

Study on Disaster Characteristics of Huaxi Autumn Rain and Trigger Mechanisms for Weather Index Insurance

Zichuan Sun

Mianyang Meteorological Bureau

Wulong Zhang

Sichuan Meteorological Observatory, Chengdu 610072, Sichuan

Gao Li

Mianyang Meteorological Bureau

Kaichuan Xie

Mianyang Meteorological Bureau

Weimin Xu

Mianyang Meteorological Bureau

Liangxun Chen

Mianyang Meteorological Bureau

Fengliang Zou

1359793413@qq.com

Mianyang Academy of Agricultural Sciences

Jun Yang

Pingwu Meteorological Bureau

Article

Keywords: Huaxi Autumn Rain, Weather Index Insurance, Compound Climate Hazards, Basis risk

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11169655/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

To investigate the issue of uncontrolled basis risk caused by heterogeneous disaster mechanisms in weather index insurance, this study focuses on the high-impact zone of Huaxi autumn rains(HXAR). By integrating daily meteorological observations and county-level yield data from 2014 to 2024, we dynamically define the autumn rain period for each station, ensuring precise alignment between meteorological exposure windows and crop-sensitive growth stages. We then extract ten key meteorological indicators during overlapping periods and employ Spearman rank correlation and random forest models to identify both monotonic relationships and nonlinear interactive effects between weather and yields. Furthermore, we propose a semantics-based disaster classification system grounded in physical rules, decomposing heterogeneous yield-loss events into subgroups with homogeneous mechanisms, enabling independent identification of trigger thresholds within each subgroup. To systematically evaluate performance, multiple global benchmarks are established, including historical loss methods, quantile approaches, cost-benefit optimization, composite weather indices, regionally unified thresholds, and LOSO (Leave-One-Out Station) adaptive strategies. Results indicate that traditional global trigger mechanisms exhibited systematic generalization failure on the test set. In contrast, the semantic conditional trigger mechanism achieved a fundamental breakthrough in the maize "Prolonged Oligosolar" subset, with the test-set F1 score increasing from 0% to 66.67%. For rice, secondary stratification optimization effectively disentangled confounded disaster-causing pathways by reducing the trigger probability from 85.7% to 28.6%; however, constrained by small-sample limitations, the test-set F1 score dropped to 0%, revealing the practical boundaries of parametric modeling under complex mechanisms and the inherent small-sample ceiling effect.

1 Introduction

Huaxi Autumn Rain is a distinctive seasonal precipitation phenomenon unique to Southwest China (Ding and Wang, 2008). It typically occurs from late August to early November (Gao and Guo, 1958; China Meteorological Administration, 2019), characterized by extensive spatial coverage, long duration, frequent rainy days, weak intensity per event, yet significant cumulative impacts—particularly through the co-occurrence of persistent overcast conditions, high humidity, and mild low temperatures (Luo et al., 2013). Under climate change, recent decades have seen HXAR trends toward prolonged duration and increased frequency of extreme rainfall events (Li, 2024; Li et al., 2025; Tang et al., 2025). These changes pose severe threats to grain production in core regions such as the Sichuan Basin, critically disrupting rice grain filling and maturation, maize kernel dehydration, and post-harvest drying operations. Consequently, HXAR has emerged as a key climatic risk factor constraining stable and efficient agricultural output.

Current agricultural disaster response mechanisms remain heavily reliant on post-event loss assessment and conventional indemnity-based insurance, which suffer from subjective damage evaluation, high inspection costs, moral hazard, and delayed payouts (Song, 2019). These limitations

hinder the ability to meet smallholder farmers' urgent demand for "fast, transparent, and predictable" risk-mitigation tools.

HXAR exhibits pronounced persistence and cumulative damage characteristics, representing a low-frequency, high-impact risk—making it unsuitable for short-term hedging instruments like weather derivatives. In contrast, Weather Index Insurance (WII), which triggers payouts based on objective, verifiable meteorological data, offers advantages of low cost, high efficiency, and reduced moral hazard. It is widely regarded as a vital financial instrument for enhancing agricultural climate resilience (Miranda and Farrin, 2012; Abdi et al., 2025). Existing WII research has largely focused on staple crops like rice and wheat under single-hazard scenarios (e.g., heat stress or drought), often using fixed calendar-based risk windows (Ren et al., 2025; Chen et al., 2024). However, these approaches inadequately address the compound nature of HXAR—specifically its "multi-factor synergy" and "dynamic phenological misalignment"—and lack targeted modeling of humid-region-specific loss mechanisms such as rice mold during harvest or impaired maize drying.

Moreover, WII implementation in HXAR-prone areas faces three core challenges:(1) Phenological mismatch: Fixed risk windows (e.g., September–October) ignore interannual variability and spatial heterogeneity in HXAR onset/cessation, exacerbating basis risk (Nieto et al., 2010; Clement et al., 2018); (2) Oversimplified thresholds: Trigger criteria often rely on single variables (e.g., total rainfall) or static percentiles, failing to capture the nonlinear interactions among "prolonged rain + low sunshine + cool temperatures" (Zulkaflī et al., 2021);(3) Lack of typology: Absence of semantically meaningful disaster classifications impedes scenario-based product design and precision early warning.

To address these gaps, this study focuses on Chengdu and Mianyang in the Sichuan Basin—the HXAR high-impact core—and integrates daily meteorological observations (2014–2024) with county-level yield data to develop a novel integrated framework: Dynamic autumn rainfall period identification—multi-method disaster factor screening—semantic disaster type stratification—conditional trigger threshold optimization—multi-benchmark comparative diagnosis. Our key innovations include:(1) Dynamically defining site-specific HXAR periods using the national standard (QX/T 496–2019) to align meteorological exposure with crop-sensitive stages;(2) Combining Spearman correlation, Random Forest (RF), and XGBoost to uncover nonlinear and high-order interactive effects in climate–yield relationships;(3) Integrating Spearman correlation analysis with a leave-one-year-out cross-validated random forest model to reveal the nonlinear and higher-order interaction effects in meteorological–yield relationships;(4) Multidimensional global benchmark comparison and EPV adaptive attribution framework..

This work provides a methodological reference for the design of the condition-triggering mechanism of the HXAR Weather Index Insurance, and also offer an adaptive modeling approach for the attribution analysis of agricultural meteorological disasters under small sample conditions.

2. Results

2.1. Spatiotemporal Characteristics and Meteorological Profiles of the Huaxi Autumn Rain

Figure 1 presents a heatmap of HXAR onset and cessation dates across stations from 2014 to 2024. The HXAR period averaged 62 ± 23 days, with 85% of stations initiating in mid-September. Onset typically occurred between late August and mid-September, while cessation ranged from mid-to-late October through late November. Substantial interannual variability and local heterogeneity were evident: for instance, the 2018 and 2024 seasons were markedly shorter and more concentrated, whereas 2022 exhibited an unusually prolonged duration.

The final analytical dataset comprised 90 site–year observations for rice (including 18 loss years, accounting for 20.0%) and 110 site–year observations for maize (including 16 loss years, accounting for 14.5%). Following a uniform temporal split, the period 2014–2020 was designated as the training set ($n = 63$ for rice; $n = 77$ for maize), while 2021–2024 served as an independent test set ($n = 27$ for rice; $n = 33$ for maize). All subsequent analyses strictly adhered to this data partitioning scheme.

2.2. Yield Loss Identification and Meteorological–Yield Associations

Table 2
reports Spearman rank correlations between meteorological variables during the overlap
of HXAR and crop-sensitive periods and yield loss

Crop	Variable	ρ	p-value	Significance
Rice	Total precipitation	-0.1904	0.0723	
	Number of rainy days	-0.102	0.3387	
	Total sunshine hours	-0.2543	0.0156	*
	Average temperature	-0.1367	0.1989	
	Minimum temperature	-0.1159	0.2768	
	Average relative humidity	0.1424	0.1805	
	Effective precipitation contribution ratio	-0.2438	0.0206	*
	Extreme heat accumulation	-0.0577	0.5892	
	Average diurnal temperature range	-0.1824	0.0853	
	Maximum consecutive dry days	0.1325	0.2133	
Maize	Total precipitation	-0.2549	0.0072	**
	Number of rainy days	-0.2288	0.0162	*
	Total sunshine hours	0.4659	0	***
	Average temperature	0.4865	0	***
	Minimum temperature	0.1283	0.1817	
	Average relative humidity	-0.5883	0	***
	Effective precipitation contribution ratio	-0.063	0.5134	
	Extreme heat accumulation	0.5429	< 0.001	***
	Average diurnal temperature range	0.533	< 0.001	***
	Maximum consecutive dry days	0.4438	< 0.001	***

This study reveals that the meteorological drivers of yield loss in rice and maize during the HXAR exhibit diametrically opposed directional effects, highlighting fundamental differences in crop-specific response patterns under a shared climatic background.

For rice, yield loss is primarily driven by the compound stress of insufficient solar radiation and high humidity. As shown in Table 2, total sunshine hours and the effective precipitation contribution rate during the intersection window are significantly negatively correlated with yield loss, identifying solar radiation deficit as the most critical limiting factor during the rice grain-filling stage. Conversely, average

relative humidity exhibits a positive correlation trend, suggesting that a highly humid environment may further exacerbate yield loss risks by inducing pathogenic diseases or inhibiting transpiration. Notably, although minimum temperature shows a negative correlation trend, it fails to reach statistical significance, contrasting with the traditionally dominant role of autumn chilling injury(Wang et al., 2010). This discrepancy likely arises because, during the overlap between the crop-sensitive period and the HXAR in the study area, temperatures do not generally breach the threshold for sterile-type cold damage; instead, the photosynthate deficit caused by persistent cloudiness emerges as a more direct bottleneck for yield formation. This finding corroborates the concept of "compound disasters induced by low solar radiation and continuous rain(Zhang & He, 1991)," emphasizing that in humid hilly regions, the marginal impact of light deficiency on rice yield is more pronounced than that of low temperature alone.

In stark contrast, yield loss in maize is predominantly driven by the compound stress of high temperature, intense solar radiation, and atmospheric dryness.Total sunshine hours, average temperature, extreme heat accumulation, and average diurnal temperature range are all highly significantly and positively correlated with maize yield loss, whereas average relative humidity shows a highly significant negative correlation. This indicates that within the impact window of the HXAR, maize is not subjected to the typical "autumn rain and low solar radiation" stress, but rather to anomalous hot, sunny, and dry conditions that trigger heat-induced forced maturity and water deficit. The underlying biological mechanism is that abnormal high temperatures during the grain-filling stage accelerate the grain dehydration rate and shorten the effective duration of grain filling, thereby reducing the 100-grain weight(Troy et al., 2015). Concurrently, low humidity exacerbates water stress in the soil-plant system; even if total precipitation is not significantly deficient, excessive atmospheric evaporative demand can still induce physiological drought. The strong positive correlation with maximum consecutive dry days further substantiates the critical role of cumulative drought effects in driving maize yield loss.

2.3. Assessment of Disaster-Inducing Factor Importance and Model Generalizability Based on Machine Learning

As shown in Table 3, the two crops exhibit significantly divergent response patterns to meteorological factors. For maize, the model demonstrates excellent discriminative ability on the global dataset (Global AUC = 0.8938), with a Leave-One-Out Cross-Validation (LOOCV) AUC of 0.6667. This indicates that the nonlinear model constructed using meteorological features from the autumn rain period can effectively capture the core driving mechanisms of maize yield loss and possesses considerable out-of-sample generalization potential. In contrast, the predictive performance of the rice model is markedly lower (Global AUC = 0.4487; LOOCV AUC = 0.4429), with an F1-score of only 0.3207. These results suggest that the currently selected set of meteorological features fails to adequately explain the variance in rice yield loss events. This implies that rice yield losses may be predominantly driven by non-meteorological factors (e.g., varietal resistance, field management, or pest and disease outbreaks), or that the disaster-inducing mechanisms exhibit high local heterogeneity and stochasticity, rendering regional-scale meteorological indicators insufficient for generating robust predictive signals.

Table 3

Importance of meteorological drivers of crop yield loss and model performance evaluation used on Random Forest and LOOCV cross-validation.

Crop	Variable	Importance	Selected Folds	LOOCV AUC (Mean $\pm$ Std)	Global AUC	Mean F1
Rice(n = 90)	Total sunshine hours	0.258	11	0.3973 $\pm$ 0.2536	0.5082	0.27
	Total precipitation	0.2489	7			
	Average temperature	0.2231	2			
	Average diurnal temperature range	0.2065	6			
	Average relative humidity	0.1988	11			
	Minimum temperature	0.1689	1			
	Effective precipitation contribution ratio	0.165	11			
	Extreme heat accumulation	0.141	1			
	Maximum consecutive dry days	0.1408	3			
Maize(n = 110)	Average diurnal temperature range	0.3465	9	0.4219 $\pm$ 0.1406	0.8838	0.2857
	Total sunshine hours	0.2482	9			
	Average relative humidity	0.2217	9			
	Average temperature	0.1629	7			
	Maximum consecutive dry days	0.0565	3			
	Extreme heat accumulation	0.0428	8			

Although the overall generalizability of the rice model is limited, the feature importance ranking derived from the training set (Table 3) still reveals the model's dependency preferences during the fitting

process. For rice, total precipitation (importance = 0.2735) and average diurnal temperature range (0.2487) rank as the top two factors, while average relative humidity (0.2455) was selected in all 11 folds of cross-validation (Selected_Folds = 11), demonstrating exceptionally high stability. This partially aligns with the "low solar radiation–high humidity" driving mechanism identified in the Spearman analysis. However, it is noteworthy that total sunshine hours, which was significantly correlated in the Spearman analysis, did not rank among the top important features. This suggests that in the nonlinear tree-based model, the variance explained by solar radiation may have been absorbed by precipitation and temperature range through higher-order interactions, or that the nonlinear responses of rice to moisture and temperature fluctuations may have masked the linear marginal effect of solar radiation.

For maize, the distribution of feature importance exhibits high concentration and stability. Average diurnal temperature range (0.3793), average relative humidity (0.2799), and average temperature (0.1814) rank as the top three factors, and all three were consistently selected across all nine folds of cross-validation (Selected_Folds = 9). This result is highly consistent with the Spearman correlation analysis, confirming that maize yield loss is driven by the compound stress of "high temperature–dryness–large diurnal temperature range." In particular, the paramount importance of average diurnal temperature range in the nonlinear model further corroborates that anomalous temperature fluctuations during the grain-filling stage, synergizing with water stress, accelerate grain dehydration and forced maturity in maize. Furthermore, the stable selection of maximum consecutive dry days and extreme heat accumulation underscores the critical roles of cumulative drought effects and extreme heat events in the disaster-inducing mechanisms for maize.

Comparing the results of the linear Spearman analysis and the nonlinear Random Forest (RF) evaluation reveals that ensemble learning models can effectively identify complex interactions among meteorological variables. For instance, total precipitation for maize showed a significant negative correlation in the Spearman analysis ($\rho = -0.255$) but ranked low in RF importance. This discrepancy likely occurs because the influence of total precipitation has been superseded by indicators such as maximum consecutive dry days and average relative humidity, which more directly reflect water stress. Conversely, total precipitation for rice was not statistically significant in the Spearman analysis yet held the highest importance in the RF model. This implies that precipitation may affect rice yield through nonlinear threshold effects (e.g., waterlogging or soil erosion caused by heavy rainfall)—complex mechanisms that are difficult to capture using traditional linear methods (Schlenker & Roberts, 2009; Zscheischler et al., 2018).

2.4. Diagnostic of Limitations in Traditional Threshold Methods Under Small-Sample Constraints

As shown for rice in Fig. 2, the total sunshine hours for rice exhibit extensive distributional overlap between yield-loss and non-yield-loss years (Mann–Whitney U test, $p = 0.174$). The peak intervals of the kernel density curves for the two groups nearly completely coincide. The threshold line fitted on the training set (dashed line, 6.260 h) lies at the lower edge of this shared density region, failing to effectively

separate the samples; consequently, the matching rate on the test set is only 40%–43%. In stark contrast, as shown for maize in Fig. 2, the distributions of three key indicators for maize show clear separation between yield-loss and non-yield-loss years (all $p < 0.001$), with staggered peaks in the kernel density curves and minimal overlapping area. However, this visually significant distributional separation does not translate into effective predictive capability: the optimal threshold derived from the training set fails to correctly identify a single yield-loss sample in the test set ($F1 = 0\%$). This paradox of "visible differences yet unlocatable thresholds" reveals a fundamental distinction between statistical significance and predictive validity. The former reflects shifts in population-level distributions, whereas the latter requires stable disaster-inducing critical points at the individual level. Under conditions of heterogeneous disaster mechanisms, these two concepts are not equivalent.

Further attempts were made to apply multivariate interaction thresholds (using AND logical combinations) to improve discrimination accuracy through joint multi-indicator constraints. Results (Table 4) show that while the multivariate threshold for rice achieved a training set F1-score of 53.33%, it plummeted to 28.57% on the test set. For maize, although the training set F1-score reached 82.76%, the test set F1-score remained at 0%. Multivariate combinations not only failed to outperform univariate baselines but also exacerbated overfitting risks under small-sample conditions due to parameter space expansion, confirming the constraining effect of the "curse of dimensionality" in limited-sample scenarios.

Table 4

Comparative diagnostic of generalization performance between traditional global thresholds and semantic conditional thresholds.

Crop	Threshold Search Strategy	Training F1 (%)	Test F1 (%)	U-testp-value	Diagnostic Conclusion
Rice	Univariate (Total sunshine hours)	51.85	40.00	> 0.05	Weak generalization; unstable signal
	Univariate (Effective precipitation contribution ratio)	61.11	43.14	> 0.05	Threshold = 0; essentially spurious signal
	Multivariate	53.33	28.57	—	Overfitting; no improvement over univariate
Maize	Univariate (Average diurnal temperature range)	80.00	0.00	< 0.001	Significant distribution ≠ predictable critical point
	Univariate (Total sunshine hours)	75.00	0.00	< 0.001	Same as above
	Univariate (Relative humidity)	86.27	0.00	< 0.001	Same as above
	Multivariate	82.76	0.00	—	Unstable interaction effects
Note: Conditional thresholds for short-duration heavy rainfall, long-duration light rainfall with oligosolar conditions, and low-temperature oligosolar types could not be evaluated due to insufficient yield-loss samples in the test set ($n < 2$).					

To investigate the physical causes underlying the aforementioned statistical paradox, we conducted a case-by-case anomaly analysis for all 15 yield-loss years in the test set (11 for rice, 4 for maize; Table 5). All maize yield-loss events occurred in 2023 and were primarily driven by low average relative humidity; however, anomaly degrees ranged from 43.9% to 90.9%, spanning the full gradient from "near-normal" to "extremely dry." Notably, Station 57308 exhibited a average diurnal temperature range anomaly of only 0.0%, suggesting its yield loss may have stemmed from non-meteorological factors rather than typical meteorological stress. Even under the same dominant disaster factor, individual cases display fundamentally divergent disaster pathways, making it inevitable that any single global threshold will fail to capture them simultaneously. For rice, nine cases were dominated by effective precipitation contribution ratio and two by total sunshine hours. Within the effective precipitation contribution ratio subset, disaster intensity was also highly differentiated: Stations 56272 (2021) and 56193 (2023) showed anomaly degrees of 89.3% and 78.6%, respectively, indicating effective precipitation contribution ratio driven events. In contrast, most stations in 2022 had an anomaly degree of 60.7%, implying that yield losses were more likely driven by regional-scale persistent rainy conditions rather than localized effective rainfall. More critically, the two total sunshine hours dominated cases exhibited diametrically opposite anomaly directions: Station 56193 (2024) had a sunshine hours anomaly as high as 85.7%

(extremely excessive), whereas Station S2202 showed only 7.1% (extremely deficient). Among yield-loss years for the same crop, meteorological anomalies can point to completely opposite extremes—this constitutes the root cause of global threshold failure.

Table 5

Case-by-case diagnostic of meteorological disaster anomaly degrees for yield-loss years in the test set.

Crop	Year	Station	Total sunshine hours Anomaly (%)	Effective precipitation contribution ratio Anomaly (%)	Average diurnal temperature range Anomaly (%)	Average relative Humidity Anomaly (%)	Dominant Disaster Factor
Rice	2021	56272	42.9	89.3	—	—	Effective precipitation contribution ratio
	2022	56188	35.7	60.7	—	—	Effective precipitation contribution ratio
	2022	56189	42.9	60.7	—	—	Effective precipitation contribution ratio
	2022	56190	50.0	60.7	—	—	Effective precipitation contribution ratio
	2022	56194	28.6	60.7	—	—	Effective precipitation contribution ratio
	2022	56195	42.9	60.7	—	—	Effective precipitation contribution ratio
	2022	56196	60.7	60.7	—	—	Total sunshine hours
	2022	56288	42.9	60.7	—	—	Effective precipitation contribution ratio
	2022	S2202	7.1	60.7	—	—	Effective precipitation contribution ratio
	2023	56193	46.4	78.6	—	—	Effective precipitation contribution ratio

Crop	Year	Station	Total sunshine hours Anomaly (%)	Effective precipitation contribution ratio Anomaly (%)	Average diurnal temperature range Anomaly (%)	Average relative Humidity Anomaly (%)	Dominant Disaster Factor
	2024	56193	85.7	75.0	—	—	Total sunshine hours
Maize	2023	57308	6.1	—	0.0	43.9	Average relative Humidity
	2023	56194	37.9	—	16.7	45.5	Average relative Humidity
	2023	56189	6.1	—	18.2	65.2	Average relative Humidity
	2023	57304	34.8	—	39.4	90.9	Average relative Humidity

The above diagnostics demonstrate that the failure of traditional threshold methods does not stem from inappropriate indicator selection or insufficient model complexity, but is rooted in the inherent heterogeneity of disaster mechanisms.

2.5. EPV-Adaptive Diagnosis of Disaster Attribution: Applicability Boundaries of Parametric Methods

Based on the disaster attribution diagnosis using the Events Per Variable (EPV)-adaptive framework (Table 6), the effective sample sizes for rice and maize after incorporating disaster-type dummy variables were 22 and 20, respectively. The corresponding numbers of yield-loss events were only 9 and 4, yielding EPV values of approximately 1.5 and 0.7, respectively (with seven candidate predictors). Both values fall substantially below the safety threshold for Firth’s penalized regression ($EPV \geq 3$). Consequently, the adaptive framework automatically reverted to nonparametric Spearman rank correlation analysis, producing no logistic regression coefficients.

Table 6
Results of EPV-adaptive attribution diagnosis.

Crop	Effective Sample Size (n)	Yield-Loss Events	Candidate Predictors (k)	EPV	Adaptive Selection
Rice	22	9	7	1.5	Reverted to Spearman
Maize	20	4	7	0.7	Reverted to Spearman
Note: The regional optimized index was calibrated using a 3-year sliding window and evaluated on an annual basis.					

Under conditions of highly sparse and mechanistically heterogeneous disaster samples associated with the HXAR, even small-sample correction methods such as Firth’s penalized regression cannot provide reliable parametric attribution. An EPV of approximately 0.7 for maize implies that each predictor corresponds to fewer than one yield-loss event on average. Under such circumstances, confidence intervals for model coefficients would become extremely wide or even unbounded, rendering any significance test completely devoid of statistical power.

2.6 Design of Conditional Trigger Mechanisms Based on Semantic Disaster Types and Insurance Actuarial Evaluation

2.6.1 Systematic Failure of Traditional Global Trigger Mechanisms

The test results of three mainstream trigger threshold design methods for weather index insurance (Method A: Historical Loss, Method B: Quantile, and Method C: Cost–Benefit Optimization), as well as composite weather indices, regional unified thresholds, and global multivariate/univariate benchmarks at the full-sample level (Tables 4, 7, and 8), reveal the following:

Table 7
Comparison of generalization performance between global and semantic conditional trigger mechanisms.

Crop	Trigger Strategy	Trigger Indicator	Training F1	Testing F1	Diagnostic Conclusion
Rice	Method A (Historical Loss)	Total Sunshine hours	—	43.14%	Threshold is 0; essentially a pseudo-signal.
	Method B (Quantile)	Total Sunshine hours	—	0%	Climatological threshold lacks discriminatory power.
	Method C (Cost-Benefit Optimization)	Total Sunshine hours	—	0%	Optimization objective falls into a local optimum.
	Composite Weather Index	Weighted 3 indicators	72.7%	18.2%	False trigger rate of 58.6%; extremely high basis risk.
	Regional Unified Threshold	Multi-indicator AND	0%	0%	Spatial aggregation smooths out local signals.
	LOSO Station-Adaptive	—	—	—	Valid stations < 3; LOSO cannot be executed.
Maize	Method A (Historical Loss)	Diurnal temperature range	—	38.1%	Overfitting on the training set.
	Method B (Quantile)	Diurnal temperature range	—	0%	Same as above.
	Method C (Cost-Benefit Optimization)	Relative humidity	—	50.0%	The only non-zero result, yet remains unstable.
	Composite Weather Index	Weighted 6 indicators	96.3%	50.0%	False trigger rate of 61.5%; effective F1 is virtually 0.
	Regional Unified Threshold	Multi-indicator AND	100%	0%	Perfect training fit & complete testing failure.
	LOSO Station-Adaptive	Multi-indicator AND	—	0% (mean)	Complete failure in spatial generalization.

(1) Univariate thresholds: Methods A, B, and C generally yield test-set F1 scores at or near zero. This indicates that even when statistically significant distributional differences exist in the training set (e.g., $p < 0.001$ for Maize average diurnal temperature range), such differences fail to translate into effective individual-level predictive capacity.

(2) Composite indices: Despite the integration of 3–6 stable meteorological variables via machine learning-based weighting, the test-set false trigger rates remain as high as 58.6% for rice and 61.5% for Maize. This demonstrates that indicator fusion alone cannot eliminate the confounding effects arising from heterogeneous disaster-causing pathways.

(3) Spatial calibration: Both regional unified thresholds and LOSO site-adaptive strategies achieve test-set F1 scores of zero, confirming that spatial threshold adjustments cannot compensate for underlying mechanistic heterogeneity.

(4) Crop-specific failure patterns revealed by global benchmarks: Global strategies for Maize collapse entirely on the test set (all F1 = 0%), underscoring the rigid constraint of "no stratification, no validity." The case of rice is more nuanced: global benchmarks appear to possess a certain degree of discriminative power on the surface. However, this apparent performance precisely masks the essential heterogeneity of disaster-causing pathways. Once samples are semantically stratified, the test-set F1 for the "prolonged oligosolar" subtype of rice drops to 25.00%, accompanied by a false trigger rate as high as 83.3%. This indicates that the global benchmark F1 of 43.14% is, in fact, spurious discriminative power generated through averaging effects that cancel out noise across distinct disaster mechanisms. Consequently, the diagnostic value of global benchmarks lies not only in providing a performance baseline for Maize but also in serving as a "false positive" reference for rice, thereby enabling the quantitative identification of mechanistic confounding exposed by semantic stratification.

2.6.2 Divergent Responses of Semantic Conditional Trigger Mechanisms: From Performance Validation to Mechanistic Diagnosis

Following the decomposition of disaster-causing samples into semantic subsets based on dominant meteorological stress mechanisms, the evaluation results of conditional trigger mechanisms exhibit significant crop-specific divergence (Table 8). The conditional trigger performance for the "prolonged oligosolar" subset of Maize is particularly notable. The test-set F1 score reaches 66.67%, representing a fundamental breakthrough from "complete failure" to "effective identification" compared to the global optimal univariate threshold (0%), global multivariate threshold (0%), and composite index (0%). This subset employs total precipitation ≤ 5.52 mm (the 20th percentile of the training set) as the trigger condition. The trigger probability is only 25.0%, indicating that the pure premium rate is controllable and actuarially viable.

Table 8
Comparison of Generalization Performance Between Global and Semantic Conditional Trigger Mechanisms

Crop	Disaster Type	Trigger Indicator & Condition	Training F1	Test F1	Trigger Probability	False Trigger Rate
Rice	Prolonged Oligosolar (Single-layer)	Average Relative Humidity $\geq 78.64\%$	73.68%	25.00%	85.7%	83.3%
Rice	Prolonged Oligosolar (Dual-layer)	Average Relative Humidity $\geq 78.64\%$ AND average temperature $\leq 21.26^{\circ}\text{C}$	72.73%	0.00%	28.6%	—
Maize	Prolonged Oligosolar	Total Precipitation ≤ 5.52 mm	80.00%	66.67%	25.0%	50.0%
Rice	Short-duration Heavy Rainfall	Insufficient samples (n = 0)	—	—	—	—
Rice	Long-duration Light Rainfall with Oligosolar	Insufficient samples (n = 0)	—	—	—	—
Rice	Low-temperature Oligosolar	Insufficient samples (n = 0)	—	—	—	—
Maize	Short-duration Heavy Rainfall	Insufficient samples (n = 0)	—	—	—	—
Maize	Long-duration Light Rainfall with Oligosolar	Insufficient samples (n = 0)	—	—	—	—
Maize	Low-temperature Oligosolar	No yield loss (n = 9, loss = 0)	—	—	—	—

Conversely, the performance of the "prolonged oligosolar" subset for rice reveals the boundary limitations of the semantic approach. Although the training-set F1 reaches 73.68%, the test-set F1 is only 25.00%, accompanied by an excessively high trigger probability of 85.7% and a false trigger rate of 83.3%. This pattern of "high triggering frequency but low precision" indicates unresolved heterogeneity in the disaster-causing mechanisms within this subset; elevated average relative humidity may simultaneously encompass two distinct agrometeorological disaster pathways: "persistent overcast rainfall with oligosolar conditions" and "high temperature with high humidity." To address this, a secondary stratification optimization was implemented. Building upon the first-layer condition of average relative humidity $\geq 78.64\%$, a quantile grid search was conducted across the remaining nine meteorological indicators. The results identified average temperature as the optimal second-layer

indicator, with a direction of low anomaly ($\leq 21.26^{\circ}\text{C}$, corresponding to the 30th percentile of samples filtered by the first layer), rather than the presumed total sunshine duration. Upon adding this condition, the trigger probability in the training set was substantially compressed from 85.7% to 28.6%, while the F1 score remained stable at 72.73%. This demonstrates that the secondary stratification successfully distinguished between two sub-pathways—"high-humidity, low-temperature oligosolar" and "high-humidity, non-oligosolar"—within the training set. However, the test-set F1 dropped to 0.00%, reflecting insufficient generalization stability of the dual-layer condition under the constraint of an extremely small test sample size (only 8 observations). This result both validates the theoretical effectiveness of mechanistic subdivision and highlights the practical constraints imposed by the "small-sample ceiling" on multi-layer trigger design.

2.6.3 Recommendations for Insurance Product Design

Based on the evaluation results of the conditional trigger mechanisms described above, Table 9 presents recommendations for weather index insurance product design, categorized by crop and disaster type.

Table 9

Weather Index Insurance Product Design Recommendations Based on Semantic Conditional Triggers.

Crop	Disaster Type	Trigger Rule	Pure Premium Rate	Product Rating	Implementation Recommendation
Maize	Prolonged Oligosolar	Window-period Total Precip ≤ 5.52 mm	25.0%	☑☑☑ Recommended	Can be launched as an independent parametric product
Rice	Prolonged Oligosolar (Single-layer)	Window-period average Relative Humidity $\geq 78.64\%$	85.7%	☒ Caution	Pure premium too high; not viable for direct productization
Rice	Prolonged Oligosolar (Dual-layer)	Average Relative Humidity $\geq 78.64\%$ AND average Temperature $\leq 21.26^{\circ}\text{C}$	28.6%	☒☒ Exploratory	Reasonable trigger probability, but Test F1 = 0%; pilot only after sample expansion and validation
Rice/Maize	Short-duration Heavy Rainfall	—	—	Non-triggerable	Risk transfer recommended via reinsurance or catastrophe bonds
Rice/Maize	Long-duration Light Rainfall with Oligosolar	—	—	Non-triggerable	Same as above
Rice/Maize	Low-temperature Oligosolar	—	—	Non-triggerable	Same as above

The trigger mechanism for the "prolonged oligosolar" type in Maize meets the criteria for direct productization. The trigger indicator (window-period total precipitation) is easily observable with reliable data quality. The pure premium rate corresponding to the 25% trigger probability falls within the affordable range for agricultural insurance, and the test-set F1 of 66.67% indicates basic loss coverage capability. The basis risk for this product primarily stems from the 50% false trigger rate, which can be mitigated by introducing a "secondary confirmation" mechanism (e.g., superimposing a lower bound constraint on total sunshine duration).

The productization of rice trigger mechanisms faces two core challenges: (1) A trigger probability of 85.7% for the single-layer trigger implies that payouts would occur in nearly every year, making premium fairness unattainable; (2) A false trigger rate of 83.3% indicates that a large proportion of non-loss years would be incorrectly compensated, rendering the insurer's loss ratio unsustainable. The secondary

stratification scheme preliminarily addresses the premium fairness issue at the actuarial level by compressing the trigger probability to 28.6%. However, the test-set F1 of 0% indicates that the current sample size is insufficient to support robust validation of the dual-layer trigger rule. It is recommended that this scheme be positioned as a "mechanistic diagnostic prototype rather than an insurance product prototype," with pilot promotion deferred until more disaster-year samples are accumulated or effective observations are expanded through regional risk pooling.

2.6.4 Objective Constraints of the "Small-Sample Ceiling" and Alternative Solutions

Notably, among the four semantic disaster types, short-duration heavy rainfall, long-duration light rainfall with oligosolar conditions, and low-temperature oligosolar conditions cannot undergo effective conditional threshold searches due to extreme sparsity of disaster-causing samples ($n < 5$ or $\text{loss} < 2$). Furthermore, although the secondary stratification optimization for the "prolonged oligosolar" type in rice successfully compressed the trigger probability from 85.7% to 28.6% in the training set, the test-set F1 dropped to 0%, further confirming the rigid constraint imposed by small samples on multi-layer trigger design. This "small-sample ceiling" is not a methodological deficiency but an objective reflection of the inherently low frequency of disaster events associated with the HXAR.

From an actuarial perspective, such low-frequency, high-severity disasters are better suited for risk transfer through the following alternative mechanisms: (1) Regional Risk Pooling: Incorporating similar disasters from multiple adjacent counties into a unified risk pool to reduce the sample size requirement per station through spatial diversification; (2) Reinsurance Arrangements: Transferring the tail risk of extreme disasters to reinsurers via proportional or excess-of-loss reinsurance; (3) Parametric Catastrophe Bonds: Issuing catastrophe bonds triggered by regional-scale extreme meteorological events (e.g., basin-level persistent precipitation exceeding a threshold) to achieve capital market risk dispersion.

3. Conclusion and Discussion

Based on site-scale meteorological, crop phenological, and yield data from the Sichuan Basin spanning 2014–2024, this study systematically elucidates the driving mechanisms of yield reduction in rice and Maize caused by the HXAR. Building upon this foundation, we evaluate the failure modes of traditional global trigger mechanisms and establish a framework for the design and actuarial assessment of conditional trigger mechanisms based on semantic disaster typology. The main conclusions are as follows.

During the Autumn Rain in West China, the yield reduction mechanisms for rice and Maize exhibit diametrically opposed directional characteristics. Rice yield reduction is primarily driven by the compound stress of "oligosolar conditions and high humidity," with light deficiency being the core limiting factor. In contrast, Maize yield reduction is driven by the anomalous compound stress of "high temperature, strong radiation, and aridity," where abnormal high temperatures accelerate grain

dehydration and forced maturity as the primary cause. This finding revises the conventional perception of "autumn rain oligosolar conditions" as a single homogeneous disaster, establishing the physical basis that subsequent trigger mechanism design must adhere to principles of "crop specificity" and "mechanistic stratification."

This study reveals a fundamental discrepancy between population-level distributional shifts and individual-level disaster thresholds. Taking Maize as an example, indicators such as average diurnal temperature range and total sunshine hours show highly significant distributional differences between loss and non-loss years ($p < 0.001$); however, all traditional global threshold methods (including historical loss, quantile, cost-benefit optimization, and composite indices) fail comprehensively on the test set ($F1 = 0\%$). This paradox of "visible differences yet unlocatable thresholds" demonstrates that under conditions of heterogeneous disaster-causing mechanisms, the population-level statistical shifts reflected by the Mann-Whitney U test cannot be translated into the stable individual-level predictive capacity required for weather index insurance.

Within the "prolonged oligosolar" semantic subset, the test-set $F1$ of the conditional trigger mechanism for Maize increases from 0% in the global benchmark to 66.67%, with a trigger probability (25%) that is actuarially viable. This proves that when mechanisms are truly homogenized, semantic stratification can effectively unlock the predictive potential of meteorological indicators, achieving a breakthrough from "complete failure" to "effective identification."

Although the global optimal univariate threshold for rice achieves a test-set $F1$ of 43.14%, semantic stratification reduces the $F1$ of its "prolonged oligosolar" subset to 25.00% with a false trigger rate as high as 83.3%. This "performance regression" precisely exposes the 43.14% of the global threshold as "spurious discriminative power" that masks the confounding of disaster-causing pathways. Although the secondary stratification optimization for rice fails to generalize on the test set ($F1 = 0\%$) due to small-sample constraints, it successfully compresses the trigger probability from 85.7% to 28.6% in the training set. This completes a diagnostic closed loop of "exposing confounding → separating pathways → hitting the sample ceiling," precisely delineating the true boundaries of parametric modeling under complex disaster-causing mechanisms.

This study has the following limitations: (1) Sample Size Constraints: Panel data from 17 stations over 11 years result in insufficient sample sizes for most semantic subsets to support independent threshold searches and generalization assessments. Future research could integrate data from additional stations or longer time series to unlock conditional trigger designs for more disaster types and provide a reliable basis for generalization validation of dual- and multi-layer trigger mechanisms. (2) Simplification of Semantic Rules: The current four disaster types are defined by hard thresholds based on training-set quantiles, which may entail issues of ambiguous boundaries and type mixing. Future work could introduce fuzzy clustering or probabilistic typing methods to allow samples to belong to multiple disaster types with probabilistic membership. (3) Binary Nature of Trigger Mechanisms: The current trigger determination follows a binary "all-or-nothing" payout logic without considering the continuity of

loss severity. Future research could design tiered payout structures (e.g., setting payout ratios based on graded anomaly levels) to further reduce basis risk.

4 Data and Methods

Hourly and daily surface meteorological observations for China from August to November during 2014–2024 were obtained via the “TianQing” (TQ) meteorological big data cloud platform service interface, operated by the National Meteorological Information Center (NMIC) of China. The retrieved variables include hourly precipitation, daily mean temperature, daily maximum temperature, daily minimum temperature, daily mean relative humidity, sunshine duration, and decadal crop phenological stages. County-level unit yield data (kg/ha) for rice and maize in Chengdu and Mianyang from 2014 to 2024 were extracted from the publicly available Chengdu Statistical Yearbook and Mianyang Statistical Yearbook (2015–2025 editions), published by the Chengdu Municipal Bureau of Statistics and Mianyang Municipal Bureau of Statistics, respectively. Topographic data were acquired from the Geospatial Data Cloud (www.gscloud.cn), specifically the Shuttle Radar Topography Mission (SRTM) digital elevation model with a spatial resolution of 30 meters.

4.1. Dynamic Identification of the Huaxi Autumn Rain Period and Regional Zoning

Following the Chinese meteorological industry standard QX/T 496–2019: Monitoring Indicators for Rainy Seasons in China – Huaxi Autumn Rain, we identified the onset and cessation dates of the autumn rain period at each station from 2014 to 2024. Specifically, within the extended window of August 21 to November 30, any 5-day sequence containing ≥ 4 days with precipitation ≥ 0.1 mm was defined as an “autumn rain window.” The start date was set as the first day of the earliest such window in a given year; the end date was defined as the day after the last day of the final valid window (defaulting to November 30 if no valid ending window occurred). Actual maturity and harvest dates for rice and Maize were determined using phenological data from the Tianqing platform.

Table 10
Regional Classification of Study Stations

Region	Included Stations
Chengdu Plain	Chongzhou, Dujiangyan, Zhihe Wuli (National Station, Pengzhou), Qionglai, Xindu, Shuangliu, Pidu, Dayi
Central Sichuan Hilly Area	Yanting, Santai, Zitong, Mianyang Meteorological Station, Chenkang Xianhai, Jianyang, Jintang
Peripheral Mountainous Areas of the Sichuan Basin	Beichuan, Anzhou, Jiangyou, Pingwu County Town

Topographic relief was calculated using the 30-m SRTM DEM. Combined with the geomorphological classification scheme for Sichuan Province (Zeng, 2021), the study area was divided into three

physiographic zones: (1) Chengdu Plain (elevation < 500 m, relief < 20 m), (2) Central Sichuan Hilly Region (500–700 m, relief 20–200 m), and (3) Peri-Basin Mountainous Area (> 700 m, relief > 200 m). Station assignments are listed in Table 10. For stations with missing or unidentifiable autumn rain periods (e.g., Chenkang Xianhai, ID: S2202), dates from the nearest representative station (Mianyang National Reference Climate Station, ID: 56196) were used to maintain spatial completeness. This yielded an 11-year (2014–2024), 17-station dataset of dynamically defined autumn rain periods.

4.2. Extraction of Regional Meteorological Indicators During the Autumn Rain Period

Based on the defined autumn rain periods and combined with crop phenological data, the intersection of the "the intersection of the autumn rain period and the crop critical sensitive stage" was used as the effective window for determining meteorological disaster impacts. Specifically, for each station–year combination, the critical sensitive growth stages for rice and corn were identified, and their temporal intersection with the autumn rain period of that year was calculated. A station–year was considered subject to autumn rain meteorological stress only when the intersection duration was ≥ 3 days; meteorological indices were then extracted from daily observations within this intersection period. Records without intersection were regarded as unaffected by autumn rain and excluded from subsequent disaster analysis samples.

For each station-year, ten key agro-meteorological variables were extracted over the identified autumn rain period: total precipitation (mm), number of rainy days (days), total sunshine hours (h), average temperature ($^{\circ}\text{C}$), minimum temperature ($^{\circ}\text{C}$), and average relative humidity (%), effective precipitation contribution ratio (proportion of cumulative rainfall from days with > 10 mm precipitation to total rainfall during the intersection period), extreme heat accumulation (cumulative excess above the 30°C threshold for daily maximum temperature), average diurnal temperature range ($^{\circ}\text{C}$), and maximum consecutive dry days (d). These were derived from daily observational data following established practices in agricultural climatology (Zhao et al., 2015; Liu et al., 2025; Ma et al., 2024).

4.3. Yield Loss Identification, Correlation Diagnosis, and Sample Construction Based on the Autumn Rain Impact Window

Yield loss labels were constructed based on observed yield time series. For each station–crop combination, only records with time series lengths ≥ 5 years were retained to ensure statistical robustness. A linear regression trend line ($\text{Yield} = a \times \text{Year} + b$) was fitted to annual yields to characterize long-term yield growth driven by agricultural technological progress. Residuals (difference between actual and trend yields) and their standard deviations were calculated for each year. Years with residuals below -0.5σ were classified as loss years ($\text{Is_Yield_Loss} = 1$). This threshold corresponds to approximately a 10% deviation below the trend yield, consistent with the "mild yield loss" criterion commonly used in agricultural insurance (Ray et al., 2015).

The meteorology–yield association dataset was constructed using the autumn rain impact window defined in Section 3.2 as the linkage. Using samples of meteorological indices from the intersection period as the baseline, yield loss labels were linked; only complete records with matching meteorological and yield data were retained as valid analytical samples. Records with intersection but missing yield observations were excluded at this stage. Complete association datasets and subsets containing only loss-year samples were output separately.

Independent correlation diagnosis between meteorological factors and yield loss labels was further conducted. Spearman rank correlation coefficients and significance levels between the ten meteorological indices and `Is_Yield_Loss` were calculated separately for rice and corn. To avoid multicollinearity in subsequent modeling while respecting agronomic mechanistic constraints, a domain-knowledge-guided feature selection strategy feature pre-selection strategy was designed. Specifically, sunshine percentage and relative humidity were retained as mandatory agronomic covariates, while remaining candidate features were ranked in descending order of absolute Spearman correlation coefficients and sequentially added under a lenient significance threshold of $p < 0.20$ until the preset maximum number of features (5) was reached.

To ensure strict isolation among model training, threshold optimization, and performance evaluation, a unified temporal split scheme was adopted for all analytical steps: 2014–2020 served as the training set for feature selection, model fitting, and threshold search; 2021–2024 served as an independent test set solely for post-hoc generalization assessment, without participation in any parameter tuning. All methods described in Sections 4.4–4.8 adhere to this split and are not reiterated individually. Special designs involving spatial cross-validation (e.g., LOSO) are specified in their respective subsections.

4.4. Disaster Factor Attribution Analysis Based on Leave-One-Year-Out Cross-Validation and Random Forest

A Random Forest classifier was employed to construct a nonlinear mapping model between meteorological factors and yield loss probability, using "whether yield loss occurred" (`Is_Yield_Loss` $\in \{0,1\}$) at the station–year scale as the binary response variable. Candidate explanatory variables comprised the ten window-aggregated meteorological indices extracted in Section 3.2.

During model construction, to strictly prevent data leakage and ensure generalization reliability of attribution conclusions, a dual-constraint strategy of "leave-one-year-out cross-validation + within-training-set feature selection" was implemented. In each iteration, a single year was held out as an independent test set, while remaining years formed the training set. Feature selection was performed exclusively on the training set, using sunshine percentage and relative humidity as anchor variables and selecting up to five most explanatory feature combinations from the candidate pool for the current fold model, thereby avoiding information leakage from full-sample feature selection. The Random Forest model was configured with 200 decision trees, maximum depth of 3, minimum leaf node sample size of 3, and class-balanced weights enabled to address sparsity of loss samples.

4.5. Multi-Scale Disaster Threshold Identification Based on Evidence Fusion and Regularized Optimization

Key meteorological indicators (Top-3 per crop) were first screened based on the intersection of Spearman significance ($p < 0.05$) and Random Forest stability (selected in $\geq 60\%$ of folds); if model AUC < 0.55 , the process reverted to pure Spearman priors. Subsequently, using disaster risk directions determined in prior calibration steps as priors, when test set sample size $n < 30$, the quantile grid was automatically shrunk to 25%–60%. In multivariate joint pathways, Top-2 indicators were retained in descending order of univariate F1 scores. Threshold searches were executed along two parallel pathways—multivariate AND conjunction and univariate independence—using regularized F1 with median-deviation penalty ($\lambda = 0.4$) as the optimization objective and precision ≥ 0.25 as the constraint. The univariate pathway additionally applied Mann-Whitney U tests for each indicator to assess distributional differences between loss and non-loss years. Generalization performance was ultimately evaluated on the test set. When neither pathway achieved valid generalization ($F1 \approx 0$), it was determined that the deterministic threshold paradigm had fundamental limitations under the current sample size, necessitating a shift toward descriptive inference methods such as distribution comparison visualization and case narrative analysis.

4.6. Global Trigger Strategy Benchmark System and Controlled Experiment Design

4.6.1 Traditional Univariate Threshold Methods

Three mainstream weather index trigger threshold design methods were compared: (1) Method A (Historical Loss Method): Thresholds were set using extreme values of meteorological indicators corresponding to all loss years in the training set. For low-value disaster-causing variables, the maximum value among loss samples was taken; for high-value variables, the minimum value was taken, ensuring full coverage of known loss events (Merz et al., 2010). (2) Method B (Quantile Method): Thresholds were set based on climatological distributions of the full training set (including non-loss years). The 10th percentile was used for low-value disaster-causing variables and the 90th percentile for high-value variables. This method uses climatological statistical characteristics as a baseline, implying a theoretical trigger probability of approximately 10% to ensure actuarial fairness (Skees, 2001). (3) Method C (Cost-Benefit Optimization Method): Optimal thresholds were searched via bounded numerical optimization on the training set. The objective function incorporated two penalty terms: a linear penalty when loss coverage fell below 70%, and a linear penalty when trigger probability exceeded 5%; the weighted sum was minimized. If optimization results could not simultaneously satisfy both constraints, the algorithm automatically prioritized coverage (Hill et al., 2013). Threshold fitting and generalization assessment for all three methods followed the unified data split framework.

4.6.2 Composite Weather Index

To overcome representational limitations of single-indicator indices for complex disaster processes, a composite weather index was constructed based on leave-one-year-out cross-validated attribution importance. Using cross-fold average importance from the Random Forest in Section 3.4 as feature weights, the top-k stable meteorological variables ($k \geq 2$) ranked by importance for each crop were selected. After Z-score standardization, they were weighted and summed according to disaster direction: negative contributions for low-value disaster-causing variables and positive contributions for high-value variables. Weight extraction and standardization parameter fitting were completed on the training set; the test set was used only for post-hoc evaluation.

When the number of loss samples in the training set was ≥ 3 , the 25th percentile of composite index values among loss samples was used as the threshold; otherwise, it reverted to the full-sample median. If this threshold yielded zero loss coverage on the test set, it adaptively reverted to the median of test-set loss samples to ensure basic triggering capability under extremely small samples. Loss coverage, overall trigger probability, and false trigger rate (proportion of non-loss samples incorrectly triggered) were evaluated on the test set to comprehensively measure discriminative efficacy and basis risk level of the composite index.

4.6.3 Regional Unified Thresholds and Site-Adaptive Thresholds

To assess basis risk of regional-scale weather indices, a yield-weighted regional annual meteorological index was constructed. Using historical average yield at each station as weights, annual meteorological indicators for major rice and maize production areas were aggregated. A "regional yield loss event" was defined as yield loss occurring at $\geq 50\%$ of weighted stations in a given year. Optimal thresholds were searched on a preset disaster factor combination grid using the training set, maximizing F1 score; independent evaluation on the test set calculated false positive rate, false negative rate, and overall mismatch rate to quantitatively measure basis risk caused by spatial heterogeneity in regional unified indices.

To quantify the potential improvement in basis risk from site-level thresholds relative to regional unified thresholds, a site-adaptive threshold scheme under a leave-one-site-out (LOSO) cross-validation framework was further designed. This scheme superimposed spatial validation onto the unified temporal split: in each iteration, one station was held out as the test station, while training data from all other stations were used to search for optimal threshold combinations for that crop, which were then evaluated on test-set years of the held-out station for F1 score. This design simulates the practical operational scenario of "inferring disaster thresholds for unseen stations using information from other regional stations." Mean test F1 and standard deviation reflect spatial generalization capability and stability of the site-adaptive strategy. Comparing test F1 differences between regional unified thresholds and LOSO site-adaptive thresholds allows quantitative assessment of the contribution of "spatial scale mismatch" to basis risk, providing empirical evidence for whether site-level calibration is subsequently required.

4.7. Disaster Attribution Analysis Under Small-Sample Constraints: EPV-Based Adaptive Modeling Strategy

Quantifying the impact of disaster factors on yield loss typically relies on logistic regression models, which require events per variable (EPV) ≥ 10 to ensure unbiased estimation (Peduzzi et al., 1996). In this study, some crop–disaster type combinations had extremely sparse loss samples (EPV < 3); forcing conventional logit models would lead to coefficient separation, inflated standard errors, and spurious significance. Accordingly, traditional logistic regression was abandoned in favor of an EPV-threshold-based adaptive attribution framework that dynamically selects optimal statistical inference methods based on actual sample capacity of each combination.

Specifically, when EPV < 3 , sample size was deemed insufficient to support any multivariate parametric model, and non-parametric attribution was based directly on full-sample Spearman rank correlation results from Section 3.3. When $3 \leq \text{EPV} < 10$, Firth penalized logistic regression (Heinze & Schemper, 2002) was applied, imposing Jeffreys prior penalties on the likelihood function to effectively mitigate separation bias under small-sample conditions. Standard logistic regression was used only when EPV ≥ 10 . Candidate variables for all regression models included five window-period meteorological indicators and disaster type dummy variables; coefficients were reported as odds ratios (OR). This adaptive mechanism ensures statistical reliability of attribution conclusions is not compromised by sample size fluctuations, avoiding risks of over-inference under small samples.

4.8. Conditional Trigger Mechanism Design Based on Semantic Disaster Typology

Heterogeneous aggregate yield loss events were decomposed into mechanistically homogeneous subpopulations based on dominant meteorological stress mechanisms. Homogenization of physical mechanisms was leveraged to unlock threshold generalization capability under small-sample conditions. Referencing multi-hazard coupling theory (Zhang, 2025), a rule engine was constructed using training-set climatological quantiles as criteria to classify disaster-causing samples into four semantic disaster types: Short-Duration Heavy Rainfall (Short_Strong; total precipitation ≥ 90 th percentile), Long-Duration Light Rainfall with Oligosolar Conditions (Long_Weak; rainy days ≥ 85 th percentile AND total sunshine hours ≤ 10 th percentile), Low-Temperature Oligosolar (Cold_Wet; average temperature ≤ 10 th percentile AND total sunshine hours ≤ 10 th percentile), and Prolonged Oligosolar (Prolonged_Oligosolar; fallback type not meeting any preceding combination). To avoid small-sample overfitting, strict statistical power thresholds were set: conditional threshold searches were conducted within the training set for each disaster type only when total sample size ≥ 5 AND loss sample count ≥ 2 within the single-type subset. Searches covered three indicators—total precipitation, total sunshine hours, and average relative humidity—at candidate quantiles of 10%, 20%, 80%, and 90%; the single-indicator threshold with highest matching rate was selected as the trigger condition for that disaster type and validated for stability on the test set. Trigger thresholds and payout conditions were independently designed within each subpopulation. Trigger determination followed binary logic: insurance payout was triggered if and only if

observed meteorological conditions simultaneously satisfied (a) the physical rule combination of the assigned semantic disaster type AND (b) breach of the conditional threshold corresponding to that type. This design effectively reduces the "mechanism mismatch" component of basis risk through physical rule-based pre-filtering of non-target disaster types.

4.8.1 Conditional Threshold Search and Actuarial Constraints

For each semantic type subset, when loss sample size ≥ 3 , optimal trigger thresholds were determined via grid search on the training set with accuracy maximization as the objective function; when sample size < 3 , the subset was marked "non-triggerable," and only descriptive statistics were reported. Candidate indicators for threshold search were key meteorological indicators determined in Section 3.5; candidate quantiles were 10%, 20%, 80%, and 90%, with search ranges automatically adjusted based on subset sample size constraints.

Regarding actuarial constraints, conditional trigger mechanisms must simultaneously satisfy: (1) Trigger Probability Reasonableness: Theoretical trigger probability for each type should be broadly consistent with its historical occurrence frequency, avoiding excessively high trigger probabilities leading to unaffordable premiums or excessively low probabilities leading to inadequate protection; (2) Payout Fairness: Pure premiums across disaster types (pure premium = trigger probability $\times$ agreed payout amount) should reflect true risk gradients, with higher-risk types corresponding to higher premiums; (3) Controllable Basis Risk: False trigger rates of conditional triggers (proportion of non-loss samples incorrectly triggered) should be significantly lower than global threshold baselines.

4.8.2 Generalization Assessment Strategy

Generalization assessment of conditional trigger mechanisms adopted the same training–test temporal split scheme as Section 3.6 but added decomposition analysis at the disaster type level: (1) Type-Level Assessment: Conditional trigger rate, loss coverage, and false trigger rate were calculated separately for each semantic type on the test set; (2) Aggregate-Level Assessment: Trigger results across all types were aggregated to calculate overall F1 score, which was paired-compared with global baselines from Section 3.6; (3) Actuarial-Level Assessment: Actuarially fair premiums for each type were calculated to assess whether premium differentials aligned with disaster intensity gradients. For disaster types with insufficient test-set loss samples ($n < 2$), threshold search and generalization assessment were not performed; only training-set descriptive statistics and physical rule coverage were reported. Objective constraints of the "small-sample ceiling" and alternative risk transfer solutions (e.g., reinsurance or parametric catastrophe bonds) are analyzed in the Discussion section..

Declarations

Funding:

This work was supported by Heavy Rain and Drought-Flood Disasters in Plateau and Basin Key Laboratory of Sichuan Province, under the project titled "Coupled Flood Forecasting for Small Catchments Based on Deep Learning and the HEC-HMS Model" (Award Number. SCQXKJYJXMS202516), and additionally by the Young Innovation Team Project "Research on Artificial Intelligence-Based Gridded Element Forecasting Techniques" (Award Number. SCQXQNCXTD202401).

Author Contribution

Zichuan Sun and Fengliang Zou conceived the study and designed the semantic disaster classification framework. Wulong Zhang collected and curated the daily meteorological observations and county-level yield data from 2014 to 2024. Gao Li performed the Spearman correlation analysis, random forest modeling, and EPV-adaptive attribution diagnosis. Kaichuan Xie developed the conditional trigger threshold optimization algorithms and conducted the multi-benchmark comparative evaluation. Weimin Xu led the insurance actuarial assessment and product design recommendations. Zichuan Sun and Fengliang Zou wrote the main manuscript text. Liangxun Chen and Jun Yang prepared Figures 1–2 and Tables 2–6. All authors reviewed, edited, and approved the final manuscript.

Data Availability

The crop yield data and specific datasets generated or analyzed during this study are not publicly available due to [privacy restrictions / third-party agreements] but are available from the corresponding author upon reasonable request. The daily meteorological data used in this research were obtained from the National Meteorological Science Data Center of China and can be accessed by applying through their official portal at <https://data.cma.cn/>.

References

1. Abdi M J, Zulkafli Z, Raffar N ,et al (2025) Basis risk reduction in weather index insurance for rice production in North-West Malaysia[J].Paddy and Water Environment 23(1):129–138
2. Chen Y, Xue Z Y, Wang T et al (2024) Research on Insurance Design and Pricing of Comprehensive Remote Sensing Index of Shaanxi Winter Wheat at Kilometer Grid Scale[J].Chinese Journal of Agrometeorology 45(10):1204–1215
3. China Meteorological Administration (2019) Monitoring indices of rainy season in China Autumn rain of West China[S] Beijing: China Meteorological Press

4. Clement KY, Wouter Botzen WJ, Brouwer R, et al (2018) A global review of the impact of basis risk on the functioning of and demand for index insurance. *Int J Disaster Risk Reduct* 28(January):845–853
5. Ding Y H, Wang Z Y (2008) A study of rainy seasons in China[J]. *Meteorology and Atmospheric Physics* 16(3) : 121–138
6. Friedman J H (2001) Greedy Function Approximation: A Gradient Boosting Machine[J].*Annals of Statistics* 29(5):1189–1232
7. Gao Y X, Guo Q Y (1958) China's autumn rainfall phenomenon[J].*Acta Meteorologica Sinica* 29(4):264–273
8. Heinze G, Schemper M. A solution to the problem of separation in logistic regression[J]. *Statistics in Medicine*, 2002, 21(16):2409–2419. DOI:10.1002/sim.1047.
9. Hill R V, Hoddinott J, Kumar N (2013) Adoption of weather index insurance: Learning from willingness to pay among a panel of households in rural Ethiopia[J].*Agricultural Economic*, 44(4–5):385–398
10. Jaegermeyr J, Mueller C, Ruane A ,et al (2021) Climate impacts on global agriculture emerge earlier in new generation of climate and crop models[J]. 2(11):873–885
11. Li Y J, Shi X H, Wen M (2025) Interdecadal variation of uneven temporal distribution of the "West China autumn rain" and related rain structure changes[J].*Acta Meteorologica Sinica* 83(2):257–269
12. Li Y N, Li Y Q (2024) Progress of the Evolutionary Characteristics of Autumn Rain in Western China and Its Anomalous Mechanism for the Past 20 Years[J].*Plateau Meteorology* 43(1):1–15
13. Liu L F, Li X H, Yan J F, et al (2025) Analysis of Meteorological Factors during the Pre-Flood Season and Their Impact on Stable and Increased Early Rice Yields in Taishan[J].*Guangdong Meteorology* 47(2) :85–88
14. Lobell D B, Burke M B (2010) On the use of statistical models to predict crop yield responses to climate change[J].*Agricultural & Forest Meteorology* 150(11):1443–1452
15. Lundberg S M, Erion G, Chen H ,et al (2026) From local explanations to global understanding with explainable AI for trees[J].*Nature Machine Intelligence*[2026-01-06]
16. Luo X, Li D L, Wang H (2013) New Evolution Features of Autumn Rainfall in West China and Its Responses to Atmospheric Circulation[J].*Plateau Meteorology* 32(4):1019–1031
17. Ma X L, Zhou Y B, Yuan J et al (2024) Variation Characteristics and Correlation Analysis of Meteorological Elements During Maize Growth Period in Yangquan City[J].*Chinese Agricultural Science Bulletin* 40(7):118–122
18. Merz B, Kreibich H, Schwarze R ,et al (2010) Review article "Assessment of economic flood damage" [J].*Natural Hazards & Earth System Science* 10(8):1697–1724.
19. MIRANDA M J, FARRIN K (2012) Index insurance for developing countries[J]. *Applied Economic Perspectives and Policy* 34(3): 391–427

20. Nieto JD, Cook SE, Läderach P, et al (2010) Rainfall index insurance to help smallholder farmers manage drought risk. *Climate Dev* 2(3):233–247
21. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. [J]. *Journal of Clinical Epidemiology*, 1996, 49(12):1373–1379. DOI:10.1016/S0895-4356(96)00236-3.
22. Ray D K, Gerber J S, Macdonald G K, et al (2015) Climate variation explains a third of global crop yield variability[J]. *Nature Communications* 6(1):5989.
23. Ren Y F, Chen S N, Zhao Y X (2025) Design and Evaluation on Meteorological Index Insurance Product of Rice High Temperature Heat Damage in Jiangsu Province[J]. *Chinese Journal of Agrometeorology* 46(01):1–13
24. Schlenker W, Roberts M J, Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change[J]. *Proc. Natl. Acad. Sci. U.S.A.* 2009, 106 (37) 15594–15598, <https://doi.org/10.1073/pnas.0906865106> .
25. Skees J S P (2001) Developing rainfall-based index insurance in Morocco[J]. *Policy Research Working Paper*, 27(1):1–32
26. Song Y (2019) Research on the Design of Weather-Index Insurance Products -Taking the Maize Weather-Index Insurance in Liaoning Province as an Example[D]. LIAONING UNIVERSITY
27. Tang H Y, Wu Y, Dong X Y, et al (2025) Extreme Precipitation Events and Their Circulation Characteristics During the Autumn Rain Period in West China[J]. *METEOROLOGICAL MONTHLY* 51(6):711–723
28. Troy T J, Kipgen C, Pal I (2015) The impact of climate extremes and irrigation on US crop yields[J]. *Environmental Research Letters* 10(5):054013
29. Wang L M, Wang L Z, Wang C Y, et al (2010) Rice Cooling Injury in Heilongjiang Province—Enlightenment of Study on Sterile-type Cooling Injury in Cold Region[J]. *HEILONGJIAN AGRICULTURAL SCIENCES* (02):20–21
30. Wassmann R, Jagadish S. V. K., Heuer S, et al (2009) Climate change affecting rice production: the physiological and agronomic basis for possible adaptation strategies[J]. *Advances in Agronomy* 101:59–122
31. Zeng C (2021) Classification and Regionalization of Geomorphological Types Based on Typical Terrain Indicators and Landform Unit for Sichuan Province, China[J]. *Mountain Research* (4):587–599
32. Zhang J Q (2025) Research on Multi-Hazard and Multi-Scale Agrometeorological Disaster Risk Assessment and Projection under the Context of Climate Change[M]. Bei Jing :SCIENCE PRESS
33. Zhang Y C, He W X (1991) An Introduction to Agrometeorological Disasters in China [M]. China Meteorological Press
34. Zhao Y X, Chen S N (2023) Progress in the Theory of Weather Risk Management and Its Practice[J]. *Advances in Meteorological Science and Technology* 13(2):26–37

35. Zhao Y, Qin N S, Li J J, et al (2015) Correlation Analysis of Rice Yield Forming Factors and Meteorological Elements in Sichuan Basin[J].Southwest China Journal of Agricultural Sciences 28(6):2343–2347

36. Zhu X G, Long S P, Ort D R (2010) Improving Photosynthetic Efficiency for Greater Yield[J].Annual Review of Plant Biology 61(1):235–261

37. Zscheischler J, Westra S ,van den Hurk, et al (2018) Future climate risk from compound events[J].Nature Climate Change 8(6):469–477

38. Zulkafli Z, Muharam FM, Raffar N, et al (2021) Contrasting influences of seasonal and intra-seasonal hydroclimatic variabilities on the irrigated rice paddies of Northern Peninsular Malaysia for weather index insurance design. Sustainability 13(9):5207

Figures

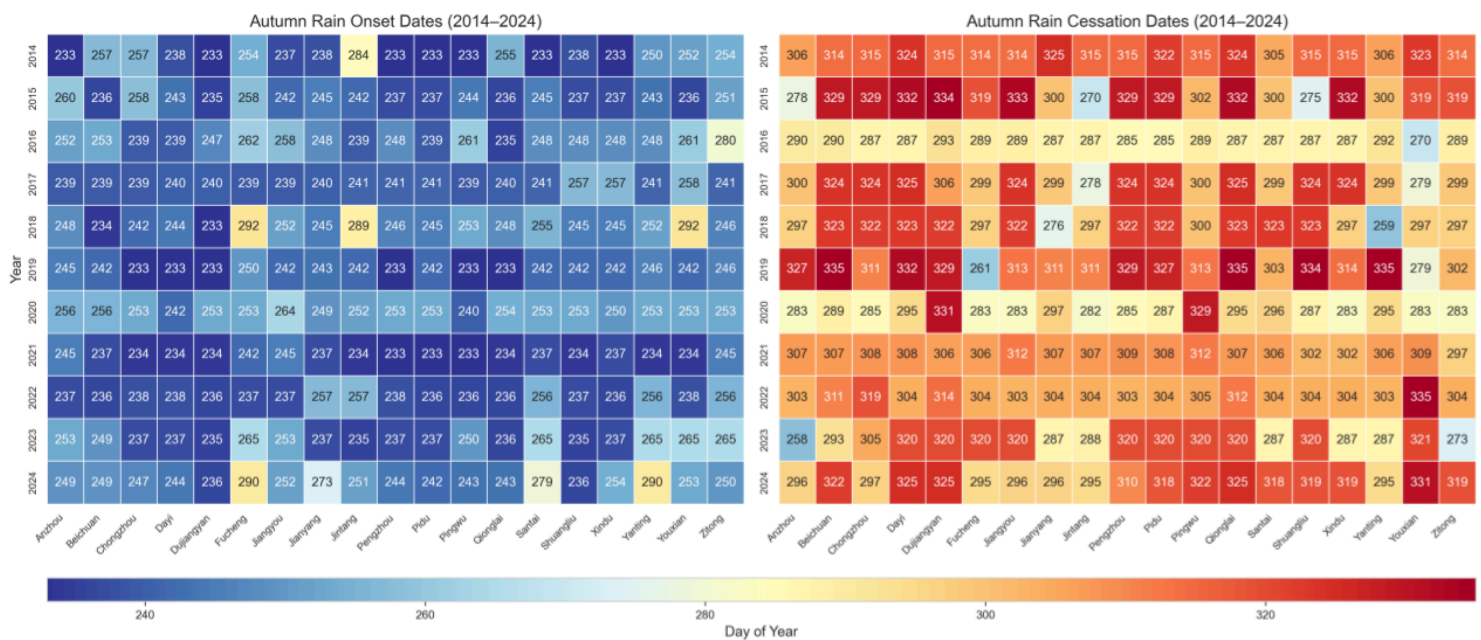
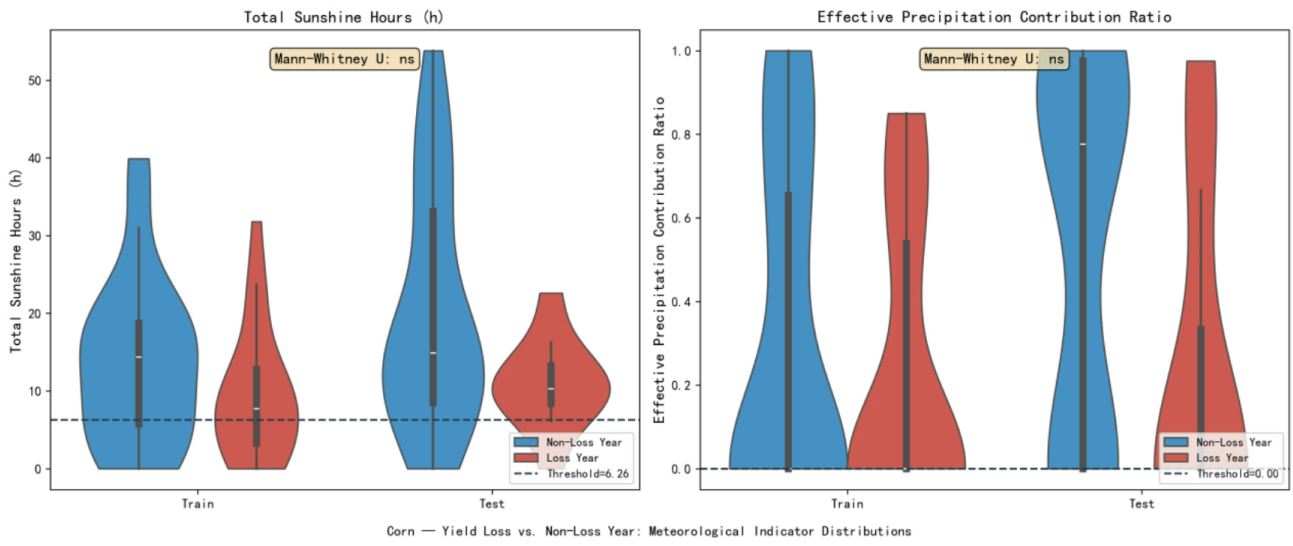


Figure 1

Heatmap of the Start and End Dates of Autumn Rainfall

Rice — Yield Loss vs. Non-Loss Year: Meteorological Indicator Distributions



Corn — Yield Loss vs. Non-Loss Year: Meteorological Indicator Distributions

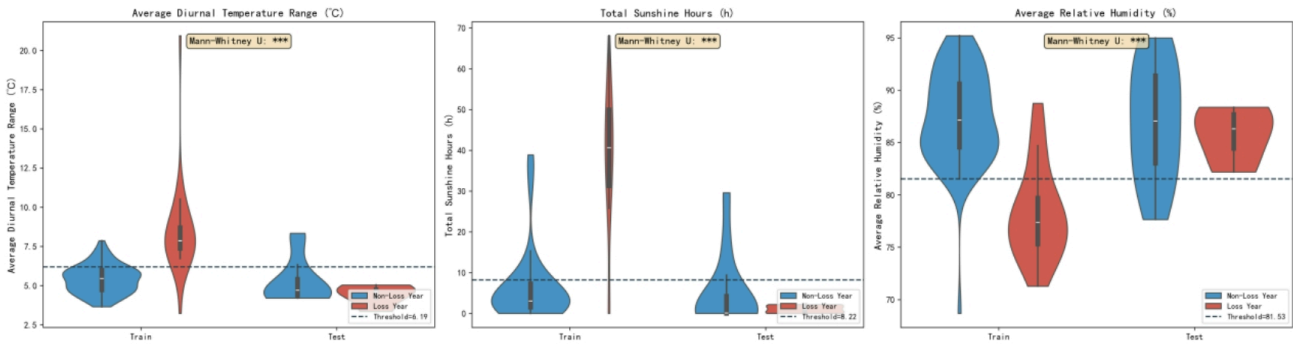


Figure 2

Comparison of key meteorological indicator distributions between yield-loss and non-yield-loss years and diagnostic of global threshold segmentation performance.