. (2018). Rice responses and tolerance to high temperature. In *Advances in Rice Research for Abiotic Stress Tolerance*. Elsevier.

Giorgi, F. (2006). Climate change hot-spots. *Geophysical Research Letters*, 33(8).

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2 edition.

Hristov, J., Toreti, A., Pérez Domínguez, I., Dentener, F., Fellmann, T., Elleby, C., Ceglar, A., Fumagalli, D., Niemeyer, S., Cerrani, I., et al. (2020). Analysis of climate change impacts on eu agriculture by 2050. Technical Report JRC119632, Publications Office of the European Union.

Huzsvai, L., Zsembeli, J., Kovács, E., and Juhász, C. (2022). Response of winter wheat (triticum aestivum l.) yield to the increasing weather fluctuations in a continental region of four-season climate. *Agronomy*, 12(2):314.

Iizumi, T. and Sakai, T. (2020). The global dataset of historical yields for major crops 1.2 and 1.3. *Scientific Data*, 7(1):1–7.

Knapen, R., de Wit, A., Buyukkaya, E., Petrou, P., Paudel, D., Janssen, S., and Athanasiadis, I. (2025). Efficient and scalable crop growth simulations using standard big data and distributed computing technologies. *Computers and Electronics in Agriculture*, 236:110392.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall, 2 edition.

MedECC (2020). Climate and environmental change in the mediterranean basin–current situation and risks for the future. first mediterranean assessment report. Technical report, Union for the Mediterranean, Plan Bleu, UNEP/MAP, Marseille, France.

Mereu, V., Gallo, A., Trabucco, A., Carboni, G., and Spano, D. (2021). Modeling high-resolution climate change impacts on wheat and maize in italy. *Climate Risk Management*, 33:100339.

Monteleone, B. and Borzí, I. (2024). Drought in the po valley: Identification, impacts and strategies to manage the events. *Water*, 16(8):1187.

Monteleone, B., Borzí, I., Arosio, M., Cesarini, L., Bonaccorso, B., and Martina, M. (2023). Modelling the response of wheat yield to stage-specific water stress in the po plain. *Agricultural Water Management*, 287:108444.

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

Musolino, D., Vezzani, C., and Massarutto, A. (2018). Drought management in the po river basin, italy. In Iglesias, A., Assimacopoulos, D., and Van Lanen, H. A. J., editors, *Drought: Adaptation and Management Strategies*. Wiley.

Ordóñez, R. A., Archontoulis, S. V., Martinez-Feria, R., Hatfield, J. L., Wright, E. E., and Castellano, M. J. (2020). Root to shoot and carbon to nitrogen ratios of maize and soybean crops in the us midwest. *European Journal of Agronomy*, 120:126130.

Poggi, G. M., Aloisi, I., Corneti, S., Esposito, E., Naldi, M., Fiori, J., Piana, S., and Ventura, F. (2022). Climate change effects on bread wheat phenology and grain quality: A case study in the north of italy. *Frontiers in Plant Science*, 13:936991.

Rembold, F., Atzberger, C., Savin, I., and Rojas, O. (2013). Using low resolution satellite imagery for yield prediction and yield anomaly detection. *Remote Sensing*, 5(4):1704–1733.

Rezaei, E. E., Siebert, S., Hüging, H., and Ewert, F. (2018). Climate change effect on wheat phenology depends on cultivar change. *Scientific Reports*, 8(1):4891.

Ringeval, B., Müller, C., Pugh, T. A., Mueller, N. D., Ciais, P., Folberth, C., Liu, W., Debaeke, P., and Pellerin, S. (2021). Potential yield simulated by global gridded crop models: using a process-based emulator to explain their differences. *Geoscientific Model Development*, 14:1639–1656.

Sage, R. F. (2004). The evolution of c4 photosynthesis. *New Phytologist*, 161(2):341–370.

Seabold, S. and Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*, volume 57, page 61.

Shi, P., Zhu, Y., Tang, L., Chen, J., Sun, T., Cao, W., and Tian, Y. (2016). Differential effects of temperature and duration of heat stress during anthesis and grain filling stages in rice. *Environmental and Experimental Botany*, 132:28–41.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.

Çakir, R. (2004). Effect of water stress at different development stages on vegetative and reproductive growth of corn. *Field Crops Research*, 89(1):1–16.

6 Appendix 1

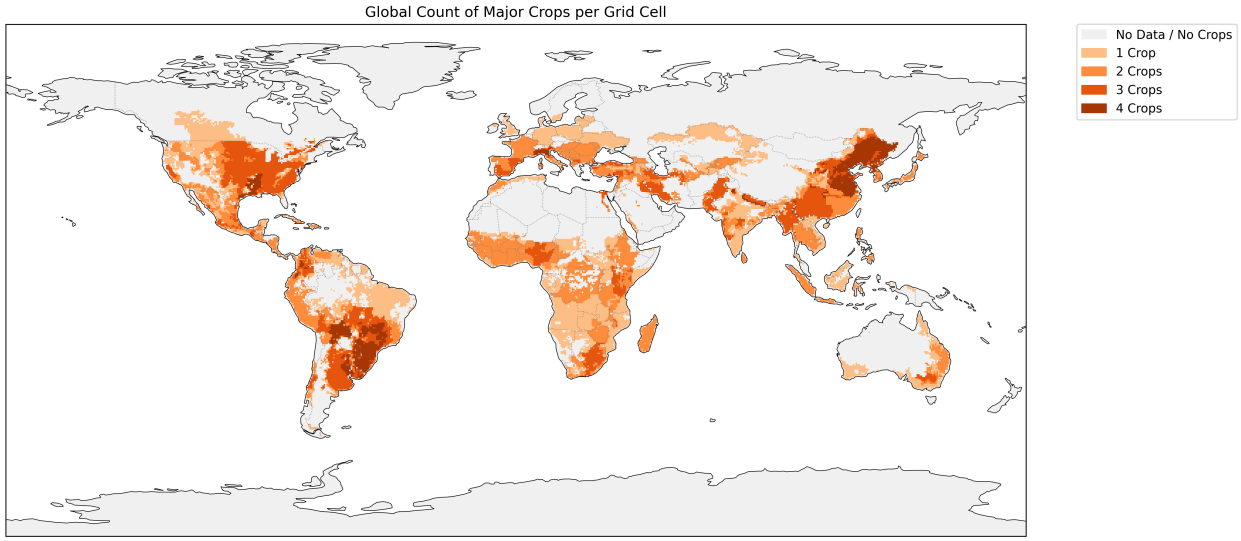


Figure 10: Global intersection of crop presence masks for the primary study crops. This spatial analysis identified Northern Italy (Po Valley) as a priority candidate region due to its rare intersection of C3 (Wheat, Rice, Soybean) and C4 (Maize) systems within a high-intensity agricultural landscape.

7 Appendix 2

Formal Model Equations

The final yield response models for each crop are defined by the following equations, where β_0 represents the intercept and ϵ represents the error term. To maintain readability, climate variables are abbreviated as: T (Temperature), P (Precipitation), SW (Soil Water), PE (Potential Evaporation), and SR (Solar Radiation).

1. **Maize:**
$$\log([\text{E}]\text{Yield}) \sim \beta_0 + \beta_1(T_{Jul}) + \beta_2(SW) + \beta_3(P_{Sep}) + \beta_4(PE_{May}) + \beta_5(T_{Jul}^2) + \epsilon \tag{2}$$

2. **Rice:**
$$\begin{aligned}\log([\text{E}]\text{Yield}) \sim \beta_0 &+ \beta_1(PE_{May}) + \beta_2(SR_{Jun}) + \beta_3(SW_{Aug}) + \beta_4(PE_{May}^2) \\ &+ \beta_5(PE_{May} \times SW_{Aug}) + \epsilon\end{aligned} \tag{3}$$

3. **Soybean:**
$$\begin{aligned}\log([\text{E}]\text{Yield}) \sim \beta_0 &+ \beta_1(SW) + \beta_2(P_{May}) + \beta_3(PE_{May}) + \beta_4(SR_{Jul}) + \beta_5(P_{May}^2) \\ &+ \beta_6(SR_{Jul}^2) + \beta_7(PE_{May} \times SR_{Jul}) + \beta_8(SR_{Jul} \times SW) + \epsilon\end{aligned} \tag{4}$$

4. **Winter Wheat:**
$$\begin{aligned}\log([\text{E}]\text{Yield}) \sim \beta_0 &+ \beta_1(T_{Mar}) + \beta_2(T_{May}) + \beta_3(T_{Nov}) + \beta_4(PE_{Jan}) + \beta_5(PE_{Apr}) \\ &+ \beta_6(T_{Mar}^2) + \beta_7(T_{May}^2) + \beta_8(T_{Mar} \times T_{May}) \\ &+ \beta_9(PE_{Jan} \times T_{May}) + \epsilon\end{aligned} \tag{5}$$

5. **Spring Wheat:**
$$\begin{aligned}\log([\text{E}]\text{Yield}) \sim \beta_0 &+ \beta_1(T_{Mar}) + \beta_2(T_{May}) + \beta_3(T_{Jul}) + \beta_4(P_{Apr}) + \beta_5(PE_{Apr}) \\ &+ \beta_6(T_{May}^2) + \beta_7(T_{Jul}^2) + \beta_8(P_{Apr}^2) + \beta_9(T_{May} \times T_{Jul}) + \epsilon\end{aligned} \tag{6}$$