JRACR

Journal of Risk Analysis and Crisis Response, 2026, 16(2), 217-253
https://jracr.com/
ISSN Print: 2210-8491
ISSN Online: 2210-8505

Article

Drought Risk Mapping and Parametric Insurance in Agriculture: A Machine Learning-Based Framework

Yousra Belhsen [1], Rim Ouhdouch [1] and Khalil Said [1,2,*]

[1] National Institute of Statistics and Applied Economics, Rabat (10000), Morocco
[2] Applied Human Sciences and Economics in Agriculture Department of Humanities, Agronomic and Veterinary Hassan II Institute, Rabat (10000), Morocco
* Correspondence: ksaid@insea.ac.ma

Received: September 21, 2025; Received in revised form: November 29, 2025; Accepted: June 3, 2026; Available online: June 30, 2026

Abstract: Climate change has increased both the frequency and intensity of agricultural droughts, reinforcing the need for risk transfer instruments that rely on objective environmental indicators. This study introduces a comprehensive framework for the development of drought index insurance using openly available climatic and satellite data. The approach builds synthetic indicators that combine precipitation-based measures (SPI, SPEI) with vegetation indices (NDVI), selected through statistical learning techniques such as Lasso regression, Random Forest, and Partial Least Squares. These indicators are then embedded in parametric indemnity functions whose shape depends on their correlation with crop yields, ensuring decreasing or increasing payouts as appropriate. Calibration with historical yield data is performed to minimize basis risk, and pure premiums are derived through simulation methods including empirical resampling, bootstrap, Monte Carlo, and kernel density estimation. The framework is applied to three Moroccan regions with contrasting agro-climatic conditions and representative crops. Results indicate that the proposed design substantially lowers basis risk while preserving transparency, interpretability, and reproducibility. More broadly, the framework provides a transferable methodological contribution by linking machine learning with actuarial tools, supporting the development of weather index insurance in contexts where data availability remains limited.

Keywords: Drought Risk; Basis Risk; Remote Sensing; Machine Learning; Parametric Insurance; Agricultural Risk Management; Composite Drought Index; Simulation-Based Pricing

1. Introduction

Agricultural production in arid and semi-arid regions is increasingly exposed to climate variability and the rising frequency of extreme weather events. Among climate-related hazards, drought is one of the main causes of agricultural losses, with major implications for food security and rural livelihoods. Recent empirical evidence shows that climate extremes have intensified yield volatility across diverse agro-ecological systems, with substantial regional heterogeneity in exposure and sensitivity. Large-scale drought-induced production losses have been documented over multiple decades [1], while stronger temperature-moisture couplings have been shown to amplify the impacts of global warming on crop yields [2]. These findings reinforce concerns regarding the disproportionate vulnerability of developing countries to climate-induced agricultural shocks [3,4].

In such contexts, conventional agricultural insurance schemes based on field-level loss assessments suffer from structural limitations when applied to systemic and spatially covariant risks such as drought. Their effectiveness is hindered by high administrative costs, asymmetric information, and moral hazard, and they remain poorly adapted to fragmented and data-scarce environments. The inherent difficulties faced by market-based agricultural insurance programs, particularly under systemic weather risk, have been documented [5], while climate variability has been shown to complicate pricing and threaten actuarial fairness [6]. These constraints have motivated the development of alternative risk-transfer mechanisms.

In response, index-based agricultural insurance has received growing attention in both research and practice. Instead of relying on individual loss verification, payouts are linked to objective and verifiable agro-climatic indicators, such as cumulative rainfall or vegetation indices derived from satellite imagery. This mechanism allows faster and more transparent compensation while reducing many of the operational constraints associated with conventional indemnity schemes. Early conceptual foundations were established in the microinsurance and rural development literature. Index-based instruments can help break poverty traps by providing more stable and predictable compensation [7,8]. The role of index insurance in developing countries has also been reassessed, showing that design quality and institutional credibility remain decisive determinants of performance and uptake [9].

However, designing reliable index-based insurance remains a major technical and actuarial challenge. The credibility of such products depends on the selection of a robust index, the specification of payout thresholds, and the choice of pricing methodology. A key obstacle is basis risk, defined as the discrepancy between actual agricultural losses and insurance payouts triggered by the index. High basis risk undermines trust and jeopardizes adoption [10,11]. The spatial and temporal resolution of climatic and phenological information also plays a central role in determining the accuracy of index-based contracts [12,13]. These insights reinforce the need for methodological frameworks that jointly integrate data quality, agronomic relevance, and actuarial consistency.

To address these challenges, this study proposes a comprehensive and operational framework for the construction, calibration, and pricing of index-based agricultural insurance. The methodological pipeline is organized into four components. First, a set of composite drought indices is built from multiple agro-climatic indicators, namely the Standardized Precipitation Index (SPI), the Standardized Precipitation Evapotranspiration Index (SPEI), and the Normalized Difference Vegetation Index (NDVI). Several statistical and machine learning methods are used for this purpose, including regularized regression (Lasso and Ridge), dimensionality reduction techniques such as Principal Component Analysis (PCA) and Partial Least Squares (PLS), simple weighted averaging, and Random Forest. These approaches reflect an emerging literature highlighting the potential of multivariate and machine learning-based methods for capturing complex drought dynamics. For example, statistical models have been validated for climate-yield relationships in Germany [14], while multi-hazard dependencies have been shown to influence agricultural risk profiles [15].

Second, the empirical correlation between each synthetic index and crop yields is evaluated to select the most predictive indicator. The chosen index then forms the basis of the insurance contract. Depending on whether this correlation is positive or negative, the indemnity function is specified as a decreasing or increasing function of the index. The payout rule involves two thresholds. The trigger threshold is identified using quantile regression applied to the conditional distribution of yields, an

approach widely recognized for its robustness under non-normality [16]. The exhaustion threshold, corresponding to the level at which indemnities reach their maximum, is calibrated by minimizing relative basis risk, a criterion designed to capture both statistical fit and economic coherence between the index and actual yield losses [17].

Third, pure premiums are estimated using four complementary simulation-based methods, including empirical resampling, bootstrap procedures, normal Monte Carlo simulation, and kernel density estimation (KDE). This plurality of approaches allows us to assess the robustness of premium estimates under different distributional assumptions and sample constraints. KDE is particularly relevant when distributions deviate from normality, while bootstrap captures sampling uncertainty and provides confidence intervals. These pricing tools build upon a longstanding actuarial tradition emphasizing the importance of distributional modeling in weather-related insurance [18].

Finally, the performance of each insurance design is evaluated along three key dimensions: basis risk reduction, premium stability, and interpretability. This framework emphasizes transparency, reproducibility, and scalability, ensuring that the methodology remains adaptable to environments where data availability is limited.

To illustrate the empirical relevance of the framework, the analysis is applied to the Moroccan agricultural sector, which has become increasingly vulnerable to systemic droughts. Recent studies provide evidence of rising climatic stress across several regions. Declines in precipitation and water availability have been documented in the Ziz Basin [19], while heightened rainfall variability has been identified in the Tensift Basin using SPI diagnostics [20]. These findings underline the urgency of developing innovative drought risk-management instruments in Morocco.

The empirical application focuses on three regions, namely Fès-Meknès, Marrakech-Safi, and the Oriental. These regions were selected based on exploratory analysis for their climatic coherence, data quality, and distinct dominant crops (olives, soft wheat, and barley, respectively). While a global index is estimated using the full dataset, the calibration of insurance contracts, including indemnity functions and pricing procedures, is carried out separately for each region. This modeling choice reflects the importance of adapting contract parameters to local agronomic characteristics while leveraging regional climatic proximity to construct a unified risk indicator.

The results show that the proposed framework yields consistent and interpretable outcomes across regions. Among the candidate indices, the Lasso-based composite performs best in terms of correlation with yields, while quantile regression proves effective for identifying appropriate trigger levels. KDE and bootstrap methods provide reliable premium estimates under non-normality and limited data conditions. Overall, the integration of statistical learning for index construction, robust threshold calibration, and simulation-based pricing produces transparent and data-driven insurance products that are well suited to environments marked by climatic volatility and weak institutional infrastructure.

This study contributes to the literature by offering a modular and transferable framework for index-based agricultural insurance that combines statistical rigor with actuarial design. While the empirical illustration is grounded in the Moroccan context, the methodology is generalizable to other crops, regions, and data ecosystems. The remainder of the article is structured as follows. Section 2 reviews the existing literature on index-based agricultural insurance. Section 3 presents the data sources and the construction of agro-climatic indicators. Section 4 details the methodological

framework, including index construction, threshold calibration, and premium estimation. Section 5 reports the empirical results, and Section 6 concludes.

2. Literature Review

Agriculture is increasingly exposed to climate-induced uncertainty, particularly in regions where it constitutes the backbone of rural economies. Farming systems in developing countries face a dual challenge: they are highly sensitive to climatic shocks while simultaneously sustaining the livelihoods of the most vulnerable populations [3]. The issue lies not only in their physical exposure to changing weather patterns but also in their limited capacity to anticipate and absorb such shocks. A complementary analysis emphasizes the intricate links between crop performance and climate variability, underscoring the difficulties inherent in forecasting agricultural outcomes under uncertain environmental conditions [4]. Empirical crop-level studies further confirm that relatively modest shifts in temperature and moisture regimes can have substantial and heterogeneous effects on yields. Summer climate exerts a decisive influence on wheat quality in the United Kingdom [21], while winter wheat phenology shows spatiotemporal sensitivity to weather and climate variability in China [22]. In this context, the need for robust and scalable tools to monitor drought and water stress has become increasingly pressing.

Over the past two decades, significant progress has been made in developing objective indicators capable of capturing both the temporal dynamics and spatial heterogeneity of drought episodes. Remote sensing technologies have played a central role in this evolution, providing continuous, large-scale measurements of vegetation responses. Among the most widely adopted tools is the Normalized Difference Vegetation Index (NDVI), which serves as a proxy for plant vigor and ecosystem stress. NDVI-based applications have been widely used across ecological systems, illustrating their relevance for drought monitoring and agricultural risk assessment [23].

Numerous studies have confirmed the effectiveness of NDVI in identifying crop phenology and signaling the early onset of water deficit. Time-series MODIS data have been used to characterize vegetation dynamics with high temporal resolution [24], while NDVI-based methods have been developed to delineate crop stages, reinforcing their applicability for agricultural observation systems [25]. Despite these advantages, NDVI alone cannot detect meteorological drought before it translates into visible vegetative stress, which limits its capacity for early warning.

To address this limitation, climate-based indices such as the Standardized Precipitation Index (SPI) and the Standardized Precipitation Evapotranspiration Index (SPEI) have gained widespread use. The SPI was introduced to capture precipitation anomalies across multiple time horizons [26], while the SPEI incorporates temperature through potential evapotranspiration, thereby increasing its responsiveness to climate change [27].

Combining NDVI with climatic indices has been increasingly recommended to improve both spatial coverage and temporal precision in drought diagnostics. Integrating satellite-based rainfall estimates with NDVI enhances the reliability of early warning systems, particularly in areas where ground-based meteorological observations are scarce [28]. Similarly, merging meteorological and vegetation-based information is important for better capturing the multidimensional nature of drought processes [29]. Taken together, these studies show that the joint use of NDVI, SPI, and SPEI provides a more comprehensive representation of drought dynamics. Building on this perspective,

the present study proposes a methodological framework aimed at developing a robust and transferable drought index for agricultural risk modeling.

Historical analyses also provide important insights for the design of drought-related insurance instruments. A systematic review of weather index insurance schemes traces their evolution from traditional indemnity-based models to more sophisticated index-based solutions, while underscoring persistent institutional and design challenges [30]. Index-based risk-transfer products can enhance microfinance by reducing default risk associated with climate shocks, thereby strengthening the resilience of local financial systems [31]. Reviews of index-based agricultural insurance initiatives in Sub-Saharan Africa identify persistent obstacles related to data quality, product design, and institutional capacity, which often constrain uptake in precisely those settings where climate vulnerability is most acute [32]. Index-based insurance has emerged as a promising alternative to conventional indemnity schemes, particularly in contexts where on-the-ground loss verification is costly, unreliable, or logistically difficult. By linking compensation to observable and objectively verifiable indicators, such as precipitation or vegetation metrics, these products seek to mitigate problems of moral hazard and adverse selection while reducing administrative burdens. Yet, despite these advantages, the adoption of such products by smallholder farmers remains limited in many regions. A key obstacle identified in the literature is basis risk.

A comprehensive review of agricultural index insurance programs shows that their effectiveness and legitimacy critically depend on the quality and local relevance of the indices employed [10]. In a related empirical contribution, livestock index insurance in northern Kenya has been analyzed, showing how sizable mismatches between payouts and realized losses can directly undermine welfare gains and perceived fairness, even when the underlying index is constructed with considerable technical rigor [33]. When the chosen index fails to reflect actual crop performance, insured farmers may either be denied compensation for genuine losses or receive payments despite not experiencing damages. Such inconsistencies erode trust in the product and discourage uptake, particularly in settings characterized by low financial literacy and limited institutional support.

A randomized field experiment in India demonstrates that farmers' willingness to pay declines sharply when they are explicitly informed about the limitations of the index mechanism [34]. This behavioral evidence underscores that technical rigor alone is insufficient; perceptions of fairness and clarity in contract design are equally critical. Experimental evidence further shows that low trust and limited understanding of weather-index insurance substantially hinder demand, emphasizing the crucial role of effective communication and institutional credibility [35]. Long-run field evidence also shows that learning from past payout performance, as well as evolving beliefs about product reliability, strongly shapes farmers' decisions to renew or abandon index-based coverage over time [36]. Complementary evidence indicates that informal risk-sharing networks can crowd out formal index insurance, particularly when community-based arrangements are perceived as partial substitutes for market mechanisms [37]. These findings highlight that adoption decisions are shaped not only by product-specific characteristics but also by the broader social structures through which farmers manage climate risk.

This behavioral perspective has also been extended to high-income countries, where willingness to purchase index-based flood insurance is strongly shaped by environmental risk perceptions and insurance literacy [38]. These findings indicate that even when market and institutional conditions

are more favorable, demand remains sensitive to trust, comprehension, and prior exposure to climate hazards.

To mitigate basis risk, recent efforts have focused on improving the spatial and temporal precision of input data. Spatial interpolation techniques, such as kriging and inverse distance weighting, have been evaluated for enhancing the geographic resolution of weather-based indices [39]. These findings suggest that refined interpolation methods can improve payout accuracy, particularly in regions with sparse meteorological coverage. Gridded precipitation data have also been combined with phenological observations, showing that aligning index construction with crop development stages can substantially reduce spatial and temporal basis risk [40].

A promising avenue for advancing index-based insurance lies in customizing indicators to reflect local cropping systems and agro-climatic conditions. A data-driven approach combining NDVI with clustering algorithms has been proposed to establish context-specific trigger and exhaustion thresholds for crop insurance in Northern Ethiopia [41]. These results show that tailoring index design to empirical patterns of yield loss can substantially reduce basis risk and strengthen the practical relevance of insurance schemes. This approach has also been extended to irrigated systems through the development of a hydrological drought index based on water storage availability, adapted to mitigate supply shortfalls in Mediterranean agriculture [42]. Multi-year index-based contracts have further been designed for water utility companies in Brazil to insure against hydrological drought, thereby adapting parametric instruments to slow-onset risks and infrastructure-oriented decision contexts [43]. From a broader portfolio perspective, price hedging and weather-shock hedging have been compared for cotton producers in Cameroon, showing that weather index insurance can complement, rather than simply substitute, existing instruments for managing income volatility [44]. These contributions demonstrate the conceptual flexibility of index insurance design across diverse agronomic contexts.

At the same time, recent advances in machine learning and statistical modeling have expanded the range of tools available for developing drought indicators adapted to agricultural applications. Traditional univariate indices, while intuitive and widely used, often fail to capture the complex and nonlinear interactions between environmental variables and crop yields. By contrast, machine learning techniques can integrate multiple predictors and model higher-order relationships, resulting in improved predictive accuracy and a closer correspondence with observed outcomes. Combining crop-model simulations with machine learning algorithms significantly enhances yield prediction accuracy, thereby improving the alignment between modeled indices and actual agricultural losses [45]. Comparisons of several machine learning algorithms for predicting building insurance outcomes illustrate how data-driven models can improve risk classification and pricing in non-agricultural lines, thereby reinforcing the broader relevance of such methods for parametric and index-based products [46]. These methodological advances further underscore the relevance of index-based insurance as a flexible instrument for managing recurring climate shocks. Such schemes are particularly well suited to moderate but frequent losses, while traditional insurance remains more appropriate for rare catastrophic events [47].

A case study from Morocco combines remote sensing data with Random Forest and Gradient Boosting algorithms to assess the impact of drought on arboricultural systems [48]. Its findings demonstrate the potential of such models to capture spatial heterogeneity in drought intensity and to estimate crop losses with greater accuracy. Random Forest has become a widely adopted algorithm

in agricultural applications due to its resistance to overfitting and its ability to handle high-dimensional input spaces [49]. In parallel, regularized regression techniques such as Lasso [50] and Ridge regression [51] provide valuable alternatives for selecting relevant predictors from large sets of climatic and spectral variables. These methods promote model parsimony and enhance numerical stability, both of which are crucial for developing transparent and reliable indices in insurance design. However, the increasing sophistication of these approaches also raises important questions regarding interpretability and trust, making calibration and stakeholder confidence central to their practical adoption.

Other dimensionality reduction techniques, such as Principal Component Analysis [52] and Partial Least Squares regression [53], have been increasingly applied to construct synthetic drought indicators. These methods condense high-dimensional climatic and spectral data into a limited number of orthogonal components, which can then be calibrated against crop yield observations to generate indices with strong explanatory and predictive power. A key challenge in employing such approaches lies in balancing transparency with predictive performance. While more complex models may deliver superior fits to historical data, their opacity can hinder acceptance by regulators and farmers alike. In the context of index-based insurance, interpretability is not merely advantageous but essential to fostering trust and enabling adoption. Building on this literature, the present study compares several statistical and machine learning methods to construct a composite drought index from SPI, SPEI, and NDVI. Among these, Lasso regression was ultimately retained for its strong predictive capacity and interpretability in our empirical setting.

However, selecting input variables is only the first step. Calibrating both the index and the associated indemnity function is equally critical to minimizing basis risk and strengthening the robustness of index-based products. Several studies emphasize that even statistically reliable indices may fail to align with realized yield losses if payout thresholds are mis calibrated or disconnected from local crop phenology. In this regard, quantile regression has emerged as a valuable tool for defining trigger levels in index insurance. Unlike traditional methods based on conditional means, quantile regression focuses on specific percentiles of the yield distribution, typically the lower tail where losses are concentrated. The foundational framework for this approach has since been adapted to drought insurance to ensure that compensation mechanisms are activated only under genuinely adverse conditions [16].

Shaping the functional form of the indemnity rule to reflect empirical behavior represents another crucial refinement. Instead of adopting rigid payout structures, several researchers advocate flexible specifications that better capture observed relationships between yield losses and the underlying index. In parametric agricultural insurance, nonlinear indemnity curves can reduce basis risk without compromising actuarial soundness [17]. Along the same lines, surrogate models derived from machine learning and remote sensing have been introduced to improve both the precision of indices and the reliability of indemnity payouts [54]. While such models are often complex, maintaining transparency remains fundamental to sustaining stakeholder confidence and securing regulatory approval. To support the calibration process, the concept of Relative Basis Risk (RBR) has been advanced as a standardized performance criterion. By jointly considering statistical goodness-of-fit and the economic consistency of payouts, RBR facilitates the selection of index specifications and exhaustion thresholds that minimize unjustified payments and coverage gaps.

While methodological refinement is essential for improving index accuracy, the practical success of index-based insurance ultimately depends on behavioral and institutional conditions. Even rigorously designed instruments may fail to achieve adoption if they are not perceived as fair, accessible, and trustworthy by their intended users. Empirical studies consistently show that farmers' willingness to participate in such schemes is strongly influenced by perceptions of credibility, previous insurance experiences, and the broader institutional environment in which these products are introduced. Early overviews of weather index insurance initiatives in Sub-Saharan Africa highlight persistent issues related to data availability and trust [55]. In the Ghanaian context, institutional trust, access to extension services, and prior involvement in public programs have been identified as key determinants of adoption [56]. These findings suggest that technical optimization must be complemented by outreach strategies and sustained engagement with farming communities through credible intermediaries. The successful implementation of index-based insurance in developing countries hinges not only on technical design but also on addressing trust, comprehension, and perceived fairness among smallholder farmers [10].

In parallel, several researchers have highlighted the importance of tailoring index insurance to local agronomic and environmental conditions. An NDVI-derived drought index (DevNDVI) combined with site-specific yield thresholds based on historical observations has been proposed to align index triggers with local crop calendars and drought dynamics [41]. Recent contributions further illustrate the potential for adapting index design to regional contexts, particularly in Morocco. Sentinel-2 imagery and machine learning have been used to capture spatial heterogeneity in drought impacts on tree crops, providing valuable guidance for localized calibration [48]. NDVI time series have also demonstrated predictive capacity for estimating cereal yields in the Fès-Meknès region, reinforcing the relevance of satellite-derived vegetation indices for parametric insurance applications [57]. Taken together, these studies confirm that Morocco constitutes a relevant empirical setting for testing index-based insurance solutions. This is not presented as an exclusive objective but rather as a practical demonstration of how data-driven methodologies can be operationalized in semi-arid and data-constrained agricultural systems. The framework developed in this article builds on these insights to offer a transferable approach that is consistent with international methodological standards while remaining responsive to local implementation challenges.

Overall, the literature demonstrates that advances in climatic and remote sensing indices, machine learning methods, and calibration strategies have significantly enhanced the potential of index-based insurance. Yet, persistent challenges remain in balancing predictive accuracy, interpretability, and stakeholder acceptance, particularly in data-scarce agricultural systems. This article contributes to this body of work by proposing a comprehensive framework that integrates statistical and machine learning approaches with agronomic knowledge, while explicitly addressing calibration and basis risk. The empirical application to Morocco does not serve as an isolated case but as a demonstration of how robust, data-driven methodologies can be operationalized in semi-arid and resource-constrained contexts, with broader relevance for global index insurance design.

3. Data

This study focuses on key agricultural regions in Morocco that are particularly exposed to climatic drought and where rainfed agriculture plays a central role in rural livelihoods. The selected regions include Oriental, Fès-Meknès, Marrakech-Safi, and Souss-Massa. Each region is associated

with a different dominant crop, such as olives in Fès-Meknès or barley in the Oriental. Despite their agronomic diversity, these areas share structural vulnerabilities that make them relevant case studies for index-based insurance modeling. Similar vulnerabilities have been documented in empirical studies linking temperature and precipitation variability to agricultural output, where weather-induced shocks generate sizeable yield losses across a wide range of crops and regions [1,2,58,59].

Annual precipitation in these areas is highly variable, often falling below 300 mm in the driest zones, and is subject to substantial interannual fluctuations. This climatic uncertainty, coupled with the prevalence of rainfed systems, has historically led to severe socio-economic consequences during drought episodes. The present analysis builds on these structural challenges to illustrate how open-access climatic and satellite data can be used to construct parametric insurance products tailored to such environments. Although the empirical focus is placed on a few Moroccan regions, the proposed framework is generalizable and replicable in other contexts where data constraints are similar. Finally, it is worth noting that the Souss-Massa region, initially considered in the dataset, will be excluded from the final results due to structural divergence in agro-climatic patterns.

3.1. Sources and Preparation

The analysis is based on a comprehensive and carefully curated dataset spanning the period from January 2010 to December 2023. It integrates meteorological observations, satellite-derived vegetation indicators, and official agricultural production statistics. This 14-year window offers a robust temporal framework for capturing both structural trends and short-term variability, which is essential for drought modeling in semi-arid agricultural systems. The use of such a multi-year panel is consistent with climate-yield assessments showing that disentangling weather-driven variability from structural trends requires sufficiently long time series and spatially explicit data [14,60,61].

Meteorological data were obtained from three complementary sources. First, the Moroccan National Meteorology Department (DMN) provided validated ground station observations with national coverage. Second, NASA's POWER project supplied a range of satellite-based atmospheric variables on a global grid, including solar radiation and evapotranspiration. Third, precipitation estimates were retrieved from the CHIRPS database, which combines satellite imagery with ground station data to generate high-resolution, bias-corrected rainfall grids. All meteorological variables, including daily minimum and maximum temperatures, cumulative rainfall, solar radiation, and evapotranspiration, were spatially interpolated and projected to a uniform grid at 1 km × 1 km resolution to ensure consistency across sources. This harmonization step was critical for allowing integrated analyses and matching the spatial scale of vegetation and yield data.

Vegetation dynamics were monitored through the Normalized Difference Vegetation Index (NDVI), derived from multiple satellite platforms. The primary data sources included SPOT-VEGETATION and Proba-V sensors, which offer high spatial resolution and long-term continuity. To enhance temporal density and fill observational gaps, MODIS products from NASA's Goddard Space Flight Center were also incorporated. NDVI series were aggregated at both monthly and seasonal scales to capture crop development across different phenological phases. Particular care was taken to avoid cloud-contaminated pixels and ensure consistency across satellite sources. Focusing on phenologically meaningful aggregation windows is in line with studies showing that yield responses are particularly sensitive to weather anomalies during critical growth stages [21,62].

Agricultural data were collected from two official channels. The Moroccan Directorate of Statistics and Surveys (DSS) provided yield and acreage data based on field surveys conducted using area sampling methods, which ensure statistical representativeness at the regional level. In addition, self-reported yield estimates were obtained from the Agricultural Extension Council. While these administrative records allow for useful cross-validation, they were treated with caution due to possible declarative bias. All datasets were cleaned and harmonized to ensure compatibility with the spatial and temporal resolution of the climatic and vegetative indicators. This preprocessing ensures that the final dataset is both representative and directly usable for empirical illustration.

The final selection of the four study regions was based on data completeness, agro-climatic diversity, and economic relevance. Each region had to meet minimum standards for data coverage across all three dimensions: climate, vegetation, and yields. This selection process ensured that the empirical illustration would be robust and replicable, while the final database was structured as a balanced regional panel ready for statistical modeling and the design of a generalizable parametric insurance framework under realistic field conditions.

In addition to assembling the multi-source climatic, meteorological, and agricultural indicators described above, the spatial data used for the illustrative maps were processed through a dedicated geospatial workflow. NDVI raster files were exported as GeoTIFF images from Google Earth Engine and imported into QGIS (QGIS 3 family), where they were combined with GADM v4.1 regional boundaries to generate publication-ready cartographic outputs. All maps were created in the QGIS Print Layout environment to ensure consistent rendering of color palettes, legends, scale bars and projection information.

Descriptive statistical analyses were performed using Python 3.11, which served as the core analytical environment throughout the study. The same Python workflow was used for correlation matrices, temporal trends, yield distributions and vegetation-climate diagnostics, ensuring a coherent treatment of the dataset. This analytical environment was subsequently maintained for all methodological components presented in the next section, including synthetic index construction, threshold calibration and simulation-based premium estimation, thereby guaranteeing a unified and fully reproducible computational pipeline.

3.2. Drought and Vegetation Indices Construction

To quantify climatic anomalies and their potential impacts on vegetation and crop yields, three key indices were constructed: the Standardized Precipitation Index (SPI), the Standardized Precipitation Evapotranspiration Index (SPEI), and the Normalized Difference Vegetation Index (NDVI). These indicators are widely used in climate impact and extremes analysis to summarize anomalies in precipitation, water balance and vegetation conditions in a standardized form [63,64].

The SPI measures deviations in precipitation relative to long-term climatological averages [26]. In this study, three-month rolling cumulative precipitation totals, denoted by Pm_3, were computed for each region. These totals were fitted to a Gamma distribution, which is widely used in hydrology due to the positive skewness of precipitation data. To account for months with zero precipitation, a mixed distribution approach was applied:

$$H(c) = \begin{cases} q_0, & \text{if } c = 0, \\ q_0 + (1 - q_0) \cdot F_{\text{Gamma}}(c), & \text{if } c > 0, \end{cases} \quad (1)$$

where $q_0$is the empirical probability of zero precipitation, and $F_{\text{Gamma}}(c)$denotes the cumulative

distribution function (CDF) of the fitted Gamma distribution. Standardization was then performed using the inverse standard normal transformation:

$$SPI = \Phi^{-1}(H(Pm_3)). \quad (2)$$

This transformation produces a normalized drought index where negative values indicate dry conditions and positive values indicate wet periods. The adequacy of the Gamma model was confirmed using Kolmogorov-Smirnov tests, with p-values consistently above the 0.05 threshold across all regions.

The SPEI extends the SPI by incorporating the effect of potential evapotranspiration, thereby capturing a more comprehensive picture of atmospheric water demand [27]. The monthly climatic water balance was computed as:

$$D_t = P_t - \mathrm{ETP}_t, \quad (3)$$

where P_t is monthly precipitation and ETP_t is monthly potential evapotranspiration. Three-month cumulative balances, denoted by Dm_3, were then fitted to a Pearson Type III distribution, which provides flexibility for skewed hydrological series. Standardization was achieved similarly:

$$SPEI = \Phi^{-1}(G(Dm_3)), \quad (4)$$

where $G(Dm_3)$ represents the CDF of the fitted Pearson III distribution. Kolmogorov–Smirnov tests also confirmed an adequate fit for all regions.

The NDVI, a widely used proxy for vegetation vigor [23], was computed from satellite imagery using the classical spectral ratio:

$$NDVI = \frac{NI}{NIR+RED}, \quad (5)$$

where NIR and RED represent the spectral reflectance in the near-infrared and red bands, respectively. To capture phenologically meaningful signals, NDVI values were averaged over the April-June period, which corresponds to the critical crop growth stage in the studied regions.

Figure 1 displays the NDVI map of Morocco for 2022, overlaid with administrative boundaries to highlight the four regions selected for further analysis: Oriental, Souss-Massa, Marrakech-Safi, and Fès-Meknès. The regional panels provide a more detailed view of vegetation patterns within these areas, which represent diverse agro-climatic zones and account for a significant share of Morocco's agricultural production. Together, these maps provide the spatial reference for the subsequent empirical analysis.

Together, these three indices offer complementary insights into climatic variability and vegetation dynamics. They will serve as central inputs for the subsequent empirical modeling of agricultural risk and the evaluation of index-based insurance performance.

To ensure biological relevance, NDVI, SPI, and SPEI values were aggregated over crop-specific seasonal windows tailored to the dominant crop in each region. NDVI was averaged over March-May for soft wheat in Marrakech-Safi, February-April for barley in Oriental, and May-July for olives in Fès-Meknès. These periods correspond to the critical vegetative development stages of each crop and were selected according to agronomic calendars. Similarly, SPI and SPEI values were computed over slightly broader windows, occasionally including preceding months to capture cumulative hydrological stress. This tailored aggregation strategy enhances the sensitivity of the indices to drought conditions most likely to affect yields, as recommended in the literature [27] and supported by empirical evidence on the role of critical phenological stages in climate-yield relationships [21,62].

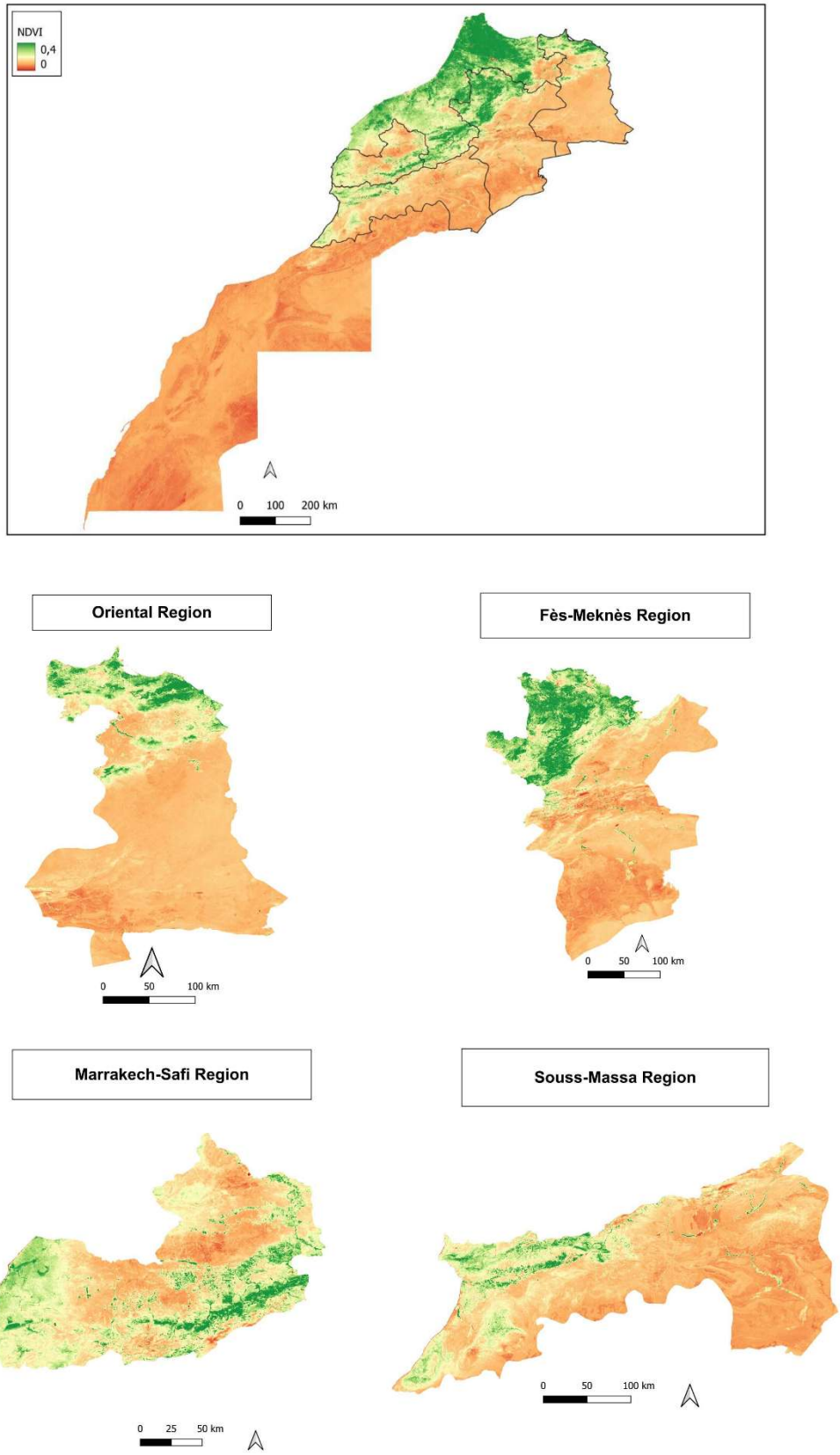

Figure 1. NDVI map of Morocco and NDVI maps of the selected regions (2022).

3.3. Descriptive Analysis

The descriptive statistics reveal pronounced heterogeneity across regions in both climatic and agricultural variables. Monthly precipitation is positively skewed in all areas, with the majority of observations below 50 mm. This distribution aligns with the semi-arid to arid conditions that characterize Morocco's main agro-climatic zones. Such positively skewed precipitation distributions, dominated by frequent low rainfall and occasional heavy events, are typical of semi-arid agro-climatic regimes and are known to exert a strong influence on yield variability [58,59,63].

The dataset includes monthly observations for four key agricultural regions (Fès-Meknès, Marrakech-Safi, Oriental, and Souss-Massa), comprising meteorological records (precipitation and evapotranspiration), satellite-based vegetation indices (NDVI), and crop yield measurements. The time series are complete, with no missing values throughout the observation period.

3.3.1. Summary Statistics by Region

Table 1 reports the mean and standard deviation of the key variables for each region. The results reveal pronounced contrasts in rainfall, vegetation cover, evapotranspiration, and agricultural yields, reflecting the climatic and agronomic diversity across the study areas.

Table 1. Regional summary statistics (mean ± standard deviation).

Region	Precipitation	NDVI	ET	Yield
Fès-Meknès	35.5 ± 29.7	0.32 ± 0.12	177.5 ± 9.5	1.54 ± 0.56
Marrakech-Safi	15.3 ± 14.4	0.09 ± 0.07	187.5 ± 10.2	1.01 ± 0.49
Oriental	21.2 ± 15.6	0.08 ± 0.07	177.5 ± 9.6	0.59 ± 0.23
Souss-Massa	13.9 ± 16.9	0.53 ± 0.10	155.0 ± 8.3	15.46 ± 4.74

3.3.2. Exploratory Distributions and Temporal Patterns

Figures B1 and B2 in Appendix B illustrate the distributions of the main variables. Precipitation displays a pronounced positive skew in all regions, with a high frequency of low or zero rainfall and sporadic extreme events, a typical feature of semi-arid climates.

NDVI distributions exhibit a clear bimodal structure, mirroring the alternation between dormancy and peak vegetative growth within the agricultural cycle. This seasonality is particularly visible in rainfed regions (Oriental, Marrakech-Safi, and Fès-Meknès), while Souss-Massa shows a distinctive profile driven by irrigated citrus production. The pronounced seasonality in vegetation signals is consistent with evidence that crop yields react most strongly to weather anomalies during specific growth phases [21,62].

Annual crop yield statistics (Table 2) highlight sharp regional contrasts. Souss-Massa achieves the highest mean yield (15.46 t/ha) and the largest standard deviation (4.74 t/ha), reflecting both the intensity and volatility of irrigated horticulture. Conversely, rainfed regions record substantially lower yields, generally below 2 t/ha, with limited interannual variation.

Figures A1, A2, and A3 in Appendix A reveal consistent seasonal dynamics across climatic and vegetation variables. Evapotranspiration and NDVI follow synchronized cycles with temperature and solar radiation, peaking in spring and early summer. Precipitation remains highly irregular, with annual maxima typically occurring in late autumn and early spring.

Table 2. Descriptive statistics of annual crop yields (2010–2023).

Region	Mean	Std. dev.	Min	Max
Fès-Meknès	1.54	0.56	0.5	2.6
Marrakech-Safi	1.01	0.49	0.4	2.0
Oriental	0.59	0.23	0.2	1.0
Souss-Massa	15.46	4.74	4.0	22.0

Yield trajectories in Figure A4 underline the structural gap between irrigated and rainfed systems. Souss-Massa combines higher average yields with greater volatility, reflecting market-driven horticultural practices. Rainfed regions display relatively flat and stable trends, indicative of their strong climatic dependence and limited technological intensification. Similar contrasts between irrigated and rainfed systems have been reported in European and global assessments of climate impacts on crop yields, where exposure to rainfall deficits and heat stress amplifies interannual variability in rainfed production [60,61,14,1].

3.3.3. Contrasts and Yield Sensitivities Across Regions

Figure B3 in Appendix B presents the regional distributions of evapotranspiration (ET). Souss-Massa displays both the lowest median and the smallest variability, illustrating the buffering effect of irrigation infrastructure. In contrast, rainfed regions exhibit wider dispersion, directly reflecting their exposure to climatic fluctuations. These contrasts highlight heterogeneous climatic pressures across regions, which can in turn shape the sensitivity of yields to environmental drivers.

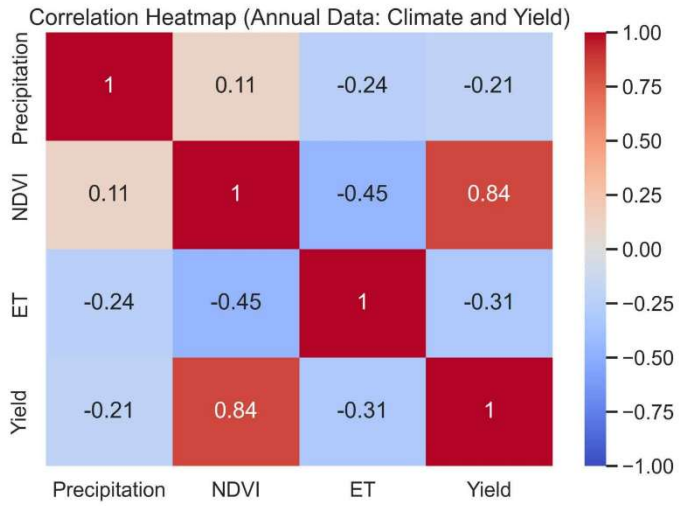

Figure 2. Annual correlation matrix among NDVI, ET, precipitation, and yield.

To further examine these relationships, we compute the annual correlation matrix of yield, NDVI, ET, and precipitation (Figure 2). NDVI clearly emerges as the most informative vegetation indicator of yield variability. In contrast, evapotranspiration shows only weak and inconsistent associations with yield, suggesting limited explanatory power under the observed conditions. While these national-level results reveal broad climatic contrasts, they do not capture the structural role of crop specialization, which is addressed in the next subsection.

3.3.4. Crop-Type Specialization and Yield Distributions

The regional yield distributions, shown in Figure 3, reveal a marked specialization, with each zone associated with a dominant crop: citrus in Souss-Massa, olives in Fès-Meknès, soft wheat in Marrakech-Safi, and barley in Oriental. From a modeling perspective, this specialization introduces a confounding structure that complicates causal interpretation. In particular, the irrigated citrus system in Souss-Massa displays substantially higher and more volatile yields compared to the rainfed cereal and olive systems. To preserve the consistency of the econometric analysis, Souss-Massa is therefore excluded from subsequent modeling. This decision is further supported by the temporal yield patterns reported in Appendix A, which confirm the structural divergence between irrigated and rainfed systems. Accounting for this specialization is essential before conducting regional correlation analyses with explicit significance testing, presented in the next subsection.

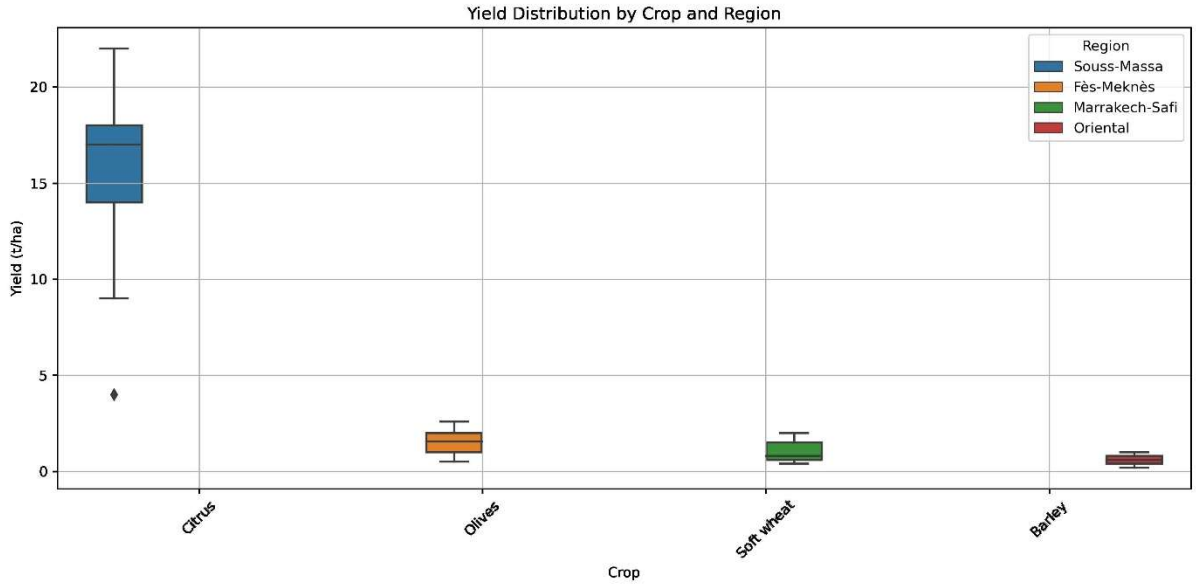

Figure 3. Yield distributions by region and dominant crop type.

3.3.5. Regional Correlation Analysis and Significance Testing

Table 3. Correlation between yield and climatic variables.
(Pearson / Spearman / Kendall; p-values in parentheses).

Region	Variable	Pearson	Spearman	Kendall
Fès-Meknès	NDVI	0.539 (0.0000)	0.546 (0.0000)	0.431 (0.0000)
	Precip.	0.407 (0.0000)	0.368 (0.0000)	0.288 (0.0000)
	ET	0.000 (1.0000)	0.000 (1.0000)	0.000 (1.0000)
Marrakech-Safi	NDVI	0.187 (0.0154)	0.070 (0.3670)	0.060 (0.3376)
	Precip.	0.413 (0.0000)	0.376 (0.0000)	0.293 (0.0000)
	ET	-0.000 (1.0000)	0.000 (1.0000)	0.000 (1.0000)
Oriental	NDVI	0.158 (0.0403)	0.067 (0.3911)	0.054 (0.3757)
	Precip.	0.481 (0.0000)	0.465 (0.0000)	0.351 (0.0000)
	ET	-0.000 (1.0000)	0.000 (1.0000)	0.000 (1.0000)
Souss-Massa	NDVI	0.643 (0.0000)	0.650 (0.0000)	0.512 (0.0000)
	Precip.	0.284 (0.0002)	0.210 (0.0064)	0.163 (0.0049)
	ET	0.000 (1.0000)	0.000 (1.0000)	0.000 (1.0000)

To complement the national-level findings, we conducted region-specific correlation analyses between crop yields and three explanatory variables, namely NDVI, precipitation, and evapotranspiration (ET), using Pearson, Spearman, and Kendall coefficients. The results, including p-values, are reported in Table 3.

NDVI consistently emerges as the strongest and most statistically significant predictor of yield, particularly in Souss-Massa and Fès-Meknès. Precipitation also shows a robust positive association with yield across all regions, with the highest correlations observed in Oriental and Marrakech-Safi. By contrast, evapotranspiration (ET) uniformly shows null correlations with p-values equal to 1.000, underscoring its lack of explanatory relevance when considered as an isolated variable at the regional scale. This does not contradict the potential usefulness of composite drought indices such as the SPEI, which integrates precipitation and potential evapotranspiration (PET) into a cumulative water balance measure. In the econometric modeling phase, ET is therefore excluded in favor of NDVI, precipitation, and derived indices that more effectively capture yield sensitivities to climatic stress.

3.3.6. Drought Index Behavior

To assess the temporal evolution of drought conditions, we analyze the behavior of two standardized indicators: the Standardized Precipitation Index (SPI) and the Standardized Precipitation-Evapotranspiration Index (SPEI). Both indices are computed at a three-month timescale to capture short-term water balance anomalies that are critical for agricultural productivity. This timescale is widely used in drought monitoring applications focusing on seasonal water balance anomalies relevant for crop growth [63,64].

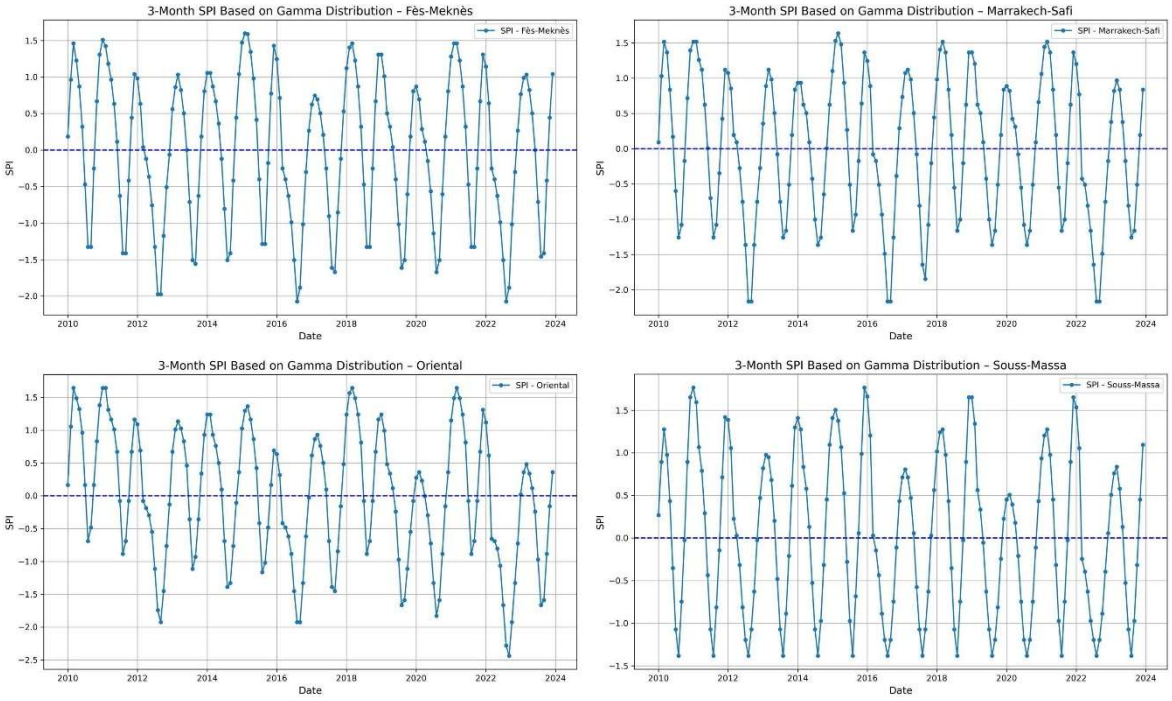

Figure 4. SPI time series (2010–2023) across the four study regions.

Figure 4 presents the SPI time series from 2010 to 2023 across the four regions. The series oscillate around zero with typical amplitudes between -2 and +2. Extended negative phases, particularly in Oriental and Marrakech-Safi, highlight recurrent drought episodes.

Figure 5 shows the SPEI time series over the same period. Compared to SPI, SPEI captures more pronounced extremes because it accounts for temperature-driven evapotranspiration in addition to precipitation deficits. In Souss-Massa and Marrakech-Safi, values occasionally fall below -8, reflecting acute hydrological stress during heatwaves or prolonged dry spells. Such extreme values result from the local standardization of cumulative water balance anomalies and emphasize the regions' exposure to compounded drought conditions. These compounded stress patterns are consistent with broader evidence that concurrent precipitation deficits and high evaporative demand substantially increase the risk of severe yield losses [2,1].

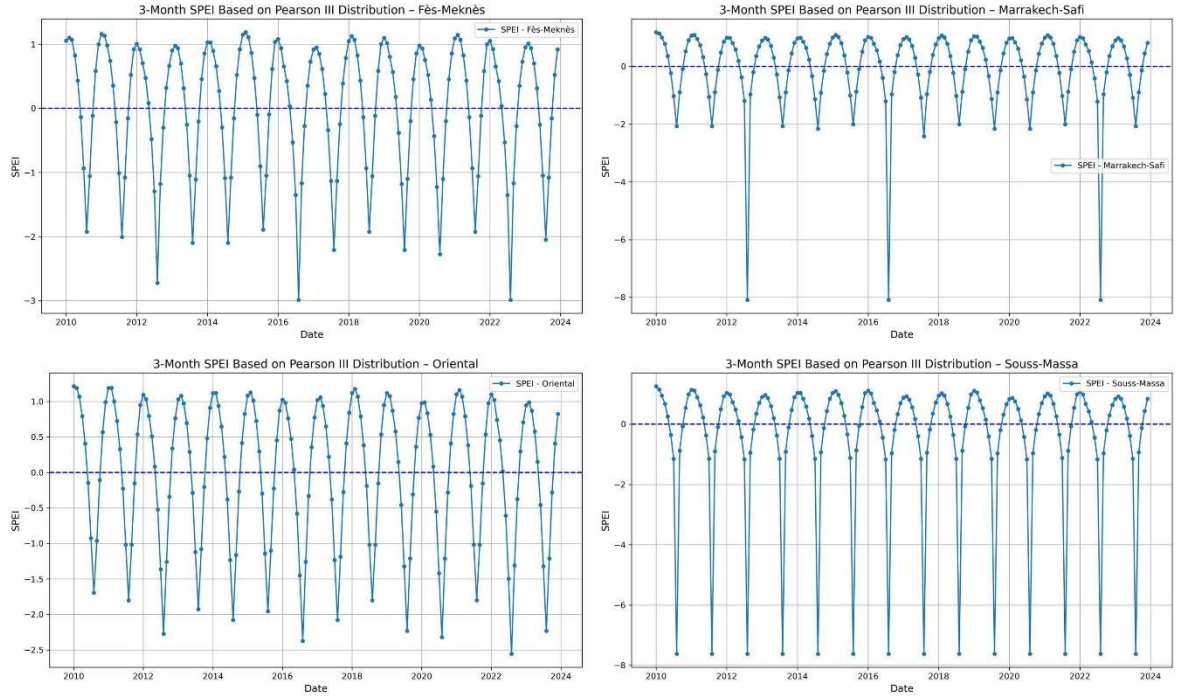

Figure 5. SPEI time series (2010–2023) across the four study regions.

Theoretical bounds and interpretation thresholds for these indices are summarized in Table 4. Although both SPI and SPEI are theoretically defined in the [-3,+3] range, empirical estimates may exceed these limits due to deviations in observed distributions.

Table 4. Expected value ranges and interpretation thresholds.

Index	Typical Range	Interpretation
SPI	$[-3, +3]$	$<$ -2: severe drought; $>$ 2: extreme wetness
SPEI	$[-3, +3]$	$<$ -2: severe drought; $>$ 2: extreme wetness
NDVI	[0,1]	Near 0: bare soil; Near 1: dense vegetation

Model adequacy was assessed by comparing the empirical cumulative distribution function (CDF) of the three-month water balance variable D_3 with the fitted Pearson Type III distribution. Diagnostics, reported in Appendix C, reveal slight deviations in the lower tail, suggesting a potential underestimation of extreme drought probabilities. Nevertheless, the Pearson III distribution remains well suited for hydrological extremes and is widely adopted in agro-climatic risk analysis.

Overall, SPI and SPEI provide complementary perspectives on drought dynamics. SPI isolates the effect of precipitation variability, whereas SPEI reflects combined water deficits driven by both rainfall shortage and high evaporative demand. Examining their joint behavior, as illustrated in

Figures 4 and 5 and summarized in Table 4, enhances the identification of critical drought episodes. This dual-indicator approach strengthens the climatological foundation for subsequent modeling and supports the design of robust parametric insurance products.

In summary, these descriptive insights on climatic variability, vegetation dynamics, and drought indicators provide the empirical foundation for the econometric modeling framework presented in the next section.

4. Methodology

This section presents a generalizable framework for the design of index-based insurance products adapted to agricultural drought. The methodology is structured in a way that allows for a comparative evaluation of multiple statistical approaches at each stage of the process: composite index construction, indemnity threshold selection, and premium estimation. This structured perspective is consistent with recent reassessments of index-based agricultural risk transfer instruments that emphasize the joint evaluation of index design, contract parameters, and pricing performance [8,9].

Although the framework is illustrated using Moroccan regional data on crop yields and climate indicators, it is conceived to be fully transferable to other contexts provided that adequate climatic and agronomic data are available. The approach emphasizes reproducibility and transparency, which are crucial requirements for both actuarial practice and regulatory validation. These aspects are particularly critical in agricultural insurance markets, where adverse selection, moral hazard, and limited take-up have been repeatedly documented for both indemnity-based and index-based schemes [5,6,65].

The methodological pipeline unfolds in three main stages:

1. Composite index construction: several techniques are used to combine climatic and vegetation indicators (SPI, SPEI, NDVI) into synthetic drought indices, including both supervised and unsupervised methods.
2. Indemnity threshold selection: alternative econometric and statistical strategies are compared to determine trigger and exhaustion thresholds that provide actuarial soundness and agronomic relevance.
3. Premium estimation: pure premiums are computed using different simulation-based approaches, allowing robustness checks under various distributional assumptions.

The purpose is not to impose a single model, but rather to compare alternative options and justify the retained choices according to statistical performance, interpretability, and practical applicability.

In the empirical application, three representative rainfed regions are analyzed. A fourth region, Souss-Massa, was excluded due to structural inconsistencies in its irrigated production system, which generates yield dynamics that are not comparable with the rainfed areas. This ensures that the modeling framework remains consistent with the principle of risk homogeneity required in insurance design.

A critical component of parametric insurance is the construction of a synthetic drought index that captures the climatic determinants of yield losses. In the following, we present six alternative methods for combining standardized indicators (SPI, SPEI, NDVI) into a composite index. These methods range from simple linear rules to regularized regressions and machine learning techniques.

4.1. Drought Index Construction Methods

The construction of a composite drought index is a central step in the design of index-based insurance products. The objective is to synthesize climatic and vegetation information into a single indicator that reliably reflects the conditions leading to crop yield losses. To achieve this, six alternative methods were considered, ranging from simple linear rules to more sophisticated supervised learning algorithms. Each method is assessed in terms of theoretical properties, statistical performance, and practical relevance for insurance applications. This multi-method strategy reflects the diversity of statistical tools proposed for drought index construction and basis-risk reduction in recent contributions to index insurance design [66,40,13].

1. **Weighted Average Method (WA).** This baseline approach defines the composite index as a fixed linear combination of normalized agro-climatic indicators:

$$\text{CompositeIndex(WA)} = 0.3 \times \text{SPI} + 0.3 \times \text{SPEI} + 0.4 \times \text{NDVI}. \quad (6)$$

The weights balance drought severity (SPI, SPEI) and vegetation response (NDVI). Although transparent and easy to implement, this method does not account for empirical yield dynamics. Its main relevance lies in serving as a benchmark, particularly in data-scarce environments.

2. **Principal Component Analysis (PCA).** PCA is an unsupervised method that transforms correlated variables into orthogonal components maximizing overall variance. The first component (PC1) is retained as the composite index:

$$\text{CompositeIndex(PCA)} = Xw_1, \quad w_1 = \arg\max_{\|w\|=1} \text{Var}(Xw), \quad (7)$$

where $X \in \mathbb{R}^{n \times p}$ is the centered data matrix and w_1 the eigenvector associated with the largest eigenvalue. PCA is effective for dimensionality reduction and addressing multicollinearity, but it ignores yield information, which limits its actuarial relevance [52].

3. **Partial Least Squares (PLS).** PLS is a supervised alternative to PCA. It identifies latent components that maximize covariance with crop yield y:

$$t_1 = Xw_1, \quad w_1 = \underset{w}{\text{argmax}} \text{Cov}^2(Xw, y). \quad (8)$$

The first component is retained as the composite index, $\text{CompositeIndex(PLS)} = Xw_1$. By construction, PLS captures the yield-relevant part of the variability in the indicators [53].

4. **Lasso Regression.** Lasso regression performs regularization and variable selection simultaneously by penalizing the absolute size of coefficients. It solves:

$$\hat{\beta} = \underset{\beta}{\text{argmin}} \left\{ \frac{1}{2n} \| y - X\beta \|_2^2 + \lambda \| \beta \|_1 \right\}, \quad (9)$$

where $\lambda > 0$ controls the sparsity of the solution. The composite index is then defined as $\text{CompositeIndex(Lasso)} = X\hat{\beta}$. Lasso promotes parsimony and interpretability, which are desirable in insurance contexts requiring transparency and robustness [50].

5. **Ridge Regression.** Ridge regression is another regularized method that penalizes the squared magnitude of coefficients:

$$\hat{\beta} = \underset{\beta}{\text{argmin}} \left\{ \frac{1}{2n} \| y - X\beta \|_2^2 + \lambda \| \beta \|_2^2 \right\}. \quad (10)$$

Unlike Lasso, Ridge does not eliminate predictors but shrinks coefficients to handle multicollinearity. The composite index is given by $\text{CompositeIndex(Ridge)} = X\hat{\beta}$. This method provides stable estimates and is particularly valuable when datasets are noisy or limited in size

[51].

6. **Random Forest (RF).** Random Forest is a non-parametric ensemble method that aggregates predictions from multiple decision trees trained on bootstrap samples:

$$\hat{f}_{RF}(x) = \frac{1}{B}\sum_{b=1}^{B} \hat{f}_b(x). \quad (11)$$

RF captures nonlinear relationships and variable interactions, offering strong predictive performance. However, its limited transparency can pose challenges in actuarial practice, particularly in regulatory or pricing contexts where interpretability is essential [49].

All indices were standardized prior to comparison. Their predictive performance was assessed through correlation with yields and visual inspection using Generalized Additive Models (GAM). Comparative results are reported in Section 5.

4.2. Threshold Determination Methods for Index-Based Insurance

The definition of payout thresholds is a crucial component in the design of parametric insurance contracts. These thresholds determine the conditions under which indemnities are activated and thus directly influence both the effectiveness and the actuarial sustainability of the product. A well-calibrated threshold must strike a balance between three objectives: (i) ensuring sufficient risk coverage, (ii) maintaining affordability through premium control, and (iii) preserving transparency and actuarial justification. Similar trade-offs between coverage, affordability, and hedging effectiveness are documented in the literature on weather insurance design under climate variability and climate change [67].

Five approaches were evaluated to determine the trigger threshold for the composite drought index, each offering distinct statistical and interpretative properties.

1. **Return Period Approach.** In this method, the trigger threshold is defined as an empirical quantile of the historical index distribution, corresponding to a chosen return period. For instance, to represent a one-in-four-year drought event, the 25th percentile is used:

$$s_{RP} = Q_{0.25}(X), \quad (12)$$

where X denotes the composite index. This approach is simple to communicate and suitable when yield data are unavailable. However, it relies only on index frequency and does not guarantee actuarial alignment with actual losses.

2. **Quantile Regression Approach.** Quantile regression links the payout trigger to the conditional distribution of yields given the index [16]. The model is specified as:

$$Q_\tau(Y|X) = \beta_0 + \beta_1 X. \quad (13)$$

To determine the index value s_{QR} associated with a critical yield level y_s (e.g., 0.5 t/ha), the equation is inverted as:

$$s_{QR} = \frac{y_s - \hat{\beta}_0}{\hat{\beta}_1}. \quad (14)$$

This method is robust to outliers and skewed distributions and ensures that payouts are triggered under conditions of genuine agricultural distress. It is particularly suited to actuarial practice, as it links index activation directly to economically meaningful yield thresholds, in line with bio-economic assessments of drought risk and insurance design [68].

3. **Maximum Correlation Method.** This approach selects the threshold that maximizes the discriminatory power of the index with respect to yield losses. Specifically, the binary

indicator $Z_s = \mathbb{1}_{\{X<s\}}$ is constructed, and the threshold is defined as:

$$s_{CorrMax} = \underset{s}{\text{argmax}}|\text{corr}(Z_s, Y)|. \quad (15)$$

Although this method can identify thresholds with strong separation power, it may lead to very frequent payouts, undermining actuarial sustainability.

4. **Logistic Regression with ROC Analysis.** Here, the probability of a loss event $Z = \mathbb{1}_{\{Y<y_s\}}$ is modeled as a logistic function of the index:

$$\mathbb{P}(Z = 1|X) = \frac{1}{1+e^{-(\alpha+\beta X)}}. \quad (16)$$

The optimal threshold is then determined using Receiver Operating Characteristic (ROC) analysis by maximizing Youden's index:

$$t^* = \underset{t}{\text{argmax}}[TPR(t) - FPR(t)], \quad (17)$$

where TPR and FPR denote true and false positive rates. The corresponding index value s_{ROC} is obtained from the estimated logistic model. This approach balances predictive power and classification accuracy, but it may be unstable in small samples or when loss events are rare.

5. **Extreme Value Theory (EVT) Method.** EVT focuses on modeling rare and severe events by fitting a Generalized Pareto Distribution (GPD) to the lower tail of the index distribution. For a high threshold u (e.g., the 5% quantile), excesses $Y = X - u$ follow:

$$F_Y(y) = 1 - \left(1 + \xi \frac{y}{\sigma}\right)^{-1/\xi}, \quad y > 0, \quad (18)$$

with ξ and σ denoting shape and scale parameters. The return level associated with a recurrence period T is then:

$$s_{GPD} = u + \frac{\sigma}{\xi}\left[(T \cdot (1 - F(u)))^{\xi} - 1\right]. \quad (19)$$

This method provides a rigorous tail-risk perspective [69]. However, short regional time series limit its reliability, leading to potentially unstable estimates. It is therefore presented here mainly for comparison rather than operational use.

Among the five approaches, quantile regression provides the strongest actuarial arguments for practical implementation. It ensures consistency with yield-based definitions of loss and remains robust to skewness and outliers. Correlation-based and EVT-based methods may produce unstable or extreme thresholds, while ROC analysis can suffer from data imbalance. The final choice of threshold method is therefore guided by both statistical performance and actuarial interpretability, and will be assessed in detail in the results section.

4.3. Pure Premium Estimation Methods

The estimation of pure premiums in index-based insurance critically depends on both the functional form of the indemnity rule and the statistical behavior of the selected drought index. To ensure robustness and generalizability, premiums are evaluated under different distributional assumptions. Central to this process are two parameters of the indemnity function: the trigger threshold s_d, at which indemnities start, and the exhaustion threshold s_e, at which the maximum indemnity $I_{\max}$ is reached. Similar pricing issues arise in the valuation of weather derivatives and

crop insurance portfolios, where statistical modeling choices must remain consistent with the structure of economic exposure [18,70].

Capped Linear Indemnity Function. The indemnity function is defined according to the sign of the correlation between the selected index and agricultural yield:

- If the correlation is positive, the index is interpreted as a performance indicator, with higher values reflecting better yields. The indemnity decreases as the index increases.
- If the correlation is negative, the index acts as a risk indicator, with higher values reflecting worse yields. The indemnity increases as the index increases.

Formally, the capped linear indemnity function is specified as follows:

For a positively correlated index:

$$\text{Indemnity}(x) = \begin{cases} I_{\max}, & x \le s_e, \\ I_{\max} \cdot \frac{s_d - x}{s_d - s_e}, & s_e < x < s_d, \\ 0, & x \ge s_d, \end{cases} \tag{20}$$

For a negatively correlated index:

$$\text{Indemnity}(x) = \begin{cases} 0, & x \le s_d, \\ I_{\max} \cdot \frac{x - s_e}{s_d - s_e}, & s_d < x < s_e, \\ I_{\max}, & x \ge s_e. \end{cases} \tag{21}$$

Here, x denotes the drought index, s_d is the trigger threshold, s_e the exhaustion threshold, and $I_{\max}$ the maximum indemnity. The parameter $I_{\max}$ is customizable and may be adjusted depending on crop type, regional policy, or budgetary constraints.

While alternative non-linear indemnity schedules (such as convex or piecewise-linear functions) can reduce basis risk, we retain the capped linear form for reasons of transparency, simplicity, and ease of implementation. This choice avoids additional parameters and supports stable calibration given the limited length of regional time series. The trade-off is acknowledged; accordingly, the linear schedule is reported as the operational baseline, consistent with actuarial practice.

Definition of Agricultural Loss. For calibration purposes, yield shortfalls are expressed in monetary terms. In region r, the physical yield threshold y_r^* is defined as a fraction of the regional mean yield:

$$y_r^* = \theta_r \cdot \bar{y}_r, \quad \theta_r \in [0.7,1], \tag{22}$$

where $\bar{y}_r$ is the average yield. The yield loss is then defined as:

$$\text{YieldLoss}(y) = \max(0, y_r^* - y), \tag{23}$$

and monetized using the regional reference price P_r (MAD/ton):

$$\text{Loss}(y) = \text{YieldLoss}(y) \times P_r. \tag{24}$$

This ensures consistency between indemnities and actual losses, both expressed in MAD/ha.

Relative Basis Risk. Basis risk is quantified through the relative basis risk (RBR) and its root mean squared version (RBR-RMSE). Conditional on loss occurrence:

$$RBR = \frac{\mathbb{E}[|\ \text{Indemnity} - \text{Lo}\quad ||\text{Loss}>\ ]}{\mathbb{E}[\text{Loss}|\text{Loss}>0]} \times 100, \tag{25}$$

$$RBR - RMSE = \frac{\sqrt{\mathbb{E}[(\text{Indemnity} - \text{Loss})^2|\text{Loss}>0]}}{\mathbb{E}[\text{Loss}|\text{Loss}>0]} \times 100. \tag{26}$$

Both measures are used to calibrate the exhaustion threshold s_e through numerical optimization,

ensuring actuarial soundness.

Simulation-Based Premium Estimation. Four complementary approaches are used to estimate pure premiums:

1. *Empirical Simulation.* Historical index values are directly passed through the indemnity function. This non-parametric method assumes that past climate realizations are representative of future risks.
2. *Monte Carlo Simulation.* A normal distribution is fitted to the index series. Synthetic values are generated (10,000 draws) and processed through the indemnity function. This approach is simple but sensitive to the normality assumption.
3. *Bootstrap Simulation.* To account for sampling uncertainty, 1,000 bootstrap samples are drawn with replacement from the observed index values. Premiums are computed for each sample, producing confidence intervals.
4. *Kernel Density Estimation (KDE).* A smooth, non-parametric density is estimated from the historical index distribution. Synthetic values (10,000 draws) are generated and used for premium estimation. KDE avoids restrictive parametric assumptions and is flexible under small datasets.

The exhaustion threshold s_e is selected by minimizing the RBR. The resulting pure premiums are expressed in MAD/ha and presented in Section 5.

These methods were applied to the three selected rainfed regions. Their empirical implementation, including the comparative evaluation of composite indices, the calibration of payout thresholds, and the estimation of pure premiums, is reported in Section 5, together with an assessment of residual basis risk.

5. Results and Discussion

This section reports the main empirical findings of the study, structured around four components: the predictive performance of the constructed drought indices, the calibration of the capped indemnity function, the estimation of pure premiums under alternative statistical assumptions, and the quantification of basis risk.

The analysis relies on annual data from three rainfed agro-climatic regions (Fès-Meknès, Marrakech-Safi, and Oriental) selected for their homogeneity in cropping systems and climatic conditions. The original dataset also included Souss-Massa, but this region was excluded due to its structurally distinct yield levels and irrigated production system, which produced unstable and non-linear index-yield relationships, as explained in Section 3. The predictive models are built on composite indices derived from climate and vegetation indicators (Section 4), and simulation-based techniques are applied to assess the robustness of premium estimates and the reliability of index-based insurance schemes.

Each subsection presents the main statistical results and discusses their implications for the operational design of agricultural insurance products.

5.1. Comparative Performance of Composite Indices

To evaluate the suitability of alternative index construction methods for parametric insurance, we assessed their ability to explain observed yield variability across the three selected rainfed regions. Performance metrics include the Pearson correlation with yields, the Root Mean Squared Error (RMSE),

the Mean Absolute Error (MAE), and the coefficient of determination (R^2). Table 5 reports the comparative predictive performance of the candidate synthetic indices.

Table 5. Predictive performance of synthetic drought indices.

Method	Correlation	RMSE	MAE	R^2
Lasso Index	**0.900**	**0.256**	**0.215**	**0.811**
Ridge Index	0.899	0.257	0.215	0.809
PLS Index	0.882	0.277	0.227	0.778
RF Index	0.864	0.296	0.238	0.747
PCA Index	-0.261	0.568	0.473	0.068
Weighted Average Index	0.102	0.586	0.503	0.010

Supervised learning techniques clearly outperform both unsupervised and fixed-weight approaches. The Lasso and Ridge indices achieve the highest correlations with observed yields (around 0.90) and the lowest prediction errors, closely followed by Partial Least Squares (PLS). These findings highlight the value of regularized and supervised methods for synthesizing climate–vegetation information into predictive drought indicators.

By contrast, unsupervised methods such as PCA and fixed-weight averages perform poorly. The PCA index, in particular, shows a negative correlation with yields and very low explanatory power ($R^2 \approx 0.07$), underscoring its limited relevance for yield-based risk modeling.

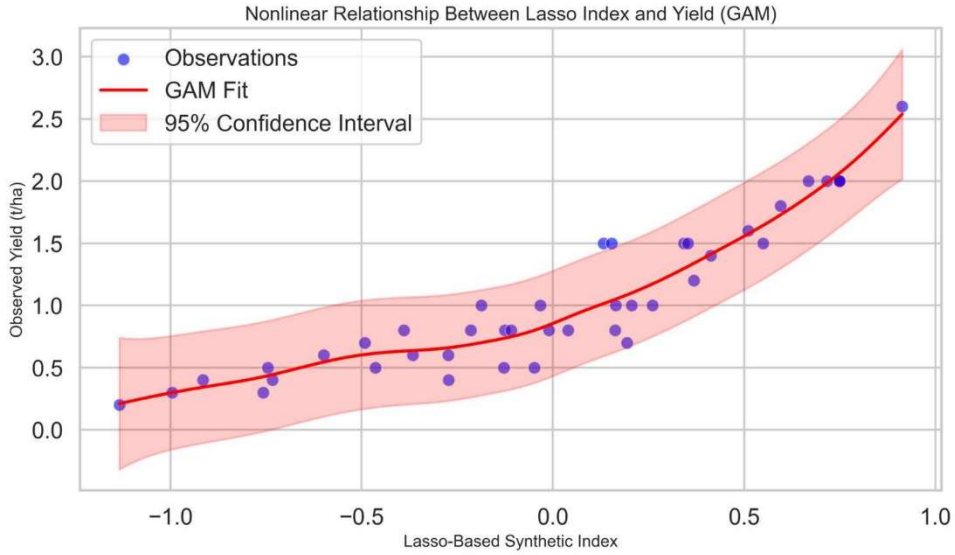

Figure 6. Nonlinear relationship between Lasso-based synthetic index and yield (GAM smoothing).

Given its strong predictive performance and interpretability, the Lasso index is retained for the subsequent design of parametric insurance. Figure 6 further depicts its relationship with yields, using a smoothed trend obtained from a Generalized Additive Model (GAM).

For visual comparison across all candidate indices, Figure 7 presents scatterplots of each synthetic index against observed yields. The superiority of supervised methods is again evident from the clarity and slope of their relationships with agricultural performance.

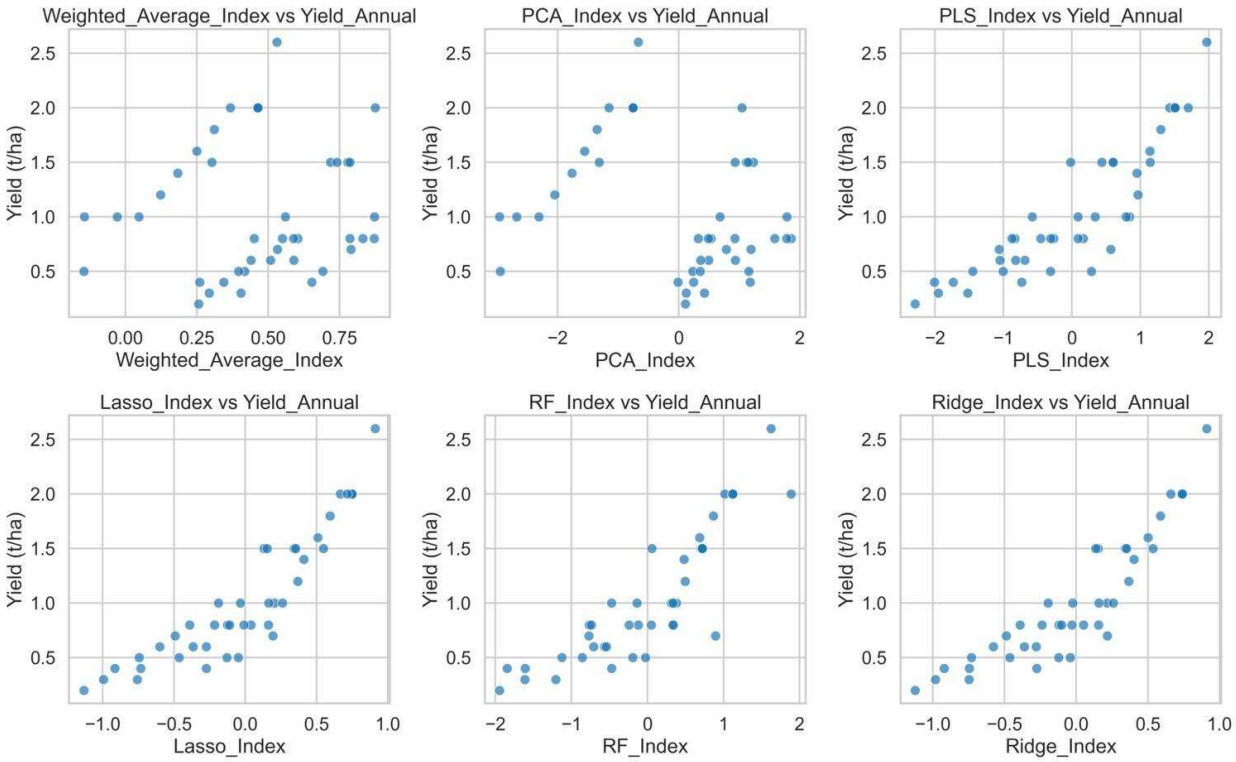

Figure 7. Observed yield vs. synthetic drought indices.

5.2. Threshold Selection for Lasso-Based Index

Once the Lasso-based index was identified as the most reliable predictor of agricultural yields, several statistical methods were evaluated to determine the optimal trigger threshold s_d, defined as the value below which indemnity payments are activated. The threshold must balance two essential criteria: actuarial sustainability and agronomic sensitivity.

Five alternative strategies are applied to the filtered dataset covering the three homogeneous rainfed regions. Their outcomes are summarized in Table 6 and visualized in Figure 8. This comparative assessment aims to identify a robust and interpretable method for defining s_d, which is subsequently applied to each region individually in the calibration stage.

Table 6. Indemnity threshold calibration across methods.

Method	Threshold s_d	Activation Frequency (%)	Average Yield (Indemnified)	Yield Gap
Maximum Correlation	0.272	69.05	0.72	**1.06**
Quantile Regression	-0.164	35.71	0.54	0.79
Return Period (4 years)	-0.343	26.19	0.48	0.76
GPD EVT (unstable)	-1.066	2.38	0.20	0.87
Logistic ROC	-1.133	0.00	–	–

The Maximum Correlation method identifies a threshold ($s_d = 0.272$) that achieves the highest yield gap (1.06 t/ha), but at the cost of an activation frequency close to 70%. Such frequent payouts would imply unsustainably high premiums.

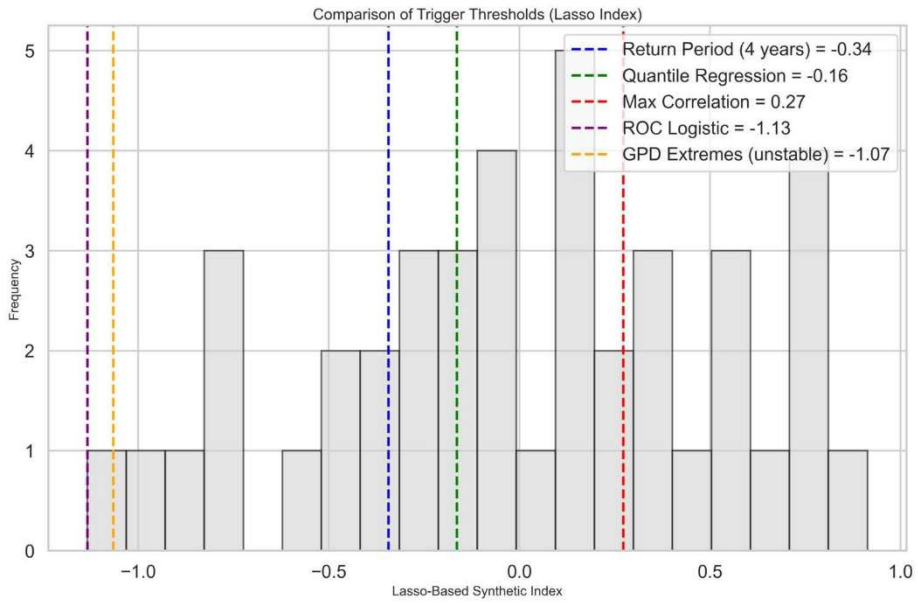

Figure 8. Trigger thresholds for Lasso-based index using different methods.

By contrast, the EVT and ROC methods yield thresholds that are excessively conservative, with negligible activation frequencies, making them operationally irrelevant in practice.

Quantile Regression offers a balanced compromise. With $s_d = -0.164$, payouts are triggered in 36% of observed years, targeting moderate droughts while preserving actuarial viability. The average yield under indemnified events (0.54 t/ha) is substantially lower than non-indemnified yields, ensuring responsiveness to meaningful agricultural losses.

In light of these considerations, Quantile Regression is retained as the preferred approach for threshold calibration. Its statistical coherence, moderate activation rate, and clear agronomic interpretation make it the most appropriate option for index-based drought insurance in Moroccan rainfed agriculture.

5.3. Premium Estimation and Basis Risk Evaluation

Following the selection of the Lasso-based index and the calibration of region-specific trigger thresholds (s_d) using quantile regression, exhaustion thresholds (s_e) were determined by minimizing the Relative Basis Risk (RBR) across observed loss years. Since the empirical correlation between the index and yields is positive in our application, the capped indemnity function is strictly decreasing. More generally, the functional form must always reflect the sign of correlation: decreasing payouts for performance-type indices and increasing payouts for risk-type indices.

Pure premiums were then estimated using four complementary simulation approaches: empirical simulation, Monte Carlo simulation, bootstrap simulation, and kernel density estimation (KDE). Applying multiple methods makes it possible to assess the stability of premium levels under different distributional assumptions and to evaluate the robustness of the pricing framework. The comparative results for the three regions are reported in Table 7.

Across all methods, KDE yields the highest premium levels, reflecting the smooth approximation of the empirical distribution. Monte Carlo produces lower premiums because of its normality assumption, while bootstrap and empirical simulations provide intermediate and closely aligned estimates, reinforcing the robustness of the evaluation.

Table 7. Premium estimates and basis risk by region.

Region	Fès-Meknès	Marrakech-Safi	Oriental
Trigger Threshold s_d	0.358	-0.285	-0.613
Exhaustion Threshold s_e	0.255	-0.576	-1.115
Max Indemnity $I_{\max}$ (MAD/ha)	1000.00	1000.00	1000.00
Pure Premium (Empirical)	281.57	212.41	165.08
Pure Premium (Monte Carlo)	275.43	234.77	128.27
Pure Premium (Bootstrap)	284.44	208.06	161.90
Pure Premium (KDE)	304.03	265.60	176.18
RBR (%)	45.18	17.57	37.61
RBR-RMSE (%)	86.97	24.43	50.02

Actuarially, Marrakech-Safi emerges as the most favorable region: premiums remain moderate and both RBR indicators are the lowest (RBR = 17.57%, RBR-RMSE = 24.43%). This reflects strong alignment between the index and actual yield losses, with limited residual basis risk.

By contrast, Fès-Meknès records the highest RBR values (45.18% and 86.97%), likely due to greater yield volatility, complex agro-climatic interactions, or nonlinear effects near the payout boundary. These findings underline the importance of region-specific calibration for both thresholds and premium-setting.

Overall, the capped indemnity function provides a stable and interpretable pricing framework. The results obtained across the four simulation methods demonstrate robustness to distributional assumptions, while offering actuarial flexibility through parameters such as $I_{\max}$ and s_e. For operational implementation in this application, KDE is retained as the primary pricing reference given its empirical robustness and flexibility.

5.4. Visual Evaluation and Regional Considerations

To complement the quantitative assessment of premium levels and residual basis risk, this subsection provides a visual inspection of the capped indemnity functions calibrated for each region.

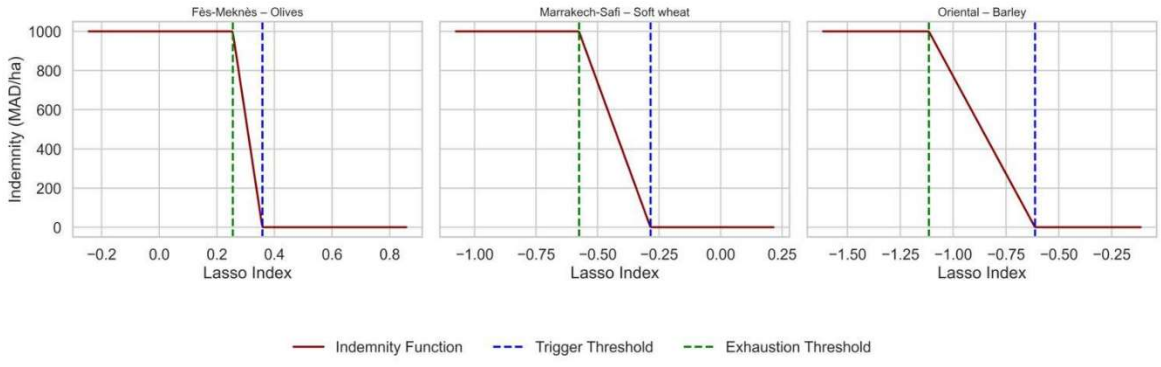

Figure 9. Capped linear indemnity functions calibrated for each region.

Figure 9 presents the capped linear indemnity functions defined by the trigger threshold s_d, the exhaustion threshold s_e, and the maximum indemnity $I_{\max} = 1000$ MAD/ha. The slope of each function reflects the interval of index values over which partial payouts occur.

The comparison highlights clear regional contrasts. In Fès-Meknès, the interval between s_d and s_e is relatively narrow, resulting in a steep slope and an abrupt transition from zero to maximum

indemnity. By contrast, the Oriental region displays a wider interval, producing a flatter slope and smoother compensation across moderate yield shocks. Marrakech-Safi lies in between, with intermediate values of s_d and s_e.

These patterns are consistent with the statistical evidence and illustrate how differences in calibration translate into distinct indemnity dynamics. While crop-specific characteristics (e.g., olive variability in Fès-Meknès, soft wheat stability in Marrakech-Safi, or barley sensitivity in Oriental) may contribute to these outcomes, the dominant driver is the statistical relationship between the Lasso-based index and observed yields in each region.

This heterogeneity underscores the limits of a uniform contract design. Region-specific calibration is essential to ensure actuarial soundness and to maintain perceived fairness among farmers.

Finally, data availability plays an additional role. Short time series restrict the application of extreme value methods such as GPD, which require larger samples for reliable estimation. By contrast, non-parametric approaches (KDE, bootstrap) remain effective with limited data. The adoption of the Lasso-based index also reflects a compromise between interpretability and predictive robustness under these constraints.

In summary, the visual analysis confirms that regional differentiation in climatic exposure, crop dynamics, and data structure requires tailored calibration strategies for both index construction and indemnity design. Such an approach improves the alignment between payouts and actual losses while enhancing the actuarial credibility and farmer acceptance of index-based insurance products.

5.5. Contextual Limitations and Implications for Scalability

The performance of parametric insurance products is inherently shaped by the agro-climatic context, the quality and length of available datasets, and the degree of homogeneity across insured risks. In this study, the analysis was restricted to three rainfed regions (Fès-Meknès, Marrakech-Safi, and Oriental), characterized by relatively low yields and distinct yet comparable climatic conditions. The selected crops (olive, soft wheat, and barley) differ in their sensitivity to water stress, but share seasonal dynamics that facilitate index calibration and cross-regional comparison.

Nevertheless, the heterogeneity observed across Moroccan regions raises important concerns for scalability. Irrigated areas such as Souss-Massa, dominated by citrus production, display structurally higher and more stable yields with limited sensitivity to climatic indices. Including such regions in pooled insurance schemes would undermine the principle of risk homogeneity, inflate basis risk, and weaken actuarial soundness in premium setting.

Data availability further constrains model complexity. With a limited number of annual observations per region over the 2010–2023 period, the potential for advanced statistical modeling is limited. While supervised learning techniques such as Lasso, Ridge, and PLS perform robustly in small samples, more flexible approaches (e.g., random forests or neural networks) risk overfitting and unstable feature selection. Parsimony and interpretability therefore remain essential design principles for operational products intended to satisfy both actuarial and regulatory standards.

These limitations highlight the importance of tailoring parametric insurance not only to agro-climatic exposure but also to crop-specific characteristics, data constraints, and market context. From a methodological perspective, potential extensions include hierarchical or Bayesian frameworks to partially pool information across regions, or spatial econometric approaches to capture inter-regional

dependencies while preserving risk differentiation. Such strategies could help scale up coverage while maintaining statistical coherence and actuarial soundness.

Overall, the empirical findings underline both the potential and the limits of index-based drought insurance in Moroccan rainfed agriculture. The Lasso-based composite index demonstrated strong predictive power, while quantile regression proved to be the most robust method for calibrating payout thresholds. The capped indemnity function ensured actuarial soundness, and the application of multiple premium estimation techniques confirmed the stability of results across varying modeling assumptions.

At the same time, residual basis risk and regional heterogeneity stress the need for region-specific calibration rather than one-size-fits-all designs. The scalability of such products remains conditional on data availability, crop-specific characteristics, and the principle of risk homogeneity. Future research should explore methods for pooling information across regions while preserving risk differentiation and actuarial fairness, thereby extending the operational viability of parametric insurance schemes.

6. Conclusion

This study develops an integrated and transparent framework for designing and pricing parametric drought insurance, combining synthetic index construction, threshold calibration, and simulation-based premium estimation within a single coherent pipeline. By aligning statistical modelling with actuarial requirements, the approach offers a structured method for translating climatic and satellite-based indicators into operational insurance products.

Beyond its methodological contribution, the framework carries several practical implications for drought risk management. For insurers and reinsurers, the explicit calibration of thresholds and the systematic evaluation of basis risk support the design of products that are actuarially sound, transparent, and easier to justify in supervisory or rating discussions. For policymakers and agricultural agencies, the framework provides a consistent basis for evaluating the cost of drought protection, quantifying trade-offs between affordability and coverage, and improving the targeting and transparency of public support schemes. The reliance on externally verifiable climate and vegetation indicators also reduces information asymmetries, which is particularly relevant in environments where traditional loss adjustment remains costly or difficult to audit.

At the same time, several limitations must be acknowledged when interpreting the empirical results. The analysis relies on approximately fourteen years of annual observations, which constrains the use of more flexible machine learning models and limits the precision of tail-risk estimation. The empirical application focuses on three rainfed regions where climatic exposure and crop cycles are relatively comparable; irrigated regions such as Souss-Massa, with structurally distinct production patterns, fall outside the scope of this modelling framework and illustrate the importance of risk homogeneity for parametric design. Residual basis risk remains non-negligible in some areas, reflecting both nonlinearities in climate–yield relationships and unobserved agronomic factors. These limitations underline the need for continuous refinement of indices, improved agronomic data, and complementary policy tools when designing large-scale drought protection strategies.

Despite these constraints, the modular structure of the framework enhances its relevance beyond the Moroccan context. Because each stage of the pipeline, from index construction to threshold calibration and premium estimation, can be re-estimated with alternative climatic, satellite and yield datasets, the methodology remains transferable to a wide range of agricultural systems. Countries

with reliable climatic archives, adequate remote sensing coverage and established agricultural insurance initiatives can adapt the framework by adjusting crop calendars, regional aggregations and production conditions. Its transparency and reproducibility also make it suitable for integration into public–private partnerships, national risk-financing strategies and donor-supported climate resilience programs.

Ultimately, this framework demonstrates how statistical and actuarial techniques can be combined to develop parametric insurance products that are both operational and scientifically grounded. By emphasizing clarity, methodological consistency, and adaptability, it provides not only an academic contribution but also a practical foundation for extending drought protection in data-constrained environments and strengthening resilience in agricultural systems exposed to increasing climatic uncertainty

Appendix A. Temporal Dynamics of Key Agro-Climatic Variables

This appendix provides a detailed visualization of the temporal trends for the main agro-climatic and agronomic variables: precipitation, NDVI, evapotranspiration (ET), and crop yields, over the 2010–2023 period across the four study regions. These series highlight seasonal patterns, interannual variability, and structural contrasts between rainfed and irrigated zones.

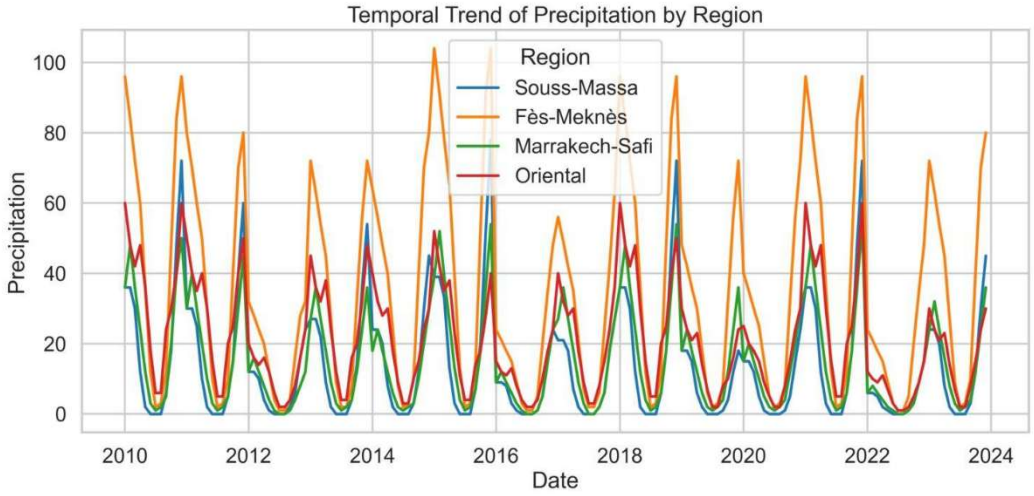

Figure A1. Monthly precipitation trends across regions (2010–2023).

A pronounced seasonality is observed in Figure A1, with precipitation peaks generally occurring in autumn and spring. Interannual variability is especially marked in rainfed regions, reflecting the stochastic nature of rainfall in semi-arid environments.

Figure A2 shows NDVI patterns that reflect well-defined phenological cycles, particularly in rainfed zones, where vegetation closely follows rainfall seasonality. In Souss-Massa, NDVI remains more stable and elevated, consistent with the continuous productivity of irrigated citrus crops.

Figure A3 confirms the strong seasonality of ET, largely synchronized with solar radiation and temperature. Although mean ET levels are similar across regions, variance is notably lower in Souss-Massa, highlighting the moderating effect of irrigation.

Figure A4 illustrates the structural divergence between rainfed and irrigated systems. While yields in Fès-Meknès, Marrakech-Safi, and Oriental remain stable over time with low variance, Souss-

Massa shows both significantly higher levels and greater year-to-year fluctuations, attributable to its intensive horticultural production under irrigation.

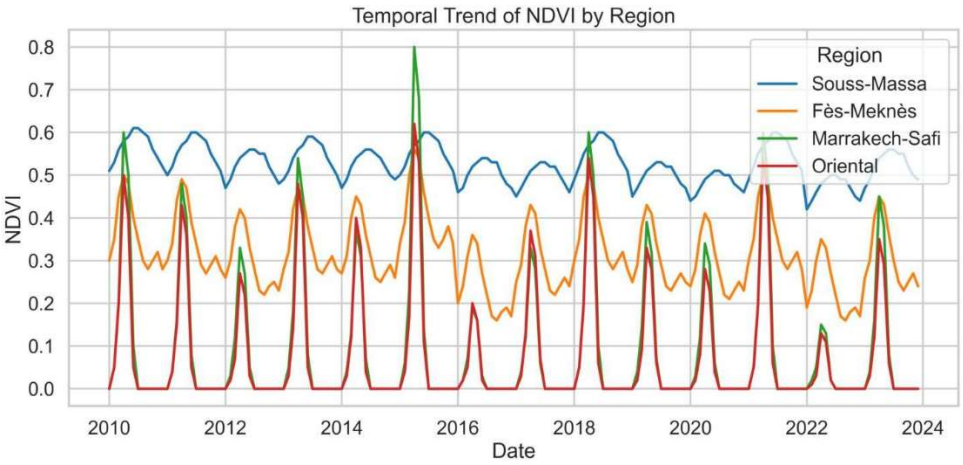

Figure A2. Monthly NDVI trends across regions (2010–2023).

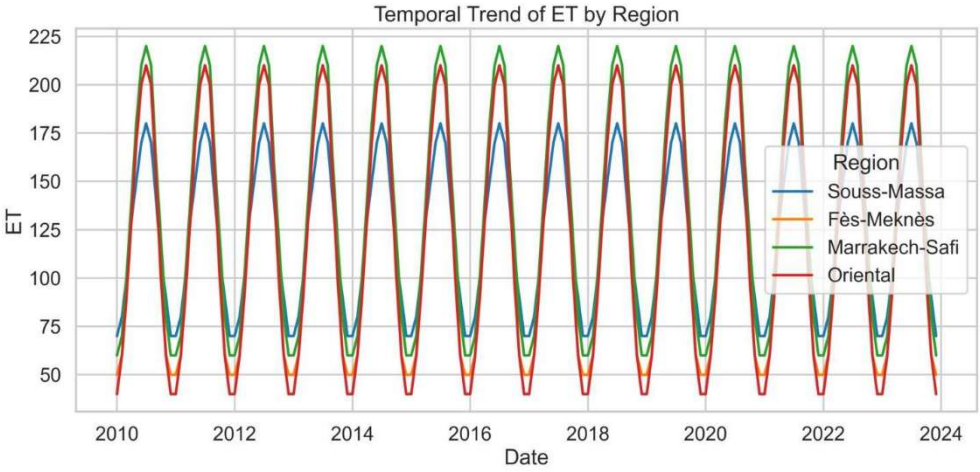

Figure A3. Monthly evapotranspiration trends across regions (2010–2023).

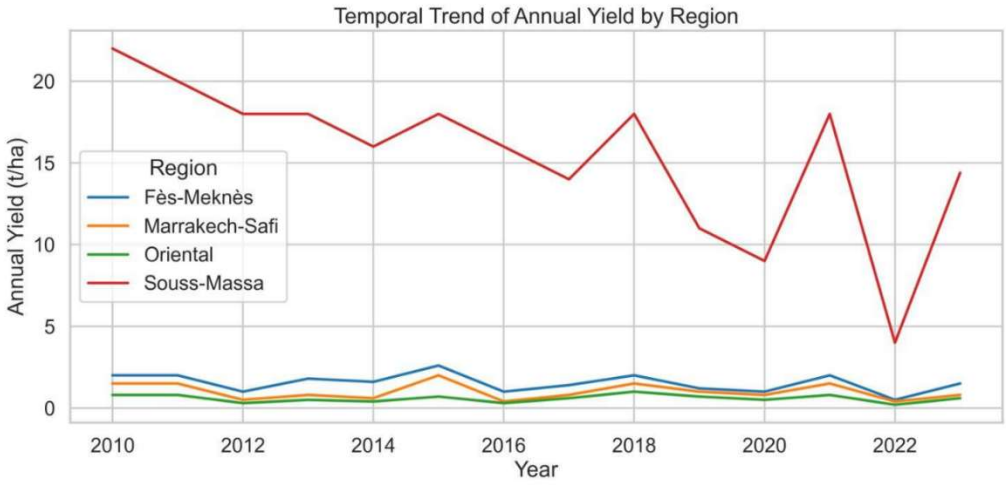

Figure A4. Annual crop yield trends across regions (2010–2023).

Appendix B. Regional Distributions of Climatic and Vegetation Variables

This appendix presents the empirical distributions of three key agro-climatic variables: NDVI, monthly precipitation, and evapotranspiration, across the four study regions. These visualizations complement the summary statistics discussed in the main text by revealing distributional shape, dispersion, and spatial contrasts.

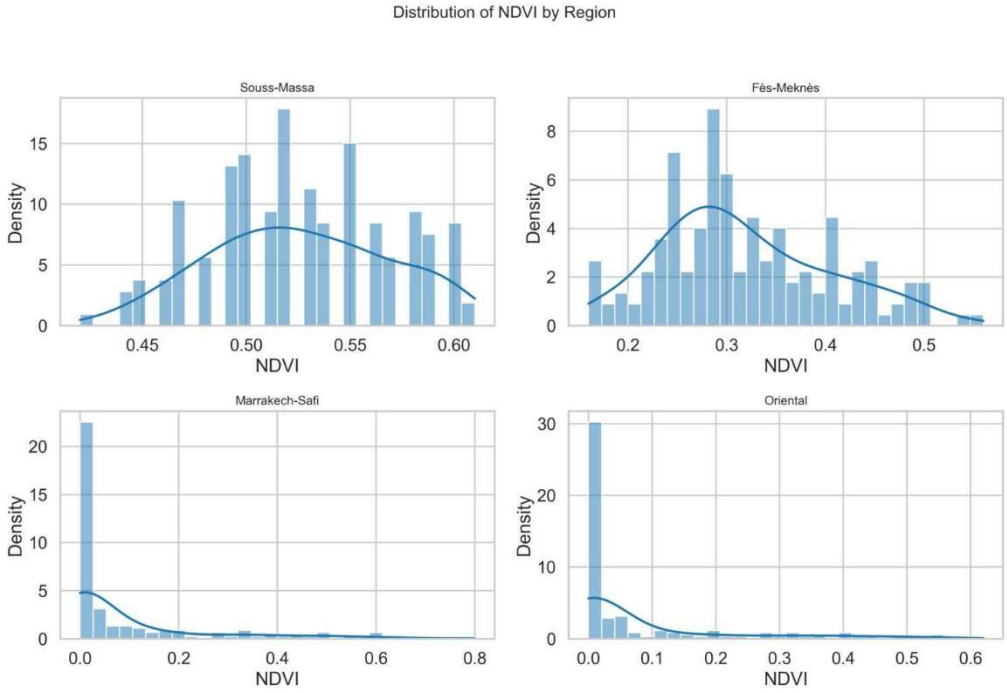

Figure B1. NDVI distributions across regions (2010–2023).

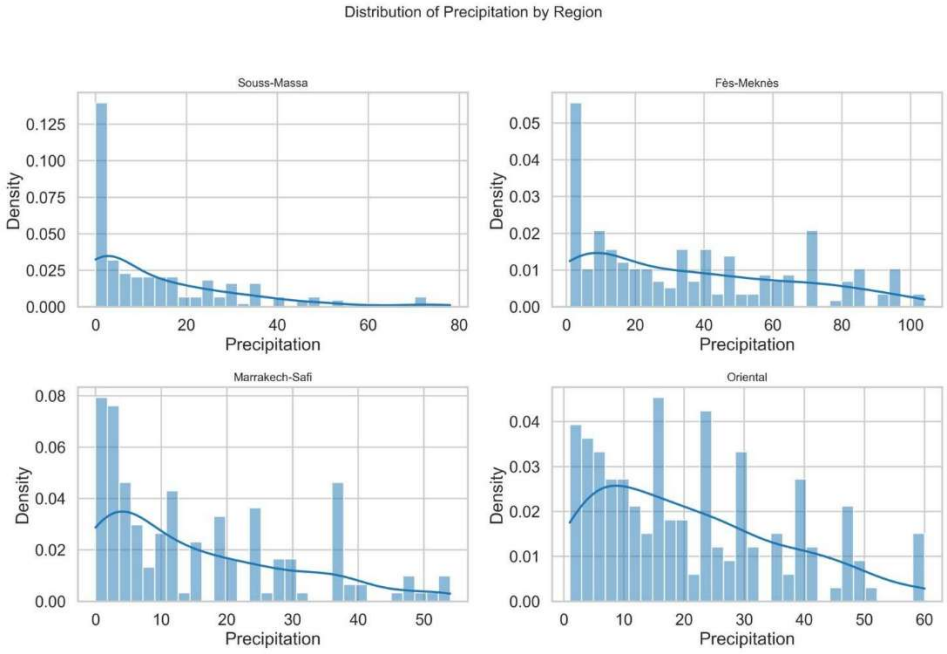

Figure B2. Monthly precipitation distributions across regions (2010–2023).

The NDVI distributions (Figure B1) show a bimodal pattern in rainfed regions, consistent with alternating phases of vegetative dormancy and active growth. In contrast, Souss-Massa exhibits a more concentrated distribution, reflecting the stability of irrigated citrus cultivation.

Monthly precipitation (Figure B2) displays a strongly right-skewed distribution across all regions, characterized by a predominance of low rainfall values and occasional extreme precipitation events. This irregularity is typical of Morocco's semi-arid climate.

The evapotranspiration boxplots (Figure B3) indicate lower medians and reduced variability in Souss-Massa, reflecting the stabilizing effect of irrigation. Rainfed regions reveal higher dispersion, driven by seasonal and interannual meteorological fluctuations. A histogram version was initially considered but was discarded due to limited readability and lack of informative value.

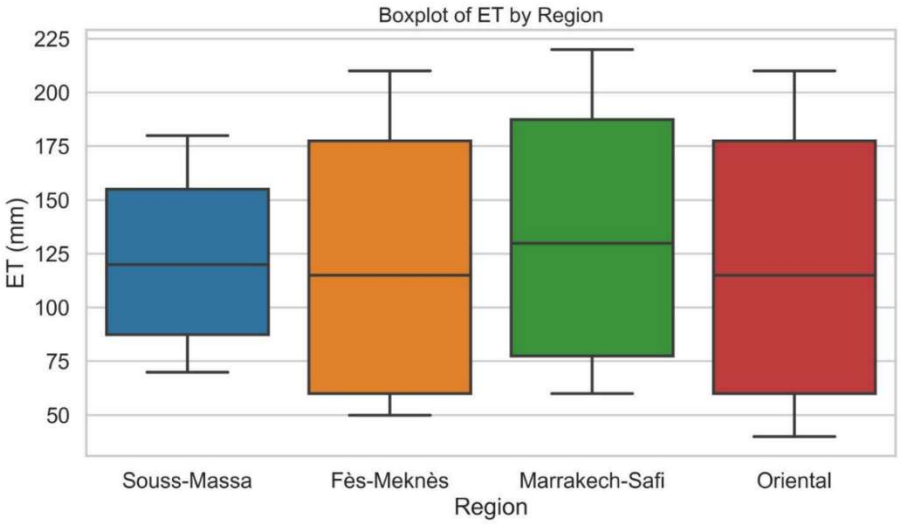

Figure B3. Monthly evapotranspiration distributions across regions (2010–2023).

Appendix C. Drought Index Diagnostics

This appendix complements the main analysis by providing model validation diagnostics for the drought indices used in the study. While the SPI and SPEI time series were already presented and interpreted in the main text (see Figures 4 and 5), we focus here on assessing the statistical adequacy of the distributional assumptions underpinning the SPEI construction.

To evaluate the goodness-of-fit of the Pearson Type III distribution used for standardizing the 3-month climatic water balance variable ($D_3 = P - \mathrm{ET}$), we compare the empirical cumulative distribution functions (CDFs) with their theoretical counterparts. Figure C1 shows these comparisons for the four study regions.

The graphs indicate that the Pearson Type III distribution provides a reasonable approximation to the empirical data in all regions. However, slight deviations are observed in the lower tail, particularly in Souss-Massa and Oriental, suggesting a mild underestimation of the frequency of extreme drought events. These discrepancies are consistent with previous studies on the limitations of parametric distributions under hydrometeorological stress.

Despite these deviations, the Pearson Type III remains a widely used model in drought analysis due to its flexibility and its ability to capture skewness. The retention of extreme values in the lower tail is deliberate, as it ensures that the probabilistic structure of the index remains consistent with

real-world drought severity, a crucial aspect for developing parametric insurance schemes based on tail risk.

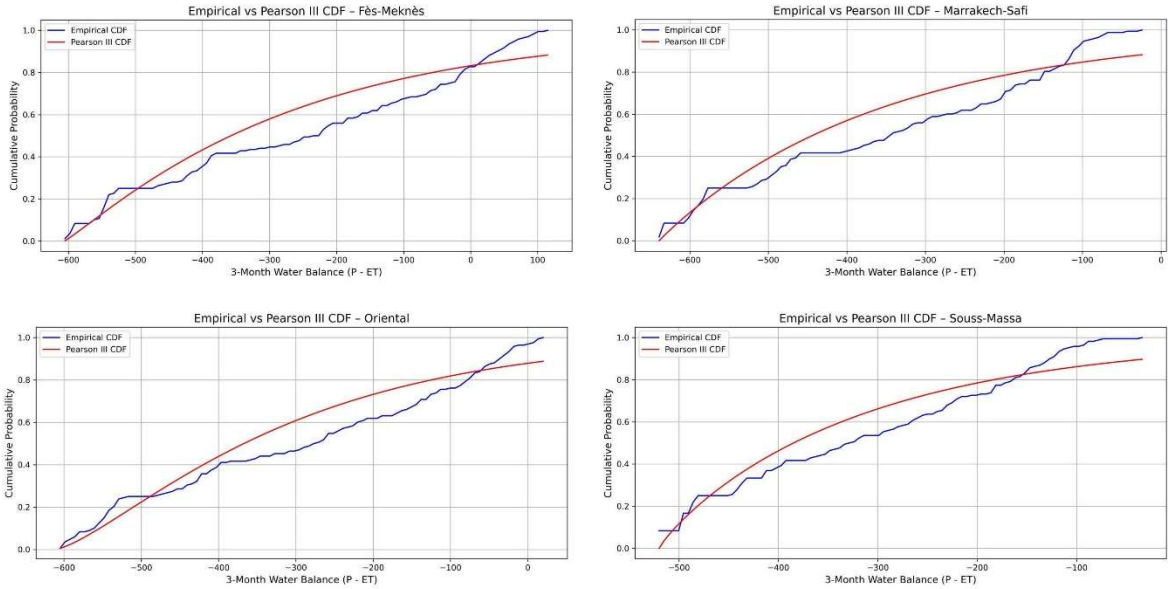

Figure C1. Empirical versus Pearson Type III cumulative distribution functions (CDFs) of the 3-month water balance D_3 across the four regions.

References

[1] Kim, W., Iizumi, T., & Nishimori, M. (2019). Global patterns of crop production losses associated with droughts from 1983 to 2009. *Journal of Applied Meteorology and Climatology*, 58(6), 1233–1244. https://doi.org/10.1175/JAMC-D-18-0174.1

[2] Lesk, C., Coffel, E., Winter, J., Ray, D., Zscheischler, J., Seneviratne, S. I., & Horton, R. (2021). Stronger temperature–moisture couplings exacerbate the impact of climate warming on global crop yields. *Nature Food*, 2(9), 683–691. https://doi.org/10.1038/s43016-021-00341-6

[3] Hertel, T. W., & Rosch, S. D. (2010). Climate change, agriculture, and poverty. Applied Economic Perspectives and Policy, 32(3), 355–385. https://doi.org/10.1093/aepp/ppq016

[4] Gornall, J., Betts, R., Burke, E., Clark, R., Camp, J., Willett, K., & Wiltshire, A. (2010). Implications of climate change for agricultural productivity in the early twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1554), 2973–2989. https://doi.org/10.1098/rstb.2010.0158

[5] Goodwin, B. K. (2001). Problems with market insurance in agriculture. *American Journal of Agricultural Economics*, 83(3), 643–649. http://www.jstor.org/stable/1245093

[6] Mahul, O. (2001). Optimal insurance against climatic experience. *American Journal of Agricultural Economics*, 83(3), 593–604. https://doi.org/10.1111/0002-9092.00180

[7] Skees, J. R. (2008). Innovations in index insurance for the poor in lower income countries. *Agricultural and Resource Economics Review*, 37(1), 1–15. https://doi.org/10.1017/S106828050000209

[8] Barnett, B. J., Barrett, C. B., & Skees, J. R. (2008). Poverty traps and index-based risk transfer products. *World Development*, 36(10), 1766–1785. https://doi.org/10.1016/j.worlddev.2007.10.016

[9] Carter, M., de Janvry, A., Sadoulet, E., & Sarris, A. (2017). Index insurance for developing country agriculture: A reassessment. *Annual Review of Resource Economics*, 9(1), 421–438. https://doi.org/10.1146/annurev-resource-100516-053352

[10] Jensen, N., & Barrett, C. (2016). Agricultural index insurance for development. *Applied Economic Perspectives and Policy*, 39(2), 199–219. https://doi.org/10.1093/aepp/ppw022

[11] Nshakira-Rukundo, E., Kamau, J. W., & Baumüller, H. (2021). Determinants of uptake and strategies to improve agricultural insurance in Africa: A review. *Environment and Development Economics*, 26(5–6), 605–631. https://doi.org/10.1017/S1355770X21000085

[12] Kath, J., Mushtaq, S., Henry, R., Adeyinka, A., & Stone, R. (2018). Index insurance benefits agricultural producers exposed to excessive rainfall risk. *Weather and Climate Extremes*, 22, 1–9. https://doi.org/10.1016/j.wace.2018.10.003

[13] Dalhaus, T., Musshoff, O., & Finger, R. (2018). Phenology information contributes to reduce temporal basis risk in agricultural weather index insurance. *Scientific Reports*, 8(1), 46. https://doi.org/10.1038/s41598-017-18656-5

[14] Gornott, C., & Wechsung, F. (2016). Statistical regression models for assessing climate impacts on crop yields: A validation study for winter wheat and silage maize in Germany. *Agricultural and Forest Meteorology*, 217, 89–100. https://doi.org/10.1016/j.agrformet.2015.10.005

[15] Hillier, J. K., Matthews, T., Wilby, R. L., & Murphy, C. (2020). Multi-hazard dependencies can increase or decrease risk. *Nature Climate Change*, 10(7), 595–598. https://doi.org/10.1038/s41558-020-0832-y

[16] Koenker, R. (2005). *Quantile Regression*. Cambridge University Press. https://doi.org/10.1002/9781118445112.stat07557

[17] Prokopchuk, O., Prokopchuk, I., Mentel, G., & Bilan, Y. (2020). Parametric insurance as innovative development factor of the agricultural sector of economy. *AGRIS on-line Papers in Economics and Informatics*, 12(3), 69–86. https://doi.org/10.7160/aol.2020.120307

[18] Müller, A., & Grandi, M. (2000). Weather derivatives: A risk management tool for weather-sensitive industries. *The Geneva Papers on Risk and Insurance. Issues and Practice*, 25(2), 273–287. https://www.jstor.org/stable/41952530

[19] Ben Salem, S., Ben Salem, A., Karmaoui, A., & Yacoubi Khebiza, M. (2023). Vulnerability of water resources to drought risk in southeastern Morocco: Case study of Ziz Basin. *Water*, 15(23), 4085. https://doi.org/10.3390/w15234085

[20] El-Hawari, J., El-Ghachi, M., Bouhafa, Y., & Sinbri, A.-E. (2023). Rainfall variability and the risk of climatic drought in the Tensift basin (Morocco). *Journal of Nature, Life and Applied Sciences*, 7(2), 92–102. https://doi.org/10.26389/AJSRP.W120123

[21] Atkinson, M. D., Kettlewell, P. S., Hollins, P. D., Stephenson, D. B., & Hardwick, N. V. (2005). Summer climate mediates UK wheat quality response to winter North Atlantic Oscillation. *Agricultural and Forest Meteorology*, 130(1–2), 27–37. https://doi.org/10.1016/j.agrformet.2005.02.002

[22] Xiao, D., Moiwo, J. P., Tao, F., Yang, Y., Shen, Y., Xu, Q., Liu, J., Zhang, H., & Liu, F. (2015). Spatiotemporal variability of winter wheat phenology in response to weather and climate variability in China. *Mitigation and Adaptation Strategies for Global Change*, 20(7), 1191–1202. https://doi.org/10.1007/s11027-013-9531-6

[23] Pettorelli, N., Vik, J. M., Mysterud, A., Gaillard, J.-M., Tucker, C. J., & Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, 20(9), 503–510. https://doi.org/10.1016/j.tree.2005.05.011

[24] Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C., Gao, F., Reed, B. C., & Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, 84(3), 471–475. https://doi.org/10.1016/S0034-4257(02)00135-9

[25] Sakamoto, T., Yokozawa, M., Toritani, H., Shibayama, M., Ishitsuka, N., & Ohno, H. (2005). A crop phenology detection method using time-series MODIS data. *Remote Sensing of Environment*, 96(3–4), 366–374. https://doi.org/10.1016/j.rse.2005.03.008

[26] McKee, T. B., Doesken, N. J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology* (pp. 179–184), Anaheim, CA, United States. American Meteorological Society. https://climate.colostate.edu/pdfs/relationshipofdroughtfrequency.pdf

[27] Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A multi-scalar drought index sensitive to global warming: The Standardized Precipitation Evapotranspiration Index. *Journal of Climate*, 23(7), 1696–1718. https://doi.org/10.1175/2009JCLI2909.1

[28] Verdin, J. P., Funk, C., Senay, G., & Choularton, R. (2005). Climate science and famine early warning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1463), 2155–2168. https://doi.org/10.1098/rstb.2005.1754

[29] Hazaymeh, K., & Hassan, Q. K. (2016). Remote sensing of agricultural drought monitoring: A state of art review. *AIMS Environmental Science*, 3(4), 604–630. https://doi.org/10.3934/environsci.2016.4.604

[30] Benso, M. R., et al. (2023). Multi-hazard risk insurance design for food security resilience: A systematic literature review. *Natural Hazards and Earth System Sciences*, 23, 1335–1362. https://doi.org/10.5194/nhess-23-1335-2023

[31] Skees, J. R., Barnett, B. J., et al. (2006). Enhancing microfinance using index-based risk-transfer products. *Agricultural Finance Review*, 66(2), 235. https://doi.org/10.1108/00214660680001189

[32] Ntukamazina, N., Onwonga, R. N., Sommer, R., Rubyogo, J.-C., Mukankusi, C. M., Mburu, J., & Kariuki, R. (2017). Index-based agricultural insurance products: challenges, opportunities and prospects for uptake in sub-Sahara Africa. *Journal of Agriculture and Rural Development in the Tropics and Subtropics (JARTS)*, 118(2), 171–185. https://www.jarts.info/index.php/jarts/article/view/2696

[33] Jensen, N. D., Barrett, C. B., & Mude, A. G. (2016). Index insurance quality and basis risk: Evidence from northern Kenya. *American Journal of Agricultural Economics*, 98(5), 1450–1469. https://doi.org/10.1093/ajae/aaw046

[34] Gaurav, S., & Chaudhary, V. (2019). Do farmers care about basis risk? Evidence from a field experiment in India. *Climate Risk Management*, 25, 100201. https://doi.org/10.1016/j.crm.2019.100201

[35] Hill, R. V., & Viceisza, A. C. G. (2012). A field experiment on the impact of weather shocks and insurance on risky investment. *Experimental Economics*, 15(2), 341–371. https://doi.org/10.1007/s10683-011-9283-8

[36] Cole, S., Stein, D., & Tobacman, J. (2014). Dynamics of demand for index insurance: Evidence from a long-run field experiment. *American Economic Review*, 104(5), 284–290. https://doi.org/10.1257/aer.104.5.284

[37] Mobarak, A. M., & Rosenzweig, M. R. (2013). Informal risk sharing, index insurance, and risk taking in developing countries. *American Economic Review*, 103(3), 375–380. https://doi.org/10.1257/aer.103.3.375

[38] Achtnicht, M., & Osberghaus, D. (2019). The demand for index-based flood insurance in a high-income country. *German Economic Review*, 20(2), 217–242. https://doi.org/10.1111/geer.12142

[39] Boyd, M., Porth, B., Porth, L., & Turenne, D. (2019). The impact of spatial interpolation techniques on spatial basis risk for weather insurance: An application to forage crops. *North American Actuarial Journal*, 23(4), 532–556. https://doi.org/10.1080/10920277.2019.1566074

[40] Dalhaus, T., & Finger, R. (2016). Can gridded precipitation data and phenological observations reduce basis risk of weather index-based insurance? *Weather, Climate, and Society*, 8(4), 409–419. https://doi.org/10.1175/WCAS-D-16-0020.1

[41] Eze, E., Girma, A., Zenebe, A. A., & Zenebe, G. (2020). Feasible crop insurance indexes for drought risk management in Northern Ethiopia. *International Journal of Disaster Risk Reduction*, 47, 101544. https://doi.org/10.1016/j.ijdrr.2020.101544

[42] Guerrero-Baena, M. D., & Gómez-Limón, J. A. (2019). Insuring water supply in irrigated agriculture: A proposal for hydrological drought index-based insurance in Spain. *Water*, 11(4), 686. https://doi.org/10.3390/w11040686

[43] Guzmán, D. A., Mohor, G. S., & Mendiondo, E. M. (2020). Multi-year index-based insurance for adapting water utility companies to hydrological drought: Case study of a water supply system of the São Paulo Metropolitan Region, Brazil. *Water*, 12(11), 2954. https://doi.org/10.3390/w12112954

[44] Leblois, A., Quirion, P., & Sultan, B. (2014). Price vs. weather shock hedging for cash crops: Ex ante evaluation for cotton producers in Cameroon. *Ecological Economics*, 101, 67–80. https://doi.org/10.1016/j.ecolecon.2014.02.021

[45] Shahhosseini, M., Hu, G., Huber, I., & Archontoulis, S. V. (2021). Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. *Scientific Reports*, 11(1), 1606. https://doi.org/10.1038/s41598-020-80820-1

[46] Ejiyi, C. J., Qin, Z., Salako, A. A., Happy, M. N., Nneji, G. U., Ukwuoma, C. C., Chikwendu, I. A., & Gen, J. (2022). Comparative analysis of building insurance prediction using some machine learning algorithms. *International Journal of Interactive Multimedia and Artificial Intelligence*, 7(3), 75–85. https://doi.org/10.9781/ijimai.2022.02.005

[47] Hott, C., & Regner, J. (2023). Weather extremes, agriculture and the value of weather index insurance. *The Geneva Risk and Insurance Review*, 48(2), 230–259. https://doi.org/10.1057/s10713-023-00081-6

[48] Oussaoui, S., Boudhar, A., Hadri, A., et al. (2025). Mapping drought severity impact on arboriculture systems over Tadla and Lower Tassaout plains in Morocco using Sentinel-2 data and machine learning approaches. *Geocarto International*, 40(1), 2471104. https://doi.org/10.1080/10106049.2025.2471104

[49] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[50] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[51] Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

[52] Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer. https://doi.org/10.1002/9781118445112.stat06472
[53] Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1
[54] Wijesena, S., & Pradhan, B. (2025). Enhancing weather index insurance through surrogate models: Leveraging machine learning techniques and remote sensing data. *Environmental Research Communications*, 7(4), 045009. https://doi.org/10.1088/2515-7620/adba2c
[55] Tadesse, M. A., Shiferaw, B. A., & Erenstein, O. (2015). Weather index insurance for managing drought risk in smallholder agriculture: Lessons and policy implications for sub-Saharan Africa. *Agricultural and Food Economics*, 3(1), 26. https://doi.org/10.1186/s40100-015-0044-3
[56] Darkoh, G. O., Laweh, N., Dasori, E. A., Abilla, M. L., Orji, N., & Amfo, S. J. (2025). Determinants of weather index-based insurance adoption among Ghanaian farmers. *EPRA International Journal of Agriculture and Rural Economic Research*, 13(1), 1–15. https://doi.org/10.36713/epra19711
[57] Belmahi, M., Hanchane, M., Krakauer, N. Y., Kessabi, R., Bouayad, H., Mahjoub, A., & Zouhri, D. (2023). Analysis of relationship between grain yield and NDVI from MODIS in the Fez-Meknes region, Morocco. *Remote Sensing*, 15(11), 2707. https://doi.org/10.3390/rs15112707
[58] Chen, C.-C., McCarl, B. A., & Schimmelpfennig, D. E. (2004). Yield variability as influenced by climate: A statistical investigation. *Climatic Change*, 66(1), 239–261. https://doi.org/10.1023/B:CLIM.0000043159.33816.e5
[59] Deschênes, O., & Greenstone, M. (2007). The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather. *American Economic Review*, 97(1), 354–385. https://doi.org/10.1257/aer.97.1.354
[60] Bakker, M. M., Govers, G., Ewert, F., Rounsevell, M., & Jones, R. (2005). Variability in regional wheat yields as a function of climate, soil and economic variables: Assessing the risk of confounding. *Agriculture, Ecosystems & Environment*, 110(3–4), 195–209. https://doi.org/10.1016/j.agee.2005.04.016
[61] Brown, I. (2013). Influence of seasonal weather and climate variability on crop yields in Scotland. *International Journal of Biometeorology*, 57(4), 605–614. https://doi.org/10.1007/s00484-012-0588-9
[62] Eitzinger, J., Thaler, S., Schmid, E., Strauss, F., Ferrise, R., Moriondo, M., Bindi, M., Palosuo, T., Rötter, R., Kersebaum, K. C., et al. (2013). Sensitivities of crop models to extreme weather conditions during flowering period demonstrated for maize and winter wheat in Austria. *The Journal of Agricultural Science*, 151(6), 813–835. https://doi.org/10.1017/S0021859612000779
[63] Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Klein Tank, A., Peterson, T. C., Trewin, B., & Zwiers, F. W. (2011). Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdisciplinary Reviews: Climate Change*, 2(6), 851–870. https://doi.org/10.1002/wcc.147
[64] Torriani, D. S., Calanca, P., Schmid, S., Beniston, M., & Fuhrer, J. (2007). Potential effects of changes in mean climate and climate variability on the yield of winter and spring crops in Switzerland. *Climate Research*, 34, 59–69. https://doi.org/10.3354/cr034059
[65] Gollier, C. (2003). To insure or not to insure? An insurance puzzle. *The Geneva Papers on Risk and Insurance Theory*, 28(1), 5–24. https://doi.org/10.1023/A:1022112430242
[66] Bucheli, J., Dalhaus, T., & Finger, R. (2021). The optimal drought index for designing weather index insurance. *European Review of Agricultural Economics*, 48(3), 573–597. https://doi.org/10.1093/erae/jbaa014
[67] Kapphan, I., Calanca, P., & Holzkaemper, A. (2012). Climate change, weather insurance design and hedging effectiveness. *The Geneva Papers on Risk and Insurance – Issues and Practice*, 37(2), 286–317. https://doi.org/10.1057/gpp.2012.8
[68] Ricome, A., Affholder, F., Gérard, F., Muller, B., Poeydebat, C., Quirion, P., & Sall, M. (2017). Are subsidies to weather-index insurance the best use of public funds? A bio-economic farm model applied to the Senegalese groundnut basin. *Agricultural Systems*, 156, 149–176. https://doi.org/10.1016/j.agsy.2017.05.015
[69] Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). Modelling Extremal Events: For Insurance and Finance. Springer. https://doi.org/10.1007/978-3-642-33483-2
[70] Frees, E. W., Derrig, R. A., & Meyers, G. (2014). Predictive Modeling Applications in Actuarial Science (Vol. 1). Cambridge University Press. https://doi.org/10.1017/CBO9781139342674

(Executive Editor: Qun Niu)