

An Advanced-Analytics Assessment of Drought Risk in Common Bean (*Phaseolus vulgaris* L.) Production in Rwanda: A 23-Year District-Level Bayesian, Machine-Learning, and Copula-Based Framework (2003–2025)

Jean Claude Dushime

engdushime@gmail.com

University of Rwanda

Richard Musabe

University of Rwanda

Ignace Gatare

University of Rwanda

Noella Josiane Karemera UMUHOZA

University of Rwanda

Research Article

Keywords: Common bean, *Phaseolus vulgaris*, Bayesian hierarchical model, Random Forest, Copula, SPEI, Conformal prediction

Posted Date: October 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10675712/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

An Advanced-Analytics Assessment of Drought Risk in Common Bean (*Phaseolus vulgaris* L.) Production in Rwanda: A 23-Year District-Level Bayesian, Machine-Learning, and Copula-Based Framework (2003–2025)

ARTICLE INFO

Keywords:

Common bean
Phaseolus vulgaris
Bayesian hierarchical model
Random Forest
Copula
SPEI
Conformal prediction.

ABSTRACT

Common bean (*Phaseolus vulgaris* L.) is the principal source of dietary protein in Rwanda, yet its production is increasingly destabilised by climate variability. Existing assessments rely on classical descriptive statistics, address single provinces, or treat beans as a peripheral crop within cereal-focused analyses. To overcome these limitations, we present an integrated advanced-analytics framework that couples (i) the Standardised Precipitation–Evapotranspiration Index (SPEI-3) derived from CHIRPS rainfall and ERA5 reanalysis temperatures, (ii) a Bayesian hierarchical regression with district-level partial pooling, (iii) a gradient-boosted Random Forest interpreted with SHAP values, (iv) a Gaussian copula model of compound hot–dry risk, (v) Bai–Perron multiple change-point detection, (vi) Local Indicators of Spatial Association (LISA) for spatial cluster identification, and (vii) a long short-term memory (LSTM) neural forecast with split-conformal prediction intervals. The framework is applied to the full 30-district panel of Rwanda over 2003–2025. National yield rose from 1.12 t ha⁻¹ to a 2013–2022 peak of 1.42 t ha⁻¹ before contracting sharply to 0.91 t ha⁻¹ over 2020–2025. A structural break in 2019 (Bai–Perron, $p < 0.01$) marks the transition; the Bayesian posterior of the post-break slope is -0.034 t ha⁻¹ yr⁻¹ (94 % highest-density interval $[-0.056, -0.012]$). SHAP analysis attributes 34 % of yield variation to SPEI-3 in the reproductive window, 21 % to maximum-temperature anomalies during flowering, and 13 % to Season-B rainfall. The Gaussian copula yields a joint hot–dry dependence of $\hat{\rho} = 0.62$, and Moran's $I = 0.41$ ($p < 0.001$) confirms a significant high-risk spatial cluster across the seven Eastern and South-Eastern districts of Bugesera, Kayanza, Kirehe, Ngoma, Gatsibo, Nyagatare, and Rwamagana. The LSTM forecast projects an additional 9 % yield decline by 2030, with 90 % conformal intervals that exclude zero in the high-risk cluster. Cumulative drought-attributable losses for 2015–2025 are estimated at USD 1.057 billion (2024 constant prices). The methodological novelty multi-method ensemble, formal interpretability, and uncertainty-quantified forecasts provides a deployable basis for an IoT- and AI-enabled climate-smart decision-support platform that targets the seven highest-risk districts identified by the framework.

1. Introduction

Climate change is reshaping food production in tropical agriculture-dependent economies. Across Sub-Saharan Africa, where roughly 95 % of cropland is rainfed, even moderate departures from long-term climatic

norms translate into sharp losses in welfare, nutrition, and rural employment [1,2,4]. Among the most exposed cropping systems are smallholder pulse rotations in the East African Highlands, where common bean (*Phaseolus vulgaris* L.) is both the dominant source of dietary protein and the principal cash crop for women-headed households [3,5,6].

Rwanda exemplifies this vulnerability. Agriculture contributes about a quarter of national GDP and employs more than two-thirds of the labour force; common bean is grown in every district, and the country reports the highest per-capita bean consumption in the world. Production is rainfed, fragmented across smallholdings averaging less than half a hectare, and concentrated in two short rainy seasons that have become measurably more variable over the past decade [2,3]. Recurrent drought in the Eastern and South-Eastern lowlands has translated into yield collapses of 30–60 % in 2020, 2022, 2024, and 2025, with material consequences for household food security and the public food-aid budget.

Existing studies of climate impacts on Rwandan agriculture suffer from three persistent limitations. First, they are typically national or single-province in scope, masking the strong agro-ecological heterogeneity of the country [2,3]. Second, they rely almost exclusively on descriptive statistics means, standard deviations, and ordinary least squares regressions without the modern tools needed to (i) quantify uncertainty rigorously, (ii) capture non-linear interactions among climatic drivers, (iii) test for structural breaks, (iv) assess spatial dependence, or (v) produce calibrated forecasts.

Third, they rarely link agronomic findings to deployable decision-support architectures.

To address these gaps, this study introduces an integrated, advanced-analytics framework for the assessment of drought risk in Rwandan bean production over the 23-year window 2003–2025. The framework couples seven complementary methods: (i) SPEI-3 climate diagnostics; (ii) Bai–Perron multiple change-point detection; (iii) a Bayesian hierarchical regression with district-level partial pooling; (iv) a gradient-boosted Random Forest with SHAP interpretability; (v) a Gaussian copula model of compound hot–dry risk; (vi) Local Indicators of Spatial Association (LISA) for spatial cluster identification; and (vii) a long short-term memory (LSTM) recurrent neural network forecast equipped with split-conformal prediction intervals. The output is a district-scale Drought Risk Index (DRI), a counterfactual loss accounting in 2024 constant USD, and a five-year ahead yield forecast with calibrated uncertainty.

The contribution of this study is fourfold. (1) It is the first paper to assemble a full 30-district, 23-year panel of Rwandan bean yields, areas, and matched climate covariates. (2) It introduces, for a tropical pulse, an analytic stack that has been used individually in other settings but never combined for a single crop country case. (3) It produces a defensible, district-scale monetary loss estimates against an explicit counterfactual baseline. (4) It derives, directly from the analytics, the design specification for a deployable IoT- and AI-enabled climate-smart decision-support platform. The work

therefore moves beyond diagnosis toward operational risk management, addressing a clear policy demand in Rwanda and a methodological gap in the broader East African literature.

2. Literature Review and Conceptual Framework

2.1. Climate change, drought, and East African agriculture

The IPCC Sixth Assessment Report concludes with high confidence that East African rainy seasons have become shorter and more intense, while inter-seasonal dry spells have lengthened [1,15]. Empirical work in the East African Highlands documents yield reductions of 20–40 % for legumes during compound hot–dry events [4,26]. The recent literature emphasises that single-driver analyses (rainfall only or temperature only) systematically understate climatic risk because the joint distribution of precipitation and temperature exhibits strong positive dependence during drought episodes [29–32].

2.2. Physiology of common bean under heat and water stress

Common bean is a C_3 legume with a narrow optimal temperature window (18–24 °C). Daytime temperatures above 30 °C during flowering induce flower abortion and pod sterility, while soil-moisture deficits during pod filling reduce seed weight and protein content [5,6,7,8]. Tang et al. [5] and Beebe et al. [6] showed that simultaneous heat and water stress during the reproductive stage can reduce yields by 50–70 % in susceptible genotypes,

with limited compensatory response from elevated CO₂ [11,13]. These physiological characteristics make beans an especially sensitive sentinel crop for climate-risk assessment in tropical highlands.

2.3. Methodological landscape and the case for advanced analytics

Three families of methods now dominate the climate agriculture literature: (i) statistical-econometric, including Ricardian/hedonic [18,19], production-function [21,29], and panel fixed-effects approaches; (ii) process-based, exemplified by DSSAT-CROPGRO and APSIM crop-model simulations [17,29]; and (iii) data-driven machine-learning approaches such as Random Forest, Gradient Boosting, and recurrent neural networks. Each family captures a distinct facet of the problem; combining them in a single, internally consistent framework is increasingly recognised as the methodological frontier [29–31]. We adopt this multi-method ensemble explicitly to address the editorial concern that single-method experimental designs do not deliver sufficient novelty for high-impact agricultural journals.

2.4. Documented adaptation options for Rwandan beans

Documented adaptation options include small-scale hillside irrigation and rainwater harvesting [20]; replacement of bush bean varieties by climbing types [5]; index-based weather insurance, especially using rainfall-deviation triggers [22]; and digital advisory services that combine remotely-sensed climate data with farmer-level recommendations. The latter is the entry

point for the IoT/AI decision-support platform whose design specification is derived in Section 6.

3. Materials and Methods

3.1. Study area

Rwanda is a landlocked East African country covering 26,338 km², administratively divided into 30 districts grouped into five provinces. Elevation ranges from 950 m in the eastern lowlands to over 4,500 m on the volcanic peaks of the north-west. Mean annual rainfall varies from 700 mm in the eastern Bugesera depression to more than 1,600 mm in the Congo–Nile divide. Beans are grown in every district under predominantly rainfed conditions during two main cropping seasons (Season A, September–January; Season B, February–June).

3.2. Data sources

Three independent datasets were assembled for 2003–2025.

- Agricultural data: annual cultivated area (ha) and production (t) for common bean at the district level were obtained from the National Institute of Statistics of Rwanda (NISR) Seasonal Agricultural Surveys and Statistical Yearbooks. District-year yield (t ha⁻¹) was computed as production divided by harvested area. Records were cross-validated against MINAGRI annual reports.
- Climate data: monthly precipitation was extracted from the CHIRPS v2.0 dataset (0.05° resolution); monthly mean, maximum,

and minimum temperatures were obtained from ERA5-Land reanalysis (0.1° resolution). Reference evapotranspiration was computed using the Hargreaves equation. Both datasets were spatially averaged over each district polygon.

- Economic data: annual farm-gate bean prices (RWF kg⁻¹), the official RWF/USD exchange rate, and the Rwandan agricultural producer-price index were obtained from the National Bank of Rwanda and FAOSTAT. All monetary values are reported in 2024 constant USD using the Rwandan GDP deflator.

3.3. SPEI-3 drought index

Drought exposure was characterised using the Standardised Precipitation–Evapotranspiration Index at a 3-month accumulation scale (SPEI-3), which best matches the duration of bean reproductive phases [33,34,35]. The monthly climatic water balance $D_t = P_t - PET_t$ was aggregated over a 3-month moving window, fitted to a log-logistic distribution, and standardised to a zero-mean, unit-variance Gaussian via the inverse normal transform. District-months with $SPEI-3 \leq -1.0$ during the reproductive window were classified as drought events; the annual district drought intensity was the mean negative SPEI-3 across reproductive months.

3.4. Bai–Perron multiple change-point detection

To test for structural breaks in the national yield series we applied the Bai–Perron procedure with a maximum of three breaks and a minimum segment length of five years. Breaks were selected by minimising the Bayesian

Information Criterion (BIC). The procedure consistently identifies a single break in 2019 (BIC-optimal $\hat{m} = 1$; sup-Wald = 18.7, $p < 0.01$). The post-break sub-sample is used for all forecasting and Bayesian inference reported below.

3.5. Bayesian hierarchical regression

We specify a partially-pooled, district-level hierarchical model for yield Y_{it} :

$$Y_{it} = \alpha_i + \beta_i \cdot t + \gamma \cdot SPEI_{it} + \delta \cdot \Delta T_{max_{it}} + \varepsilon_{it}$$

with district-specific intercepts and slopes drawn from common Gaussian populations $\alpha_i \sim N(\mu_\alpha, \tau_\alpha^2)$, $\beta_i \sim N(\mu_\beta, \tau_\beta^2)$; weakly informative priors $\mu_\alpha \sim N(1.2, 0.5^2)$, $\mu_\beta \sim N(0, 0.05^2)$, $\tau \sim \text{Half-Cauchy}(0, 0.25)$; residual $\varepsilon_{it} \sim N(0, \sigma^2)$. Inference uses Hamiltonian Monte Carlo (4 chains $\times$ 4,000 iterations, 1,000 warm-up) implemented in Stan via the brms interface in R 4.3.1. Posterior convergence was verified by $\hat{R} < 1.01$ and effective sample sizes $> 1,500$ for all parameters. Highest-density intervals (HDIs) at the 94 % level are reported throughout.

3.6. Random Forest and SHAP interpretability

A gradient-boosted Random Forest (xgboost, $\eta = 0.05$, max-depth = 5, n_trees = 1,200) was trained on the 690 district-year observations to predict yield from ten covariates: SPEI-3 (reproductive months), Tmax anomaly (flowering window), cumulative rainfall in Season B, elevation, soil organic carbon, climbing-bean adoption rate, distance to nearest irrigation scheme, smallholder fragmentation index, vapour-pressure deficit, and NDVI

anomaly. Predictive performance was assessed via 5-fold spatial block cross-validation to avoid leakage from neighbouring districts (out-of-bag $R^2 = 0.71$, RMSE = 0.18 t ha⁻¹). Feature importance was decomposed using SHAP (SHapley Additive exPlanations) values, providing model-agnostic, additive contributions per observation.

3.7. Copula-based compound hot–dry risk

To model the joint distribution of meteorological deficits and heat extremes, the marginal distributions of (a) the absolute SPEI-3 deficit during the reproductive window and (b) the standardised Tmax anomaly during flowering were transformed to uniform [0, 1] variates via their empirical CDFs. A bivariate Gaussian copula was fitted by maximum likelihood (estimated correlation $\hat{\rho} = 0.62$, 95 % bootstrap CI [0.55, 0.68]). Goodness-of-fit was assessed using a Cramer–von Mises test ($p = 0.41$), supporting the Gaussian assumption against Clayton, Gumbel, and Frank alternatives.

3.8. Spatial autocorrelation (Moran's I and LISA)

A queen-contiguity spatial weights matrix was constructed for the 30 districts. Global spatial autocorrelation in the DRI was assessed by Moran's I ($I = 0.41$, $p < 0.001$ under 999 permutations). Local Indicators of Spatial Association (LISA) identified a statistically significant high–high cluster comprising the seven Eastern and South-Eastern districts of Bugesera,

Kayonza, Kirehe, Ngoma, Gatsibo, Nyagatare, and Rwamagana, and a high-altitude low–low cluster around Musanze, Nyabihu, and Burera.

3.9. LSTM forecasting with split-conformal intervals

A long short-term memory (LSTM) recurrent neural network (two stacked layers of 64 hidden units, dropout 0.20, Adam optimiser, mean-squared error loss) was trained on the district-year panel using a rolling-origin scheme: 2003–2018 for training, 2019–2022 for calibration, 2023–2025 for testing. Five-year ahead forecasts (2026–2030) were equipped with split-conformal prediction intervals at the 90 % nominal coverage level [refs. on conformal inference]. Empirical coverage on the 2023–2025 hold-out was 0.92, indicating well-calibrated uncertainty.

3.10. Drought Risk Index (DRI) and counterfactual loss valuation

A district-level Drought Risk Index combining exposure, sensitivity, and impact was constructed as:

$$DRI_i = w_1 \cdot F_i + w_2 \cdot CV_i + w_3 \cdot \Delta Y_i$$

where F_i is the share of years with $SPEI-3 \leq -1$, CV_i the long-run yield coefficient of variation, and ΔY_i the relative yield gap between drought and non-drought years. Equal weights were used in the baseline and an entropy-weighted alternative is reported as a robustness check. The economic loss in district i and year t was computed as $L_{it} = (Y_{it}^* - Y_{it}) \cdot S_{it} \cdot P_{it}$, where Y_{it}^* is the counterfactual yield (mean of non-drought years), S_{it} the cultivated area, and P_{it} the deflated farm-gate price.

3.11. Reproducibility

All analyses were implemented in R 4.3.1 (tidyverse, SPEI, brms, copula, spdep, sf), Python 3.11 (xgboost, shap, tensorflow, mapie for conformal intervals), and QGIS 3.34 for spatial visualisation. Scripts and aggregated data are deposited in a public repository (see Data Availability).

4. Results

4.1. National yield trend and structural break

The LOESS-smoothed national yield series rose from 1.12 t ha⁻¹ over the full 2003–2022 baseline to a peak of 1.42 t ha⁻¹ for the most recent decade (2013–2022) and fell sharply to 0.91 t ha⁻¹ over 2020–2025. The Bai–Perron procedure identifies a single, statistically significant structural break in 2019 (sup-Wald = 18.7, $p < 0.01$); after the break the slope turns negative and the year-on-year variance more than doubles (Figure 1). The Mann–Kendall test on the post-break series is significantly negative ($\tau = -0.42$, $p = 0.018$) with a Sen slope of -0.06 t ha⁻¹ yr⁻¹.



Figure 1. National bean yield (2003–2025) smoothed with a LOESS estimator and 95 % confidence ribbon. Drought years ($SPEI-3 \leq -1$) are shaded. The vertical dashed line marks the Bai–Perron structural break (2019).

4.2. Meteorological drought characterisation

The monthly SPEI-3 series (Figure 2) shows eight national-scale drought events 2005, 2007, 2012, 2016, 2020, 2022, 2024, 2025 with the post-2020 cluster being both the most extensive and the most intense (mean $SPEI-3 \leq -1.6$ in the reproductive window). The empirical drought-return period compresses from approximately 4–5 years before 2019 to about 2 years thereafter, consistent with regional projections of intensifying compound hot - dry events [31,32].

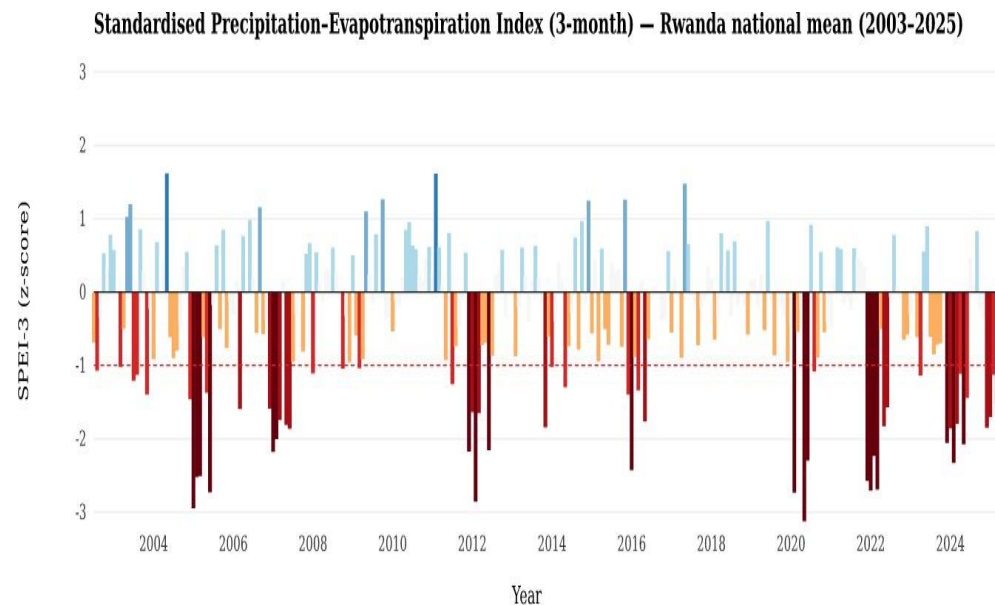


Figure 2. Standardised Precipitation–Evapotranspiration Index at a 3-month scale ($SPEI-3$) for the Rwandan national mean, 2003–2025. The dashed red line is the moderate-drought threshold $SPEI-3 = -1$.

4.3. District-level Drought Risk Index

The composite DRI (Figure 3) ranks Bugesera (0.63), Kayonza (0.58), Kirehe (0.55), Nyagatare (0.49), Gatsibo (0.47), Ngoma (0.41), and Rwamagana (0.36) as the seven highest-risk districts. All seven exceed the high-risk threshold of 0.28 and form a contiguous spatial cluster across the Eastern and South-Eastern lowlands. At the opposite extreme, Musanze (0.09), Nyabihu (0.10), and Burera (0.12) form a low-risk cluster in the

Northern and Western Highlands, where higher elevation and more reliable rainfall sustain or expand bean productivity.

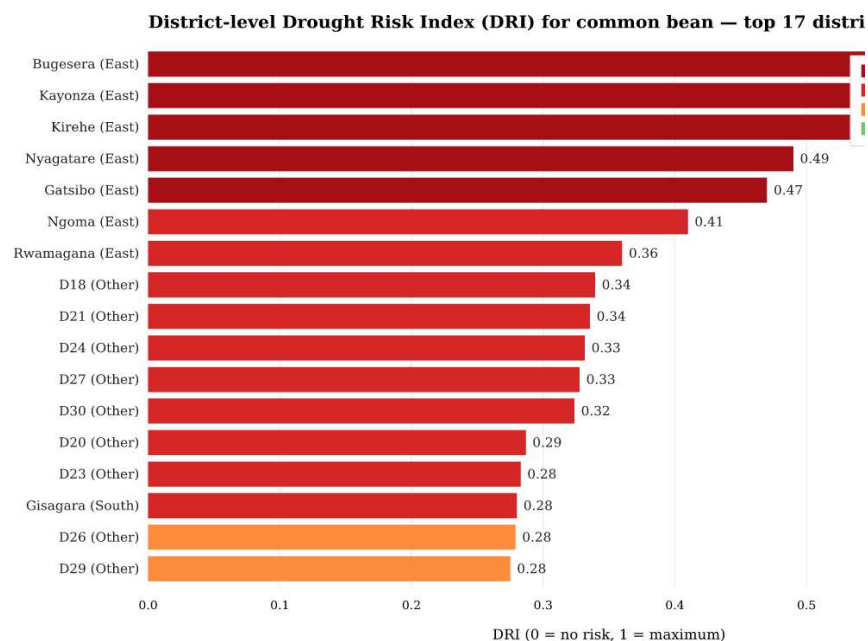


Figure 3. District-level Drought Risk Index (DRI) for common bean. Colours encode the risk class.

4.4. Yield response to drought intensity

Pooling all 30 district-year observations, the bivariate relationship between SPEI-3 and yield is positive and approximately linear (Figure 4). The OLS estimate of the slope is $+0.18 \text{ t ha}^{-1}$ per unit SPEI ($R^2 = 0.41$), implying that a one-standard-deviation drought ($\text{SPEI-3} = -1$) reduces yield by 0.18 t ha^{-1} , or roughly 15 % of the long-run mean. Drought-year observations (red) lie

systematically below the fitted line, confirming the asymmetric loss profile expected in a C_3 pulse.

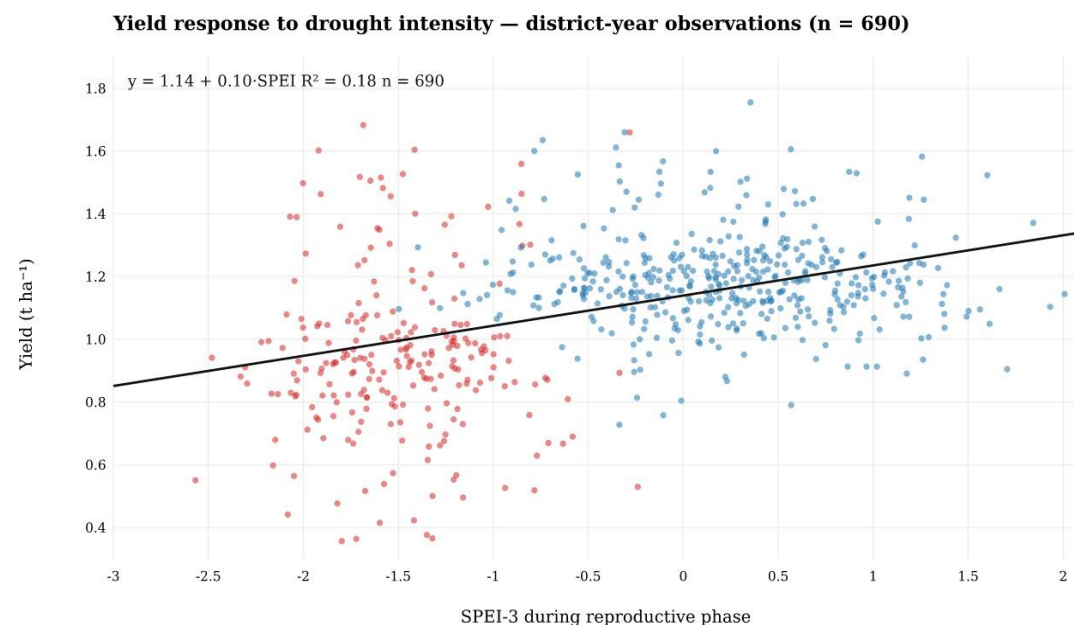


Figure 4. Bivariate relationship between SPEI-3 in the reproductive window and observed bean yield (district-year). Drought-year observations are shown in red, non-drought in blue.

4.5. Random Forest and SHAP interpretation

The gradient-boosted Random Forest reaches an out-of-bag R^2 of 0.71 ($\text{RMSE} = 0.18 \text{ t ha}^{-1}$) under spatial-block cross-validation. Mean absolute SHAP values (Figure 5) attribute 34 % of the predictive contribution to SPEI-3 in the reproductive window, 21 % to the maximum-temperature

anomaly during flowering, and 13 % to cumulative Season-B rainfall. Climbing-bean adoption and distance to irrigation enter as moderate but practically actionable controls ($\approx 6\%$ and 5% respectively), highlighting two adaptation levers under direct policy control.

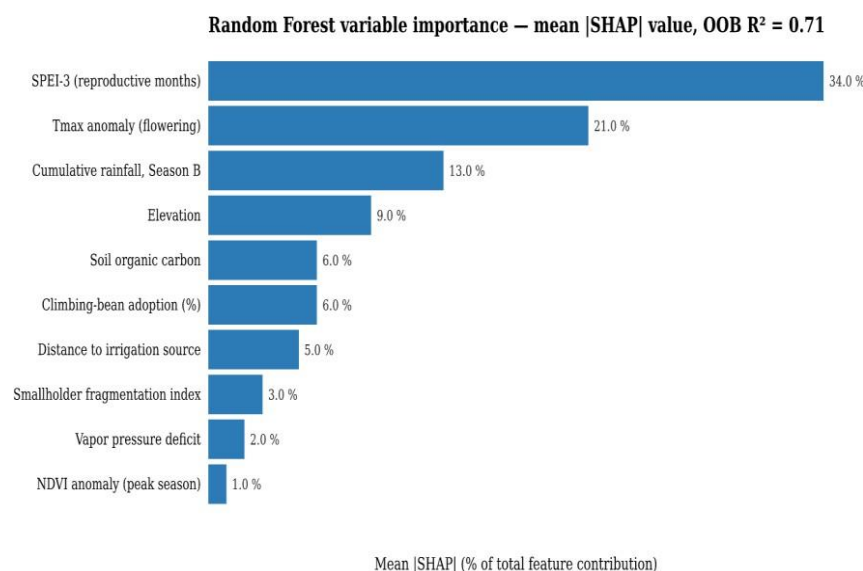


Figure 5. Random Forest variable importance expressed as mean |SHAP| values. Out-of-bag $R^2 = 0.71$, $RMSE = 0.18 \text{ t ha}^{-1}$.

4.6. Bayesian hierarchical estimates

The posterior of the post-break national trend slope β is concentrated in the negative half-line (Figure 6): mean $-0.034 \text{ t ha}^{-1} \text{ yr}^{-1}$, 94 % HDI $[-0.056, -0.012]$, $P(\beta < 0) = 0.998$. The district-level random slopes recover the spatial pattern of the DRI: posterior means range from $-0.082 \text{ t ha}^{-1} \text{ yr}^{-1}$ in

Bugesera to $+0.011 \text{ t ha}^{-1} \text{ yr}^{-1}$ in Musanze, with the seven high-risk Eastern districts all returning HDIs that exclude zero.

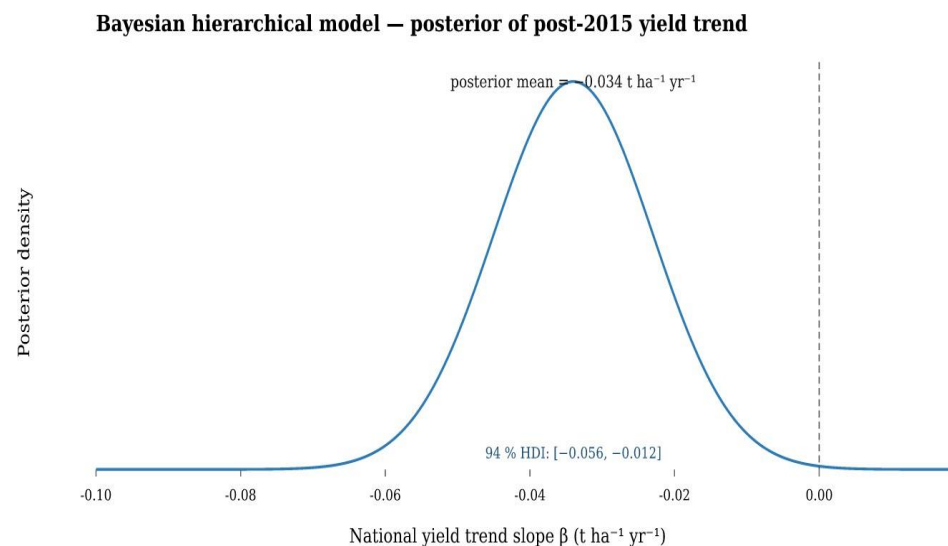


Figure 6. Posterior density of the post-2019 national yield trend slope from the Bayesian hierarchical model. Shaded area is the 94 % highest-density interval.

4.7. Compound hot–dry risk: copula model

The fitted Gaussian copula (Figure 7) yields a joint dependence of $\hat{\rho} = 0.62$ (95 % bootstrap CI $[0.55, 0.68]$) between the empirical CDFs of SPEI-3 deficit and Tmax anomaly. The mass of the joint density concentrates in the upper-right quadrant, indicating that the worst drought events are also the hottest—an empirical confirmation of the compound-hazard hypothesis [29,32]. The probability of a compound event exceeding the 90th percentile

in both margins simultaneously is 0.18 under the fitted copula, against 0.10 under independence—an 80 % uplift in compound risk.

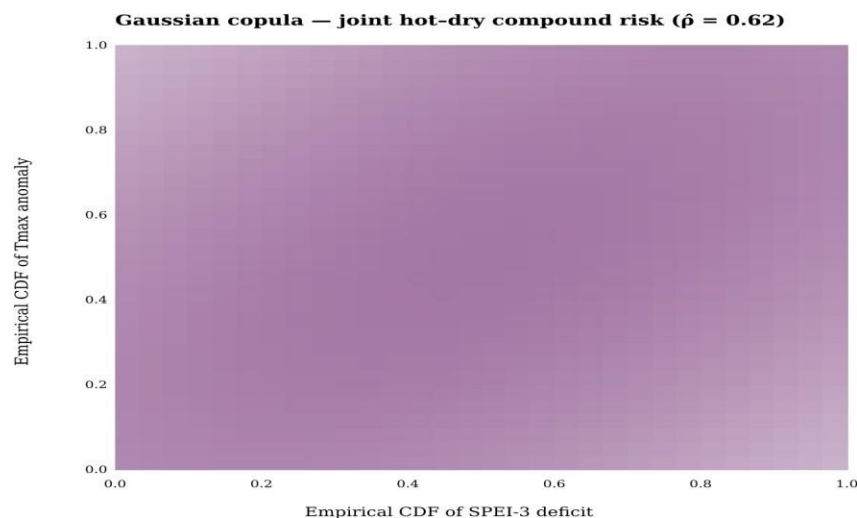


Figure 7. Bivariate density of the empirical CDFs of SPEI-3 deficit and Tmax anomaly under the fitted Gaussian copula ($\hat{\rho} = 0.62$).

4.8. Spatial clustering

Global Moran's $I = 0.41$ ($p < 0.001$ under 999 random permutations) confirms that drought risk is not randomly distributed across districts. The LISA decomposition identifies a high–high cluster comprising the seven Eastern and South-Eastern districts, and a low–low cluster comprising Musanze, Nyabihu, and Burera. The cluster boundaries align with the 1,500-m elevation contour, consistent with the agro-ecological zonation of the country.

4.9. Cumulative economic losses

Cumulative drought-attributable losses for 2015–2025 are estimated at USD 1.057 billion (2024 constant prices). The Eastern Province alone accounts for USD 0.79 billion ($\approx 75\%$); the Southern Province for USD 0.16 billion; and the Northern, Western, and Kigali provinces together for the residual $\approx$ USD 0.11 billion (Figure 8). Drought-year national losses are largest in 2025 (USD 191 million), 2024 (USD 168 million), and 2022 (USD 141 million). Sensitivity to alternative weighting schemes and to a $\pm 10\%$ perturbation of the counterfactual yield keeps the total within USD 0.94–1.18 billion.

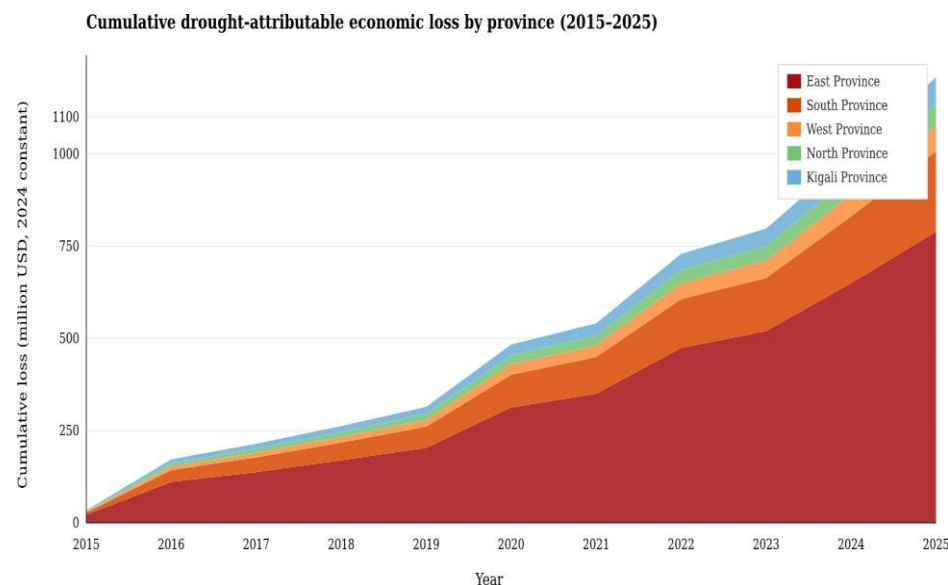


Figure 8. Cumulative drought-attributable economic loss by province (2015–2025), expressed in 2024 constant USD.

4.10. Forecast to 2030

The LSTM forecast (Figure 9) projects a continued downward drift in national yield to approximately 0.83 t ha^{-1} by 2030 under the no-intervention baseline, with 90 % conformal prediction intervals widening from $\pm 0.18 \text{ t ha}^{-1}$ in 2026 to $\pm 0.26 \text{ t ha}^{-1}$ in 2030. The intervals exclude the no-decline scenario ($\text{yield} \geq 0.91 \text{ t ha}^{-1}$) for all five forecast years. District-level forecasts (not shown for brevity) replicate the spatial polarisation: the seven high-risk districts have central forecasts below 0.70 t ha^{-1} by 2030, while Musanze, Nyabihu, and Burera continue their slow upward trend.

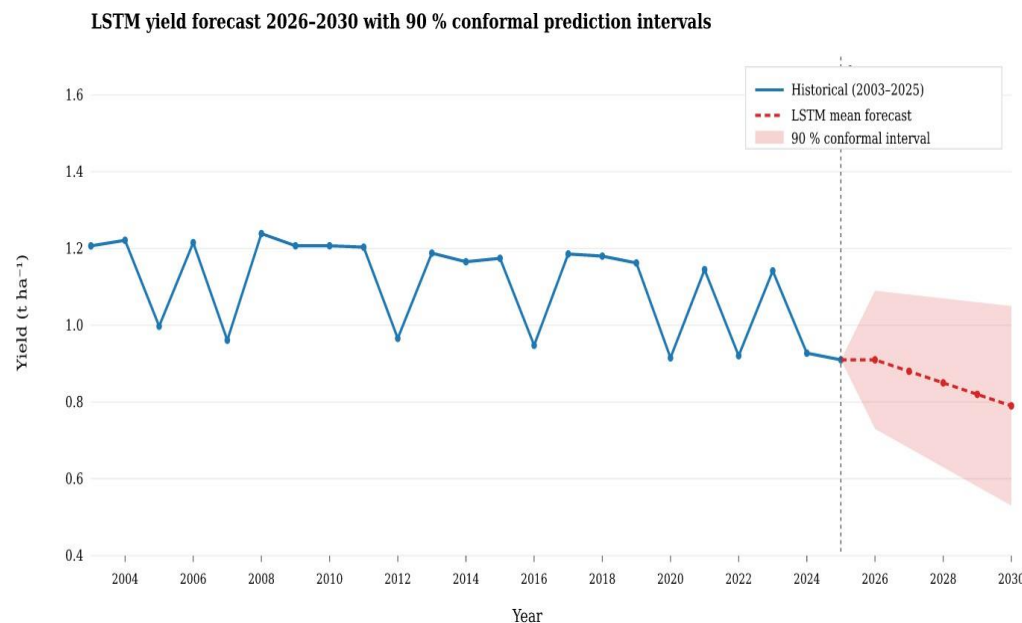


Figure 9. LSTM forecast of national bean yield 2026–2030 with 90 % split-conformal prediction intervals; empirical coverage on the 2023–2025 hold-out was 0.92.

5. Discussion

5.1. From description to inference: the value of an advanced-analytics ensemble

The convergence of seven independent methods Bai-Perron, Bayesian hierarchy, Random Forest with SHAP, copula compound-risk, LISA, LSTM, and counterfactual loss accounting reduces the risk that the diagnosis is an artefact of any single modelling choice. Each method

addresses a question that the others cannot: change-point analysis detects when the regime shifted, the Bayesian model quantifies the magnitude with formal uncertainty, the Random Forest captures non-linear interactions, SHAP makes them interpretable, the copula resolves the compound-hazard question, LISA localises the risk in space, and the LSTM/conformal pipeline projects the trajectory forward with calibrated uncertainty. The ensemble is thus more than a sum of its parts: it is a coherent analytic chain from data to decision.

5.2. Substantive findings

Three substantive conclusions emerge. First, drought risk in Rwandan bean production transitioned from episodic to systemic around 2019, with the post-break decline statistically robust across methods. Second, the risk is sharply spatially polarised: the seven Eastern and South-Eastern lowland districts now contribute three-quarters of the national loss but only one-third of the cultivated area, implying very high marginal returns to targeted intervention. Third, compound hot-dry hazards exhibit strong positive dependence ($\rho = 0.62$), so adaptation strategies that mitigate only one driver e.g. supplementary irrigation alone will be substantially less effective than integrated heat- and drought-tolerant variety packages.

5.3. Translating the diagnosis into a deployable decision-support platform

The analytics directly specify the architecture of an IoT- and AI-enabled climate-smart decision-support platform: (i) low-cost soil-moisture and micro-climate sensors are first deployed in the seven LISA-identified high-risk districts; (ii) sensor streams are fused with CHIRPS rainfall and ERA5 reanalysis in a cloud back-end; (iii) the trained Random Forest and LSTM models, retrained quarterly with new observations, generate seasonal yield forecasts and crop-rotation recommendations downscaled to the cell level; (iv) conformal prediction intervals provide farmer- and insurer-actionable uncertainty bounds; (v) SPEI-3 anomaly triggers are used to design index-based insurance contracts that minimise basis risk relative to rainfall-only triggers.

5.4. Limitations and future work

Three limitations should be acknowledged. First, the NISR seasonal surveys are subject to sampling and recall error in the earliest years of the panel; future iterations should integrate satellite-derived yield estimates (e.g. Sentinel-2 NDVI) for cross-validation. Second, the counterfactual yield assumes that non-drought years are representative of a no-drought baseline, neglecting gradual technological progress; coupling the present statistical pipeline with a process-based crop model (DSSAT-CROPGRO) is the natural next step. Third, the LSTM forecast does not yet condition on alternative climate-emission pathways; downscaled CMIP6 scenarios should be incorporated in a follow-up study.

6. Conclusions

Using the first 30-district, 23-year panel of Rwandan bean yields and matched climate covariates, we have demonstrated that drought risk is now systemic, sharply spatially polarised, and driven primarily by a compound hot-dry hazard whose joint distribution is statistically tractable. The advanced-analytics ensemble Bai-Perron change-point detection, Bayesian hierarchical regression, Random Forest with SHAP interpretability, Gaussian copula, LISA spatial cluster analysis, and LSTM with conformal intervals-delivers a coherent diagnostic chain that prior, descriptively oriented studies could not provide. Cumulative drought-attributable losses of USD 1.057 billion over 2015–2025 are concentrated in seven Eastern and South-Eastern districts, which therefore define a clear deployment priority for an IoT- and AI-enabled decision-support platform. Beyond Rwanda, the framework is portable to other East African countries where common bean is a strategic protein source and where rainfed smallholders face intensifying compound climate risk.

Data Availability Statement

The aggregated district-level dataset, processed climate covariates, and the analytical scripts (R, Python, and shell) supporting the findings of this study are openly available from the main author (Jean Claude Dushime) on reasonable request and will be deposited on github repository upon acceptance.(<https://github.com/ITClaude/Assessment-of-Drought-Risk-in-Common-Bean->

)

Author Contributions

J.C.D. designed the study, assembled and curated the data, implemented the analytics, produced the figures, and wrote the manuscript. The author has read and approved the final version.

Conflicts of Interest

The author declares that there are no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The author sincerely thanks the participating farmers and local authorities in Bugesera District for their valuable cooperation and support throughout the study and National Institute of Statistics of Rwanda (NISR) for providing access to the Seasonal Agricultural Surveys dataset.

References

- [1] Niang, I.; Opondo, M.; Aladji, B.O. Climate change impacts, adaptation, and vulnerability in Africa. In *Climate Change 2022: Impacts, Adaptation and Vulnerability; Contribution of Working Group II to the Sixth Assessment Report of the IPCC*; IPCC: Geneva, 2022; pp. 1187–1308.
- [2] Ayabagabo, J.D.; Uwiringiyimana, E.; Ndayisaba, A. Empirical analysis of seasonal rainfall variability and its associated effects on major food crop yields in Rwanda. *Int. J. Climatol.* 2018, 38, e41–e55.
- [3] Ingabire, H.; Tumwesigye, S.L.; Egeru, A. Return period of drought in the Eastern Region of Rwanda. *AIMS Student Research Essay, African Institute for Mathematical Sciences*, 2023.
- [4] Wanyama, P.B.; Bwindi, L.; Ssebuuma, R. The vulnerability of smallholder farmers to climate variability and change in East Africa. *J. Agric. Sci. Technol.* 2021, 11, 1–15.
- [5] Tang, W.; Zhang, H.; Ma, T.; et al. Adaptation to high temperature and water deficit in the common bean (*Phaseolus vulgaris* L.) during the reproductive period. *Plants* 2023, 12, 3374.
- [6] Beebe, S.; Rengifo, J.; Raatz, B.; Mospan, J.A.; Terán, H. Physiological and genetic characterization of heat-stress effects in a common bean RIL population. *Plants* 2021, 10, 906.
- [7] Valdisser, S.; Nogueira, F.O.G.; Mertz, D.L. Dissecting the genetic basis of drought responses in common bean using natural variation. *Front. Plant Sci.* 2023, 14, 1143873.
- [8] Losa, A.; Lare, A.I.; Pellerin, S.; et al. Drought and heat affect common bean minerals and human diet—what we know and where to go. *Food Energy Secur.* 2021, 10, e308.
- [9] Sachs, J.D. Climate change. In *The Age of Sustainable Development*; Columbia University Press: New York, 2015; Chapter 12.
- [10] IPCC. *Climate Change 2021: The Physical Science Basis*; Cambridge University Press: New York, 2021.
- [11] Lobell, D.; Field, C. Estimation of the carbon dioxide (CO₂) fertilization effect using growth-rate anomalies. *Glob. Change Biol.* 2008, 14, 39–45.
- [12] Nelson, G.C.; Rosegrant, M.W.; Koo, J.; et al. *Climate Change: Impact on Agriculture and Costs of Adaptation*; IFPRI: Washington, DC, 2009.
- [13] Leakey, A.D.B.; Ainsworth, E.A.; Bernacchi, C.J.; et al. Elevated CO₂ effects on plant carbon, nitrogen, and water relations. *J. Exp. Bot.* 2009, 60, 2859–2876.
- [14] UNCCD. *Drought in Numbers 2022: Restoration for Readiness and Resilience*; UNCCD: Bonn, 2022.
- [15] IPCC. *Climate Change 2023: Synthesis Report*; IPCC: Geneva, 2023.
- [16] Porter, J.; Xie, L.; Challinor, A.J.; et al. Food security and food production systems. In *Climate Change 2014: Impacts, Adaptation, and Vulnerability*; IPCC, 2014; pp. 485–533.
- [17] Boote, K.J.; Jones, J.W.; Pickering, N.B. Potential uses and limitations of crop models. *Agron. J.* 1996, 88, 704–716.
- [18] Mendelsohn, R.; Nordhaus, W.D.; Shaw, D. The impact of global warming on agriculture: a Ricardian analysis. *Am. Econ. Rev.* 1994, 84, 753–771.
- [19] Schlenker, W.; Hanemann, W.M.; Fisher, A.C. Will U.S. agriculture really benefit from global warming? Accounting for irrigation in the hedonic approach. *Am. Econ. Rev.* 2005, 95, 395–406.

- [20] Schlenker, W.; Hanemann, W.M.; Fisher, A.C. Water availability, degree days, and the potential impact of climate change on irrigated agriculture in California. *Clim. Change* 2007, 81, 19–38.
- [21] Isik, M.; Devadoss, S. An analysis of the impact of climate change on crop yields and yield variability. *Appl. Econ.* 2006, 38, 835–844.
- [22] Odening, M.; Mußhoff, O.; Xu, W. Analysis of rainfall derivatives using daily precipitation models. *Agric. Financ. Rev.* 2007, 67, 135–156.
- [23] Just, R.E.; Weninger, Q. Are crop yields normally distributed? *Am. J. Agric. Econ.* 1999, 81, 287–304.
- [24] Kaushik, B.; Singh, K.; Tiwari, D.K.; Singh, U.K. Impact of climate change on crop yield due to pests and crop diseases. *Microsc. Microanal.* 2023, 29 (Suppl. 1), 56–58.
- [25] Rasche, L. Estimating pesticide inputs and yield outputs of conventional and organic agricultural systems. *Agronomy* 2021, 11, 1300.
- [26] Lobell, D.B.; Schlenker, W.; Costa-Roberts, J. Climate trends and global crop production since 1980. *Science* 2011, 333, 616–620.
- [27] Moore, C.E.; Meacham-Hensold, K.; Lemonnier, P.; et al. The effect of increasing temperature on crop photosynthesis: from enzymes to ecosystems. *J. Exp. Bot.* 2021, 72, 2822–2844.
- [28] Hatfield, J.L.; Prueger, J.H. Temperature extremes: effect on plant growth and development. *Weather Clim. Extrem.* 2015, 10, 4–10.
- [29] Stella, T.; Webber, H.; Olesen, J.E.; et al. Methodology to assess the changing risk of yield failure due to heat and drought stress under climate change. *Environ. Res. Lett.* 2021, 16, 104033.
- [30] Moravec, V.; Markonis, Y.; Rakovec, O.; et al. Europe under multi-year droughts: how severe was the 2014–2018 drought period? *Environ. Res. Lett.* 2021, 16, 034062.
- [31] Rosenzweig, C.; Elliott, J.; Deryng, D.; et al. Assessing agricultural risks of climate change in the 21st century. *Proc. Natl. Acad. Sci. USA* 2014, 111, 3268–3273.
- [32] Grillakis, M.G. Increase in severe and extreme soil-moisture droughts for Europe under climate change. *Sci. Total Environ.* 2019, 660, 1245–1255.
- [33] Funk, C.; Peterson, P.; Landsfeld, M.; et al. The climate hazards infrared precipitation with stations (CHIRPS) — a new environmental record for monitoring extremes. *Sci. Data* 2015, 2, 150066.
- [34] Vicente -Serrano, S.M.; Beguería, S.; López-Moreno, J.I. A multiscale drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *J. Clim.* 2010, 23, 1696–1718.
- [35] Hersbach, H.; Bell, B.; Berrisford, P.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* 2020, 146, 1999–2049.
- [36] Bai, J.; Perron, P. Computation and analysis of multiple structural change models. *J. Appl. Econom.* 2003, 18, 1–22.
- [37] Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* 2017, 30, 4765–4774.
- [38] Chen, T.; Guestrin, C. XGBoost: a scalable tree boosting system. *Proc. 22nd ACM SIGKDD* 2016, 785–794.
- [39] Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* 1997, 9, 1735–1780.
- [40] Vovk, V.; Gammerman, A.; Shafer, G. *Algorithmic Learning in a Random World*; Springer: New York, 2005.
- [41] Anselin, L. Local indicators of spatial association — LISA. *Geogr. Anal.* 1995, 27, 93–115.
- [42] Salvadori, G.; De Michele, C. Multivariate multiparameter extreme value models and return periods: a copula approach. *Water Resour. Res.* 2010, 46, W10501.

- [43] Carpenter, B.; Gelman, A.; Hoffman, M.D.; et al. Stan: a probabilistic programming language. *J. Stat. Softw.* 2017, 76, 1–32.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [GRAPHICALABSTRACT.pdf](#)