Agricultural Resilience in the Po Valley: A Data-Driven Analysis of Crop Vulnerability to Climate Volatility

Esther Helga and Anders Lydig Kristensen

December 2025

1 Introduction

1.1 Aims and Objectives

The primary goal of this project is to evaluate the capacity of a unified computational pipeline to characterize and compare the climate vulnerabilities of five major crops in Northern Italy. While the Po Valley serves as a critical agricultural hub, climate change is introducing unprecedented thermal and hydrological volatility. Traditional modeling approaches often focus on isolated crop species; this study seeks to determine if a standardized, data-driven framework can successfully identify the divergent environmental "signals" that drive yield across multiple biological systems.

Research Questions To achieve this goal, 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: Harmonize high-resolution gridded yield data ? with climate reanalysis data (ERA5-Land) into a spatially-explicit modeling framework for Northern Italy.

- 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 explanatory power (Pseudo-R^2) across crops to identify the boundaries between climate-driven biological responses and human-managed agricultural variance.

2 Methods

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)** (?). 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**. We utilized the **monthly averaged reanalysis product** at a resolution of $0.1° \times 0.1°$ (9 km) to match the seasonal scale of agricultural production. We retrieved five variables representing water supply/demand and energy inputs: **2m Temperature** (T_{2m}), **Total Precipitation** (TP), **Volumetric Soil Water Layer 1** ($SWVL1$), **Surface Net Solar Radiation** (SSR), and **Potential Evaporation** (PEV).

2.2 Study Area Identification and Scouting

2.2.1 The Scouting Process

A rigorous scouting process was implemented to identify a focus region with high agricultural intensity and reliable data. We first generated global presence/absence masks for the four primary crops to identify spatial intersections supporting multi-crop rotations. This analysis highlighted four candidate regions for further evaluation: the North China Plain, Northern Italy (Po Valley), coastal Peru, and central Chile.

2.2.2 Selection of Northern Italy

The Po Valley in Northern Italy (approx. 43°N–47°N, 6°E–13°E) was selected as the final study area based on the following 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.
- **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_{2m}, $SWVL1$) were averaged over the season, while **Flux Variables** (TP, SSR, PEV) 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**. Standard Ordinary Least Squares (OLS) regression assumes Gaussian errors and allows for negative outcomes. However, crop yield data is strictly non-negative and frequently exhibits non-normal characteristics.

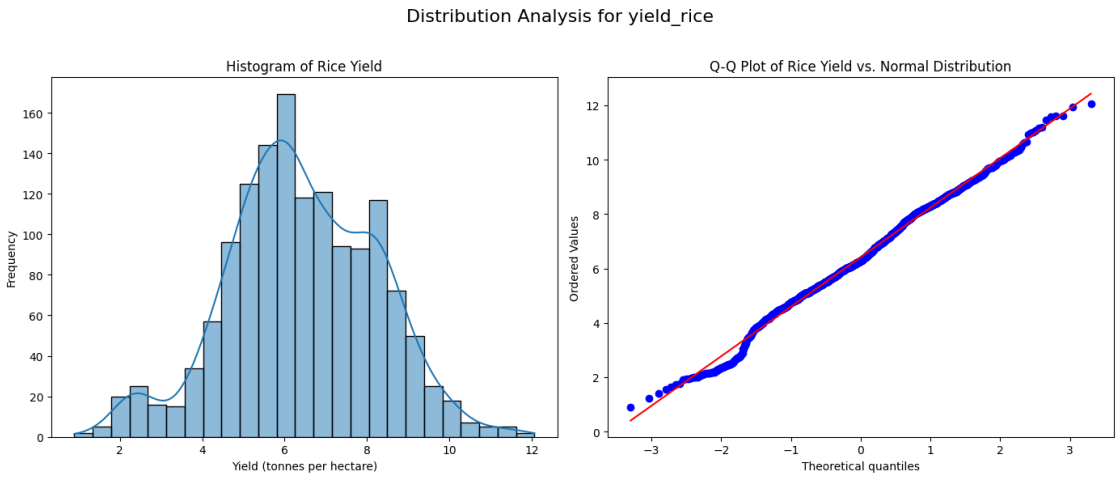
Our distributional assessment via histograms and Q-Q plots (see Fig. 1 for representative examples of Rice and Winter Wheat) demonstrated that for all five systems, the yield data departed significantly from normality. The histograms revealed clear skewness and distinct modes, while the Q-Q plots showed significant upward curvature in the tails, indicating that the residuals of a standard OLS model would be biased. This evidenced non-normality justifies the use of the Gamma distribution, which is better suited for strictly positive data where the variance typically increases with the mean.

Furthermore, the log link function $\ln(\mu)$ implies that climate impacts are multiplicative (proportional yield reductions) rather than additive, which is an intuitive and widely accepted assumption in agricultural vulnerability research.

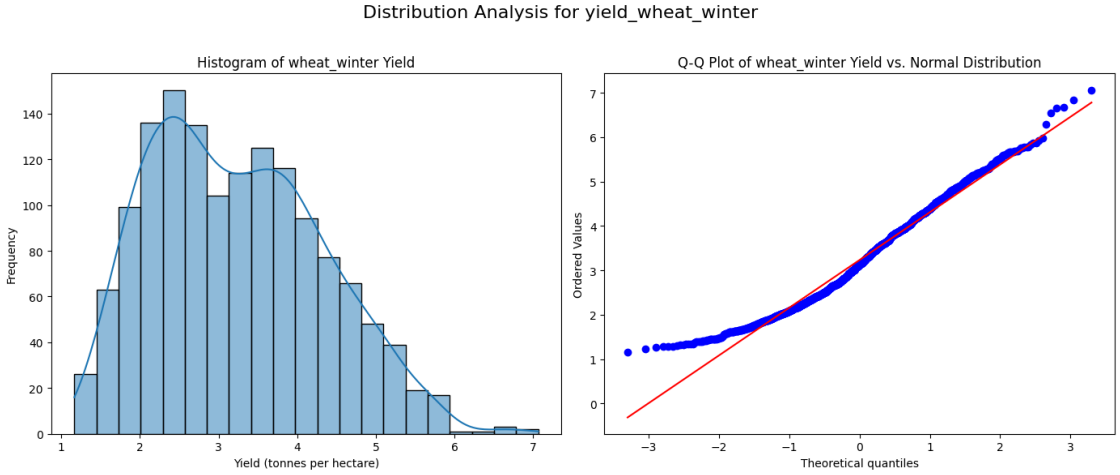
2.5.2 Spatial Modeling and Technological Trends

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*.



(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**. All predictors were Z-score standardized for this phase to ensure a fair penalty application. 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 Numerical Stability and Variable Standardization

The development of the Soybean model revealed a technical hurdle: the magnitude of Surface Net Solar Radiation (SSR) in Joules per square meter (J/m^2) is astronomically larger than temperature or soil water. During non-linear testing, squared SSR terms (SSR^2) caused extreme values in the model's Hessian matrix, leading to total convergence failure. To resolve this, we implemented targeted **Z-score standardization** for solar radiation variables in affected models to stabilize the mathematical optimization while preserving the underlying biological signal.

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 $\times$ Water) and **Temporal Compounding** (Early-season $\times$ 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.

All analyses were conducted in Python (`v3.10.18`) using `statsmodels`, `scikit-learn`, and `patsy`.

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 Paths (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.

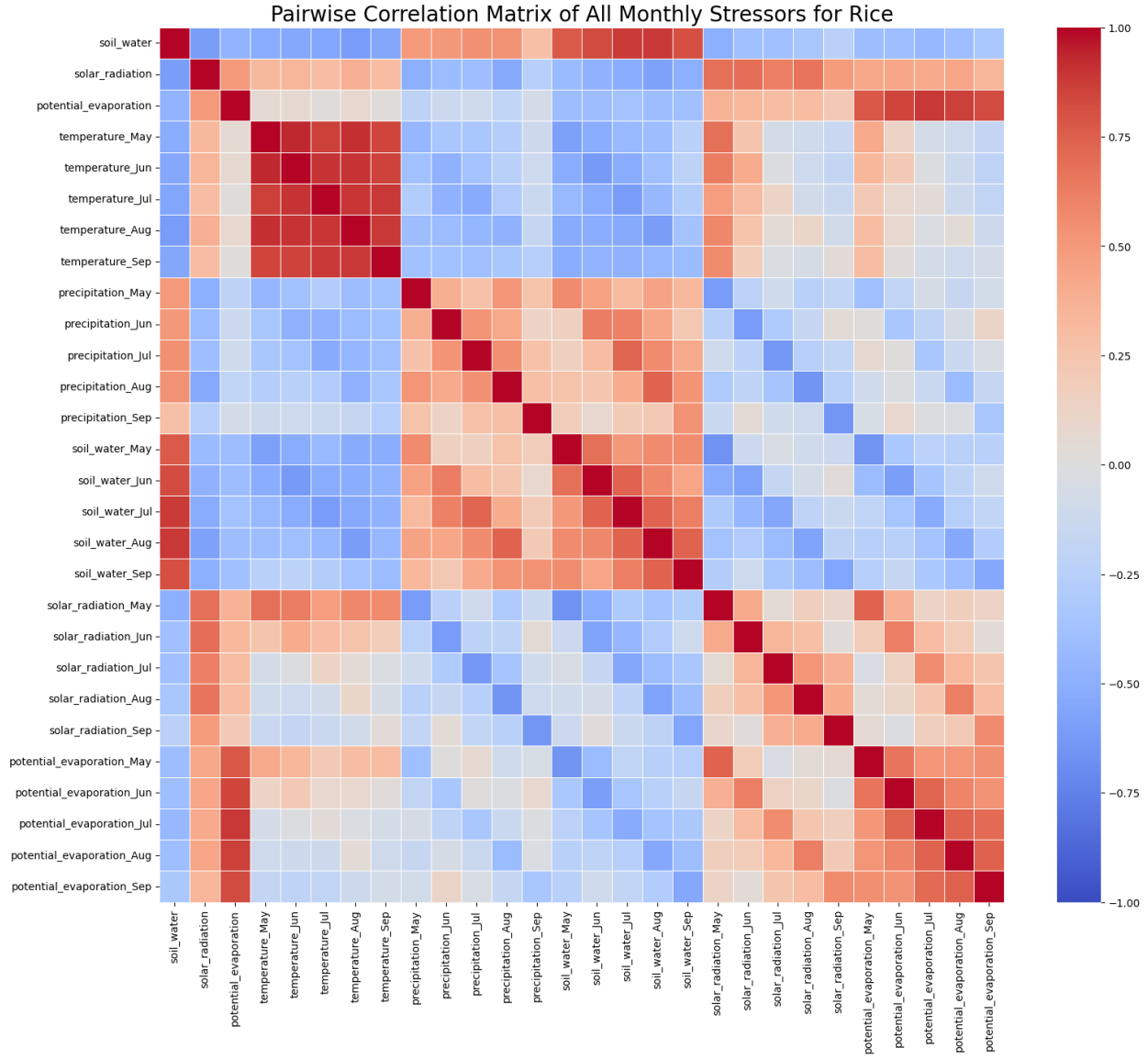
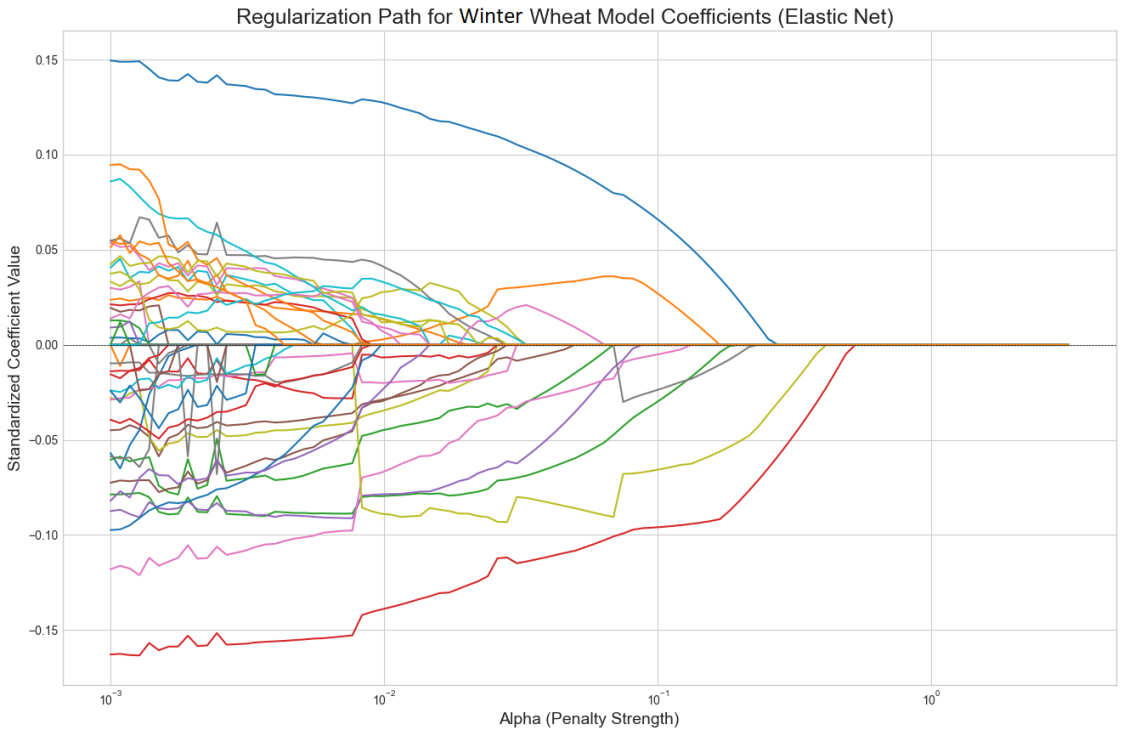
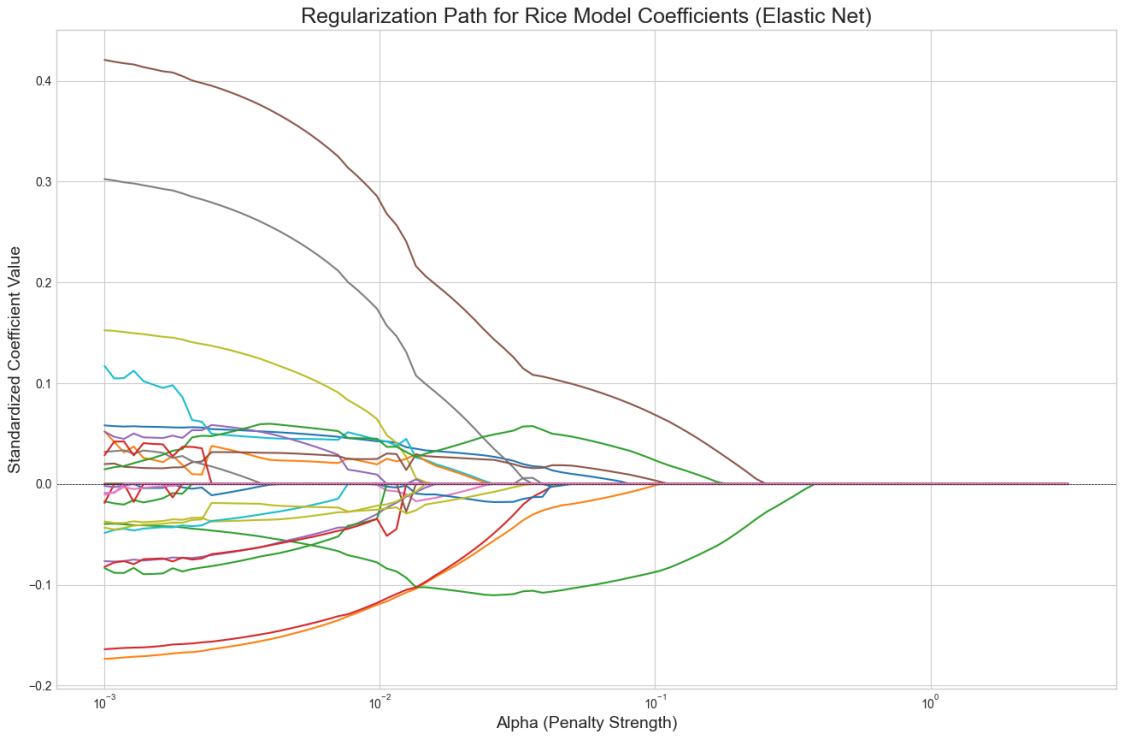


Figure 2: Pairwise Correlation Matrix of all monthly stressors for Rice. The heatmap illustrates strong intra-seasonal and inter-variable correlations among climate variables, such as Temperature, Potential Evaporation, and Soil Water, highlighting the high degree of multicollinearity in the initial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



(a) Regularization path for the winter season model.



(b) Regularization path for the rice model.

Figure 3: XXX

3.1.3 Technical Refinements and Pre-processing

During the model refinement phase, a technical hurdle was identified regarding the numerical scale of the predictors. In the Soybean model, the raw values for Solar Radiation (J/m^2) were of a significantly larger magnitude than other climate stressors, which led to convergence failures when non-linear (quadratic) terms were introduced. To resolve this, the Solar Radiation variable was Z-score rescaled prior to model fitting for the Soybean model to ensure numerical stability. To allow for a direct comparison of relative 'signals' across all 5 crops and disparate stressors in the final analysis, all resulting model coefficients were subsequently Z-score standardized. This dual approach ensured both the mathematical validity of the individual models and the interpretability of the comparative results.

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 the Non-Linear Yield Response Plot (Figure 4), the model identifies a convex (upward-curving) relationship where yield potential accelerates at higher evaporation levels.

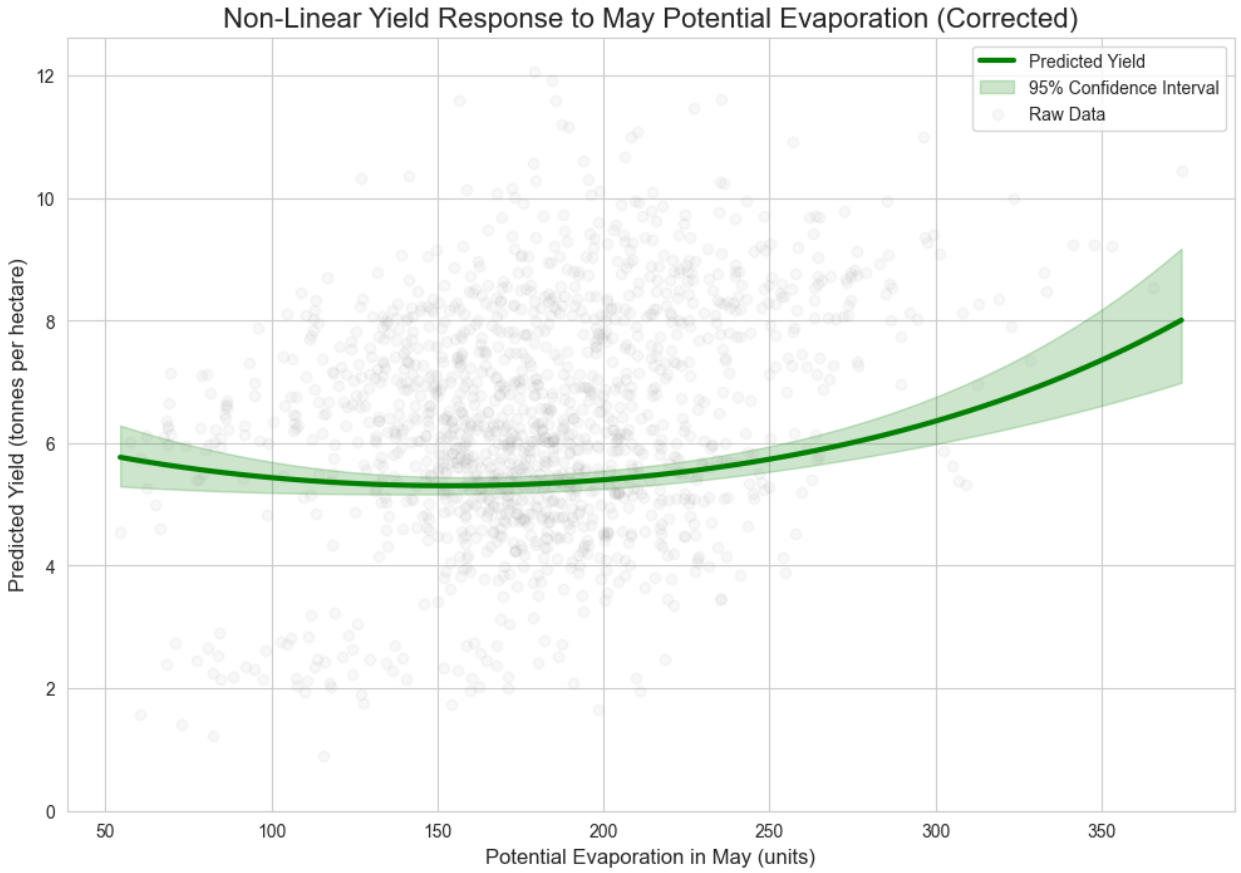


Figure 4: XXX

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.

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. The interaction

term lowered the AIC by an additional 9.20 units, capturing the synergistic effect of energy and water balance that a purely additive model would overlook.

As shown in the Interaction Effect Plot (Figure 5), the yield benefit derived from August soil water is highly 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 significantly 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.

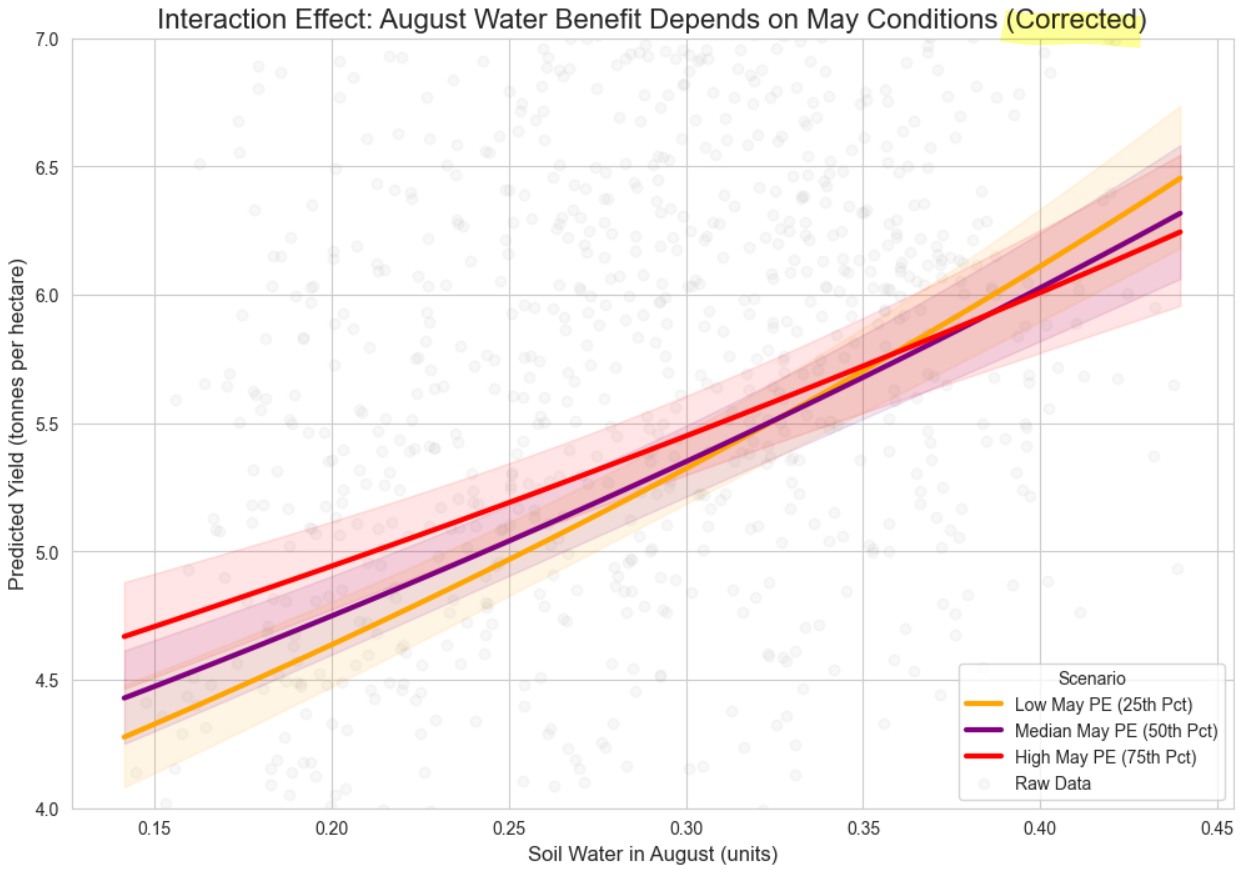


Figure 5: XXX

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 1) 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. 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.

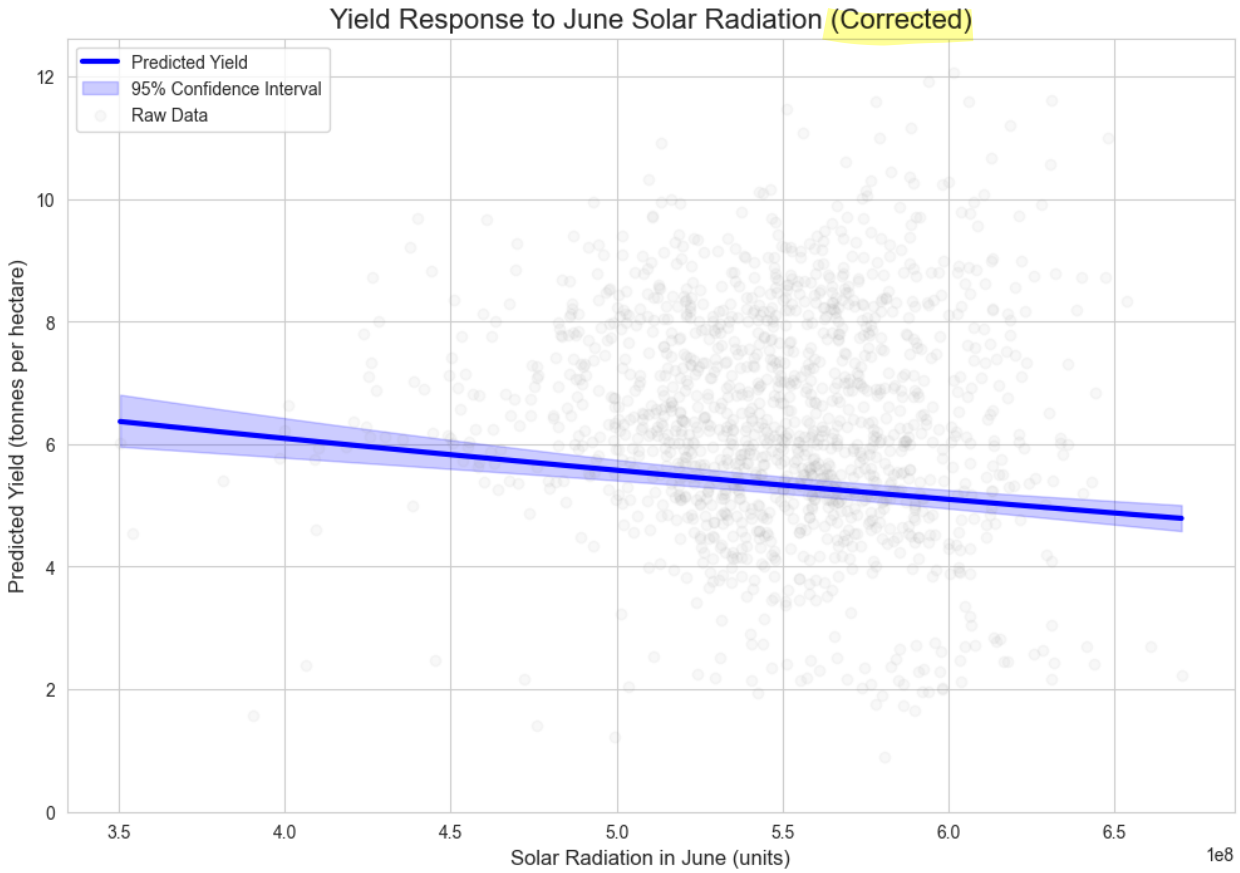


Figure 6: XXX

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 XXX.

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 the rapid thermal transitions occurring between March and July. While summer crops (Maize, Soybean, and Rice) are predominantly grown under standardized irrigation regimes in the Po Valley, which buffers them against regional climate fluctuations, Spring Wheat is often a rain-fed or secondary crop (frintier paper). Its yield is highly dependent on the precise timing of spring heat spikes during the narrow flowering window, a 'micro-timing' factor that may be obscured by the 55km spatial resolution of the gridded yield 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 is rooted in the distinct developmental stages of the two crops during this window:

- Winter Wheat (The Head Start): Sown in autumn, winter wheat enters March as an established plant. Warmth in March facilitates rapid spring tillering and biomass accumulation, allowing the plant to "take off" early. As long as the vernalization requirement (cold-exposure) was met during the winter months, a warm March acts as a catalyst for yield.
- Spring Wheat (The Vulnerable Start): Sown in March, this crop is in its most fragile state during this window. High March temperatures increase soil evaporation and can cause "seedling thirst," harming germination and early root establishment in the gridded 55km cells of the Po Valley.

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.

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.

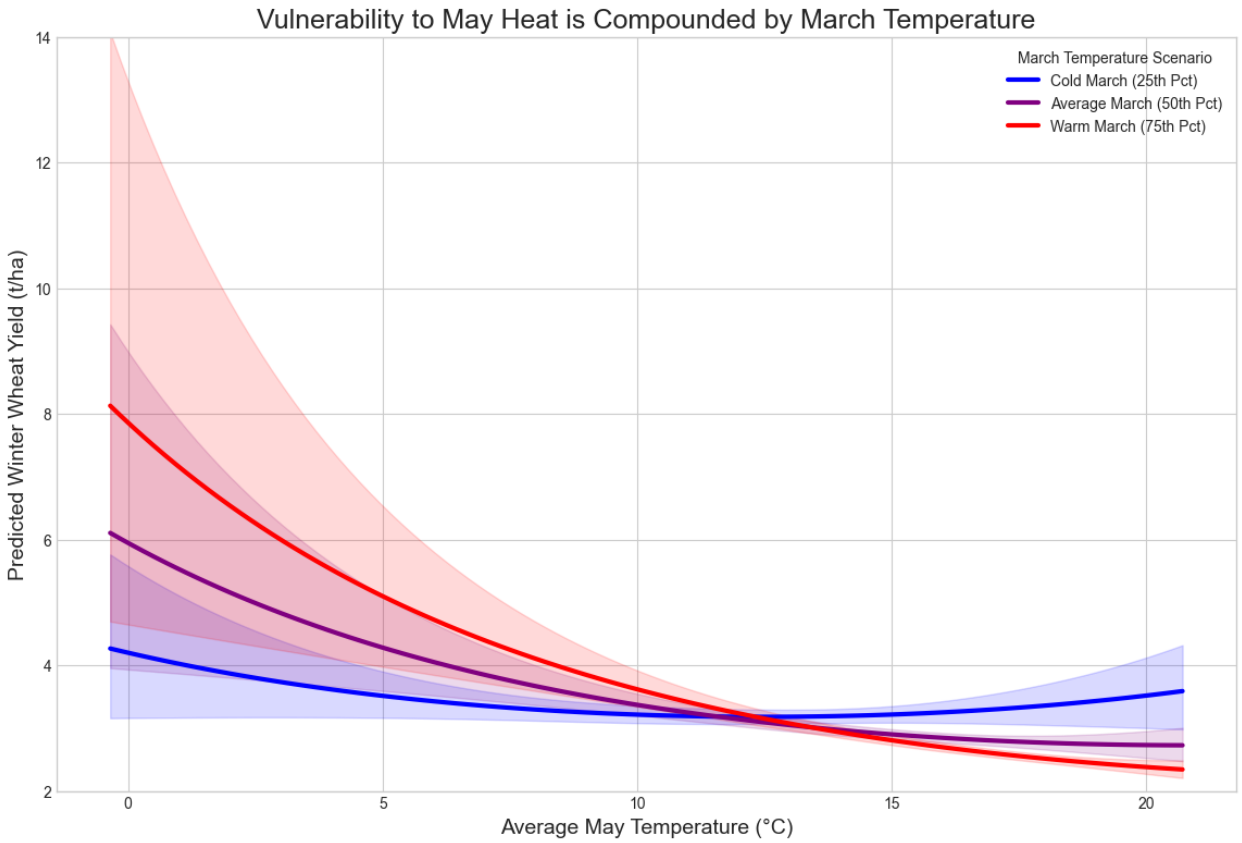


Figure 7: XXX

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. 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.

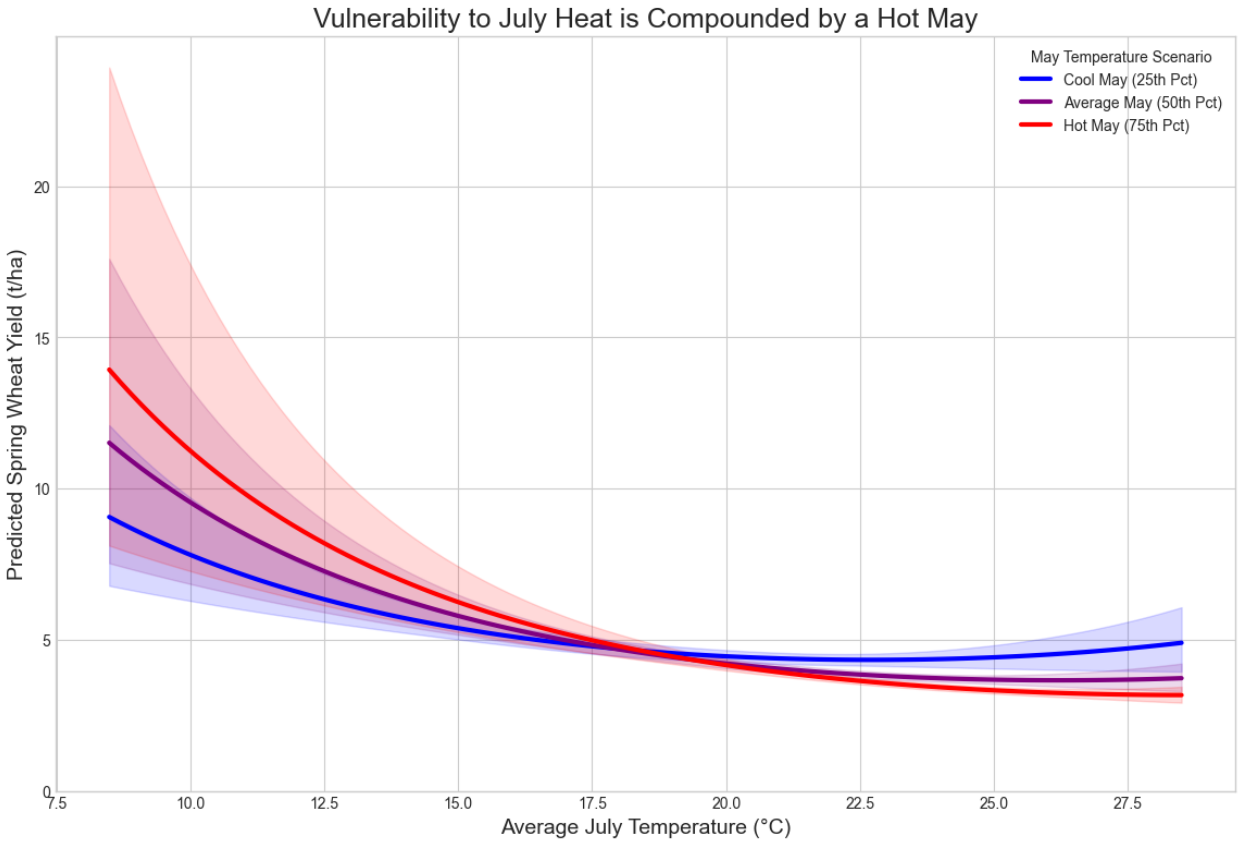


Figure 8: XXX

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.

In Northern Italy, summer precipitation is often erratic, but the region is characterized by an extensive irrigation network fed by the Po River and Alpine runoff. 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 visualized in the Yield Response Plot for Soybean (Figure 9), there is a clear, positive linear relationship: as soil moisture increases, predicted yield rises significantly, confirming that these systems are essentially "water-limited" in their current climatic context.

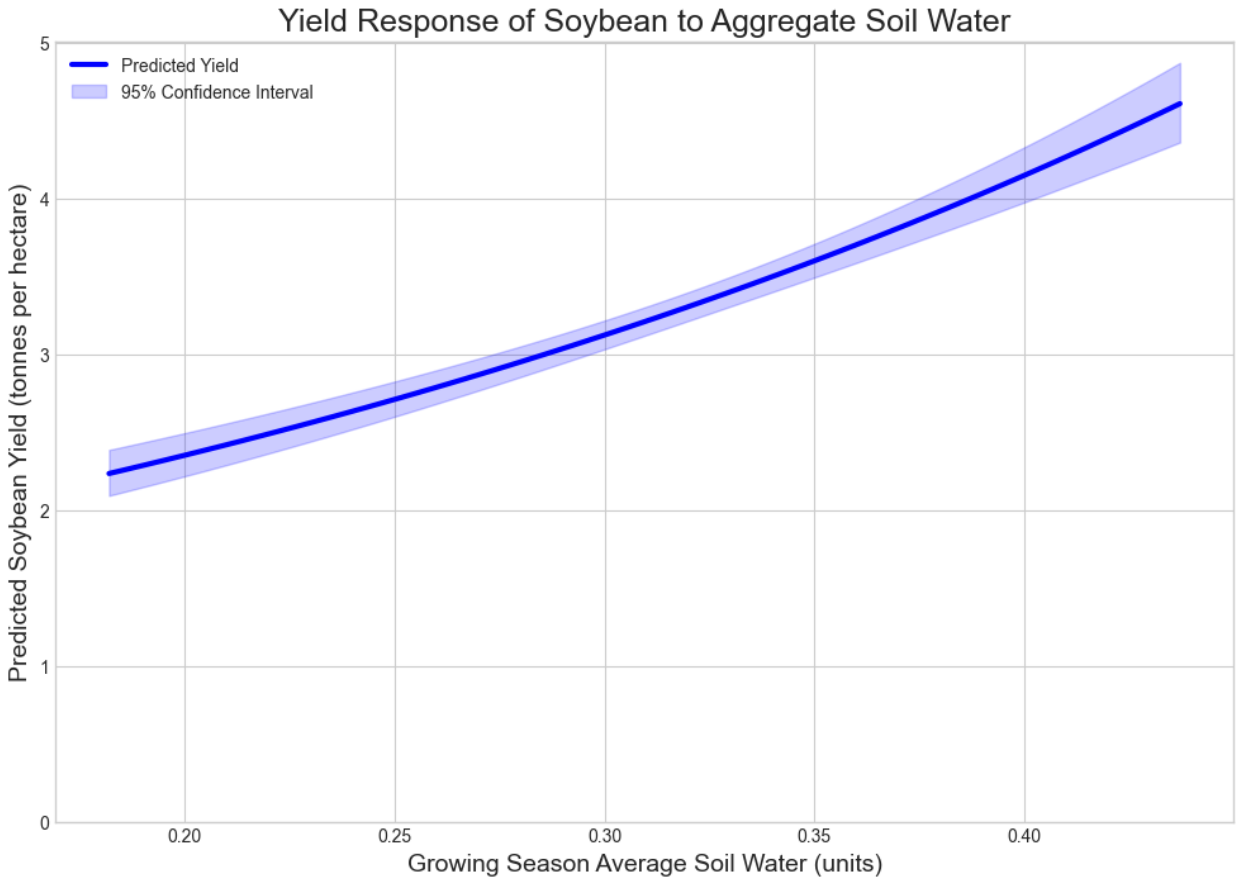


Figure 9: XXX

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 strong positive response to Precipitation in May (0.0706***). This suggests that early-season moisture is critical for soybean establishment, likely to ensure a deep root system 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.

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. 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.

In Northern Italy, June is the critical period for rooting and early tillering. During this phase, rice paddies are typically submerged in a shallow layer of standing water (approximately 5–10 cm). High levels of solar radiation in June often coincide with intense heatwaves and high atmospheric pressure. In these shallow, stagnant water systems, excessive radiation can lead to the rapid overheating of the floodwater, which can exceed the biological threshold for healthy root development and nutrient uptake. Unlike the beneficial warmth of May (PE_{May}) which helps establish the crop, the extreme radiative energy of June represents a thermal bottleneck that can "scald" the sensitive young plants before they reach the more resilient reproductive stages.

A Proxy for Hydrological Stress Additionally, June Solar Radiation may serve as a proxy for regional water scarcity. High radiation levels in June are historically correlated with periods of low precipitation and reduced Alpine snowmelt, the primary sources for the Po River and the Cavour Canal irrigation network. The negative yield response likely captures the "double-stress" of high radiative demand coupled with a potential reduction in the discharge volumes available to maintain the cooling flow of water through the paddies.

By isolating June as a month-specific stressor, the pipeline successfully characterized a vulnerability that is often masked when using growing-season averages.

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.

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 ($Temp^2_{\text{Jul}}$): 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 ($Temp_{\text{Mar}} \times Temp_{\text{May}} = -0.1641^{***}$).

- The Interpretation: In a "Cold March" scenario (Blue Line), the yield remains relatively stable as May temperatures rise. However, in a "Warm March" scenario (Red Line), the yield is significantly higher at first but then declines much more sharply as May heat increases.
- Biological Logic: 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 ($PE_{\text{May}} \times SW_{\text{Aug}}$): 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, apable 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.

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. This introduces significant "Management Noise": a term describing yield variance caused by human decisions (sowing dates, irrigation intensity, variety choice) rather than climatic stressors. 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 a 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 contradiction in March temperature sensitivity (Table 4) represents a fundamental "establishment bottleneck" in Italian cereal systems.

- Winter Wheat benefits from March warmth ($+0.3051^{***}$) 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 (-0.0731^{***}). 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.

A defining feature of the summer crop models (Maize and Soybean) is the overwhelming significance of Soil Water compared to raw Precipitation. In a natural, rain-fed system, these two variables would be tightly coupled; however, the pipeline's preference for the Soil Water signal reflects the managed nature of the Northern Italian landscape.

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.

The negative response of Rice to June Solar Radiation (-0.0209^{***}) identifies a specific thermal bottleneck unique to flooded paddy systems. While rice is a heat-loving crop, the shallow water environment of Italian paddies can become a liability during June heatwaves.

Excessive radiative energy in June, when the plants are still in the early vegetative stage, can lead to the overheating of the floodwater. High water temperatures at the soil-water interface can cause root-zone heat stress and facilitate the release of toxic soil metabolites, which prioritize plant survival over growth. 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.

The decision to use a Gamma GLM over a standard Ordinary Least Squares (OLS) regression on log-transformed data was a deliberate choice for statistical and interpretability reasons. As identified in the EDA phase, crop yields are strictly positive and typically right-skewed. While a log-transformation can "normalize" such data, it changes the fundamental nature of the model: OLS on log-data models the geometric mean, whereas a Gamma GLM with a log link models the arithmetic mean on the original scale.

By avoiding transformation bias, the Gamma GLM provides more accurate predictions of expected yield. Furthermore, this approach naturally handles heteroscedasticity, where the variance in yield grows with the mean, making the model's coefficients more robust across different regional productivity levels.

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.

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. 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.

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.

6 Appendix 1

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

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

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

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