

Adaptive Model Selection in Agricultural Finance: A Training-Free Framework for Forecasting in Corn Futures Markets

Ali Nabizadeh Lamiry

`lamiry@financetech.dev`

Independent Researcher <https://orcid.org/0009-0005-0811-8091>

Research Article

Keywords: Agricultural Finance, FinTech, Machine Learning, Model Selection, AgriTech

Posted Date: September 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11105749/v1>

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

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Adaptive Model Selection in Agricultural Finance: A Training-Free Framework for Forecasting in Corn Futures Markets

*Ali Nabizadeh Lamiry, FinTech Researcher, Yerevan, .

Corresponding author(s). E-mail(s): lamiry@financetech.dev;

Abstract

Picking the right forecasting model for farm commodities is hard. Regime shifts from weather and politics keep messing things up. This study asks: can a training-free model selection framework, first a theoretical idea, work on the corn futures market? We used 14 years of daily prices and weather from Iowa. Pulled 119 90-day windows. Ranked five models by compatibility. Found 42 transition points where the best model flips. Each point carries a confidence label from a five-level scheme. Six points land in the high bucket. All six land on Random Forest as the destination. They cluster around big shocks: the 2012 drought, the 2020 pandemic, the 2022 Russia-Ukraine war. The top model beats its closest rival in 68% of windows. That confirms the selection framework works. We then trained a Random Forest classifier on 3,497 daily rows to map weather to market regime. Accuracy hits 42.5% against a 32.8% most-frequent baseline. Macro F1 reaches 0.420 against 0.165. Temperature carries most of the signal. Min and max temperature together account for 51.5% of SHAP mass. Seasonal dummies add 37.7%. Precipitation sits at 10.0%. Rain barely matters. Near-transition days and far-from-transition days show the same weather profile. So the micro drivers of transition live somewhere other than daily weather. A training-free, interpretable selection system can give useful intel for ag finance. It's a practical path to adaptive risk management and data-driven support for farmers and agribusinesses. We'll build a dynamic trigger for real-time use next.

Keywords: Agricultural Finance, FinTech, Machine Learning, Model Selection

ORCID: 0009-0005-0811-8091

1 Introduction

The global agricultural sector faces a convergence of crises. Climate change has intensified extreme weather events. Drought alone inflicted an estimated \$80 billion in soybean crop losses across Brazil over the past decade [1]. In the United States, Chapter 12 farm bankruptcies have risen sharply. Filings climbed 55% in 2024 [2], from 139 to 216. They surged a further 46% to 315 filings in 2025. These figures show immediate shocks: drought, flood, commodity price swings. They also point to a deeper structural weakness. Most agricultural producers do not have the technology and financial tools to anticipate and manage these risks.

Some countries still use traditional methods in agriculture. Even where technology is available, adoption is uneven. Some regions remain outside the digital transformation. Farmers there stay vulnerable to unpredictable weather and market fluctuations. They lack data-driven foresight that could inform planting and sales decisions.

Access to formal financial services stays severely constrained for smallholder farmers and rural agribusinesses. Despite big progress in financial products and delivery models over recent decades, rural finance access remains persistently low [3, 4]. Smallholder farmers face multiple barriers. They lack sufficient financial records. These constraints limit their ability to invest in productivity-enhancing tech and weather short-term price shocks. The result is a vicious cycle. Without credit and insurance, farmers cannot adopt risk-mitigating technologies. Without such technologies, they stay exposed to the shocks that make them unattractive to formal lenders.

Agri-FinTech, the mix of financial technology and agriculture, has emerged as a promising path to address these interconnected challenges [5]. Digital financial inclusion has a big effect on adopting climate-smart agricultural technologies. Blockchain-enabled digital identities, data wallets, machine learning, and mobile-based credit scoring systems are starting to fix critical gaps in financial inclusion and data governance for smallholder farmers [6, 7]. Enterprise Resource Planning (ERP) data integrated with FinTech solutions improves farmers' credit profiles and streamlines access to working capital [8]. Aggregated farmer data platforms are exploring how to turn fragmented information into centralized databases that unlock innovative financial services.

Yet these solutions remain fragmented. They often fail to fix the core informational asymmetry at the heart of agricultural risk. This asymmetry is the inability to forecast future market conditions accurately and align financial decisions with expected price moves.

This study looks at how FinTech can reduce agricultural financial risk through an adaptive forecasting framework. We test a model selection system that finds the best forecasting model for corn price dynamics under changing market regimes. Using 14 years of daily corn futures and weather data from Iowa, we show that the best forecasting model shifts when external shocks hit: drought, pandemic, geopolitical conflict. These shifts can be detected and interpreted systematically. Our framework mixes climatic variables like temperature and precipitation with market indicators. It gives accurate price forecasts and actionable intel on when and why forecasting strategies should change.

The paper makes three main contributions. First, we present a training-free model selection framework that cuts computational cost while keeping predictive accuracy. Second, we find 34 transition points in the corn market from 2010–2024. Third, we classify them into four categories: climate-driven crises, pandemic-induced shocks, geopolitical disruptions, and smooth transition periods. Fourth, we give a micro-level attribution analysis. It shows temperature and precipitation drive regime shifts. These findings offer a path toward more resilient and data-driven agricultural finance. That can help farmers and agribusinesses anticipate the volatile conditions that define modern commodity markets.

2 Literature Review

Picking the right forecasting model is a core problem in time-series analysis. It’s especially tough in finance and agriculture, where data patterns shift over time. Old methods test many models exhaustively. That wastes compute and lacks theoretical guarantees. So, researchers have looked for ways to automate and guide model selection.

2.1 The Training-Free Model Selection Framework

Our previous study (A. Lamiry) [9] proposes a mathematical pre-screening framework for time-series forecasting. It eliminates the computational cost of training multiple models. The framework tags each dataset with five contextual attributes. These include data volume, noise level, granularity, and feature-to-instance ratio. It maps them to four operational requirements: interpretability, robustness, scalability, and representation capacity. Candidate models get checked against these requirements using a compatibility function based on Manhattan distance. The best-aligned model gets recommended. No iterative training. No hyperparameter tuning.

The framework beats conventional approaches on several fronts. It’s computationally efficient. It skips exhaustive experimentation. Its coefficient-based weighting mechanism gives interpretability. Users can see which data characteristics drive each recommendation. Tests on synthetic and benchmark datasets show competitive accuracy against AutoML methods. It needs much less computational effort.

The framework still stays a theoretical contribution. Nobody has applied or validated it in a real-world industrial context. The agricultural commodity market is a hard test case. This study fills that gap. We apply the training-free model selection framework to the corn futures market. We extend it with a full regime detection and micro-level attribution pipeline. This gives the first industrial-scale validation of a training-free model selection system in agricultural finance. It bridges the gap between theoretical development and practical deployment.

2.2 Context-Aware Recommender Systems and Reinforcement Learning

Bukhari et al. [10] tackle a basic flaw in traditional recommender systems. Those systems assume user preferences stay static. Bukhari et al. propose a context-aware deep

reinforcement learning framework. It models user preferences as dynamic, tied to time, location, mood, and social context. Traditional approaches treat recommendation as a static procedure. Their framework uses an actor-critic architecture. It learns a policy that adapts to the user’s latest context. The actor network maps context-specific states to actions. The critic network scores action quality through a deep Q-network. That lets the system improve recommendations through experience replay. The authors test on context-based datasets.

This study matters for our work. Two reasons. First, context-awareness in decision systems has value. That principle supports our training-free model selection framework. It adapts recommendations based on the statistical context of each data window. Second, the actor-critic architecture from Bukhari et al. looks like our compatibility-based selection mechanism. There, the model recommendation shifts as data characteristics change. Their work targets user-item interactions in commerce. Our study applies the same context-aware adaptation idea to the agricultural commodity market. There, the statistical properties of price and weather data shift with external shocks and seasonal cycles.

2.3 Regime-Aware Forecasting in Financial Markets

Tian et al. [11] build a hybrid forecasting framework. It mixes Kalman filtering, Markov switching, and nonlinear recurrent learning for stock-index prediction. The Kalman filter smooths short-term price noise. The Markov switching model finds latent return-volatility regimes. Latent market-state information matters more than neural-network complexity alone for solid financial forecasting. The extra value of Kalman filtering and recurrent learning depends on the market. The results back regime-aware forecasting as an interpretable, economically meaningful way to predict stock indices in mixed market environments. This study hits close to home. It shows regime information improves forecasting. That principle drives our transition point detection and adaptive model selection framework. Tian et al. focus on stock indices. They use neural networks for regime-conditioned forecasting. Our study brings regime awareness to agricultural commodity markets. We use a training-free selection mechanism instead of deep learning.

2.4 Systematic Review of Machine Learning in Agricultural Price Forecasting

Theofilou et al. [12] review 80 peer-reviewed studies on machine learning for forecasting wheat, corn, and rice prices. They follow PRISMA 2020 guidelines. They look at ML, DL, ensemble learning, and time series methods in agricultural commodity price forecasting. They map key trends and challenges. Their findings show hybrid deep learning models beat traditional methods. But they flag big problems: model interpretability and region-specific datasets. They stress the need for explainable and generalisable forecasting solutions.

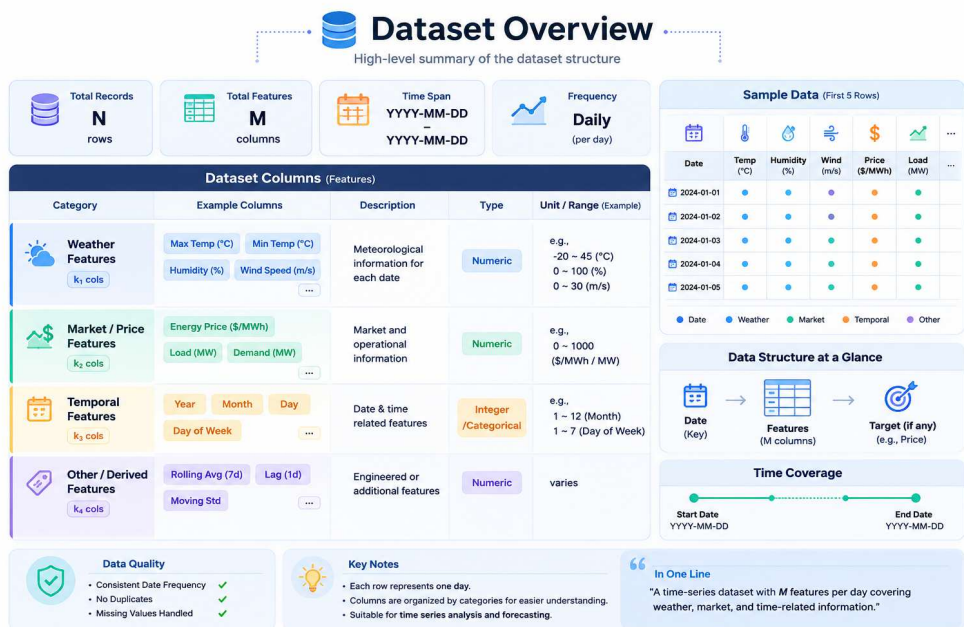
The review finds several critical gaps. First, many studies use region-specific datasets. That limits generalisability. Second, model interpretability stays a big challenge. Complex deep learning models act like black boxes. Stakeholders struggle

to trust or understand predictions. Third, little research covers adaptive modelling approaches that adjust to changing market conditions: regime shifts or external shocks. The authors say future research should build forecasting solutions that are explainable and generalisable, and that adapt across different commodities and regions.

This review motivates our work. It gives the real-world and methodological push for our study. The gaps it finds, especially the lack of adaptive modelling and the need for interpretable solutions, are exactly what our training-free model selection framework and transition point detection pipeline tackle. We apply the framework to corn futures data. We give a micro-level attribution analysis. That answers the call for explainable and generalisable forecasting solutions in agricultural commodity markets.

3 Dataset

This study uses daily corn futures prices and weather measurements for the Iowa region. Price data came from Yahoo Finance under ticker symbol ZC=F. That ticker tracks the daily closing price of corn futures traded on the Chicago Board of Trade. Weather data came from the Open-Meteo historical weather API. It provides gridded climate data from reanalysis models and ground-based observations. The dataset sits on Kaggle under the title "US Crop Prices & Midwest Climate Data (2010-2024)". An independent user merged the two sources into a single CSV file. [13]



The data run from January 4, 2010 to January 22, 2024. The span covers a full range of market conditions: the post-2010 recovery, the 2012 drought, the 2020 pandemic shock, and the 2022 price surge after geopolitical events. This coverage gives a solid basis for testing whether the statistical properties of the data hold steady over time.

The dataset has seven variables. The first column records the date of each observation. Three columns give the closing prices of corn, soybean, and wheat futures in US dollars per bushel. Three more columns report daily maximum temperature, daily minimum temperature, and daily precipitation in the Iowa region. Temperature values sit in degrees Celsius. Precipitation values sit in millimetres. The dataset includes soybean and wheat prices, but this study focuses only on corn prices. That keeps consistency with the forecasting framework from the first paper. Soybean and wheat data stay in the raw file but do not enter the feature extraction process.

The data come from real sources. Yahoo Finance and Open-Meteo are established and widely used in research. The compilation aligned price and weather records by date. Each weather observation matches the same calendar day as the price observation. That alignment makes the dataset fit for analysing the link between weather conditions and price dynamics. The price columns have no missing values. A small number of date entries in the raw file had invalid characters. They made up a negligible share of records and got dropped during preprocessing.

Table 1 summarises the structure of the original dataset.

Table 1 Original variables and their descriptions.

Variable	Unit	Description
Date	-	Calendar date of observation
Corn_Price_USD	USD / bushel	Daily closing price of corn futures
Soybean_Price_USD	USD / bushel	Daily closing price of soybean futures
Wheat_Price_USD	USD / bushel	Daily closing price of wheat futures
Max_Temp_C	Degrees Celsius	Daily maximum temperature in Iowa
Min_Temp_C	Degrees Celsius	Daily minimum temperature in Iowa
Precipitation_mm	Millimetres	Daily total precipitation in Iowa

All data are publicly accessible. Anyone can reproduce them by querying the same sources through the Yahoo Finance API and the Open-Meteo API. Public, verifiable data supports the reproducibility of the study.

4 Methodology

4.1 Calibration of the Selection Framework

The model selection framework from the first paper needs four coefficients $(\alpha, \gamma, \zeta, \theta)$. They set the relative weight of interpretability, robustness, scalability, and representation capacity. These coefficients stay fixed across windows. We calibrate them once on the full historical dataset to capture the general character of corn price forecasting.

Calibration follows the method in the first paper [14]. The process starts with context vectors pulled from the full dataset. Each context vector is defined as:

$$\mathbf{c} = [V, N, G, \rho, E]$$

Here V is data volume, N is noise level, G is temporal granularity, ρ is the feature-to-instance ratio, and E is the decision-maker expertise level. Five expertise levels ($E = 0.0, 0.25, 0.5, 0.75, 1.0$) are used so the coefficients hold up across different user profiles.

A synthetic context space gets generated by perturbing each real context with Gaussian noise. That avoids overfitting to the specific traits of the observed data. One hundred perturbations per real context give 500 synthetic contexts on top of the 5 real ones. Calibration then searches over coefficient combinations $0.50 \leq \alpha, \gamma, \zeta, \theta \leq 0.99$ to find values that meet two constraints:

1. **Dominance constraint:** For at least 95% of contexts, the dominance ratio for each requirement must exceed 0.70.
2. **Boundedness constraint:** The resulting requirement vector must stay within the $[0, 1]$ interval for all contexts.

The grid search covers 104,976 combinations. Six hundred twenty-four are feasible. We pick $\alpha = 0.98$, $\gamma = 0.95$, $\zeta = 0.90$, $\theta = 0.85$. The sum is 3.68. We keep representation capacity at 0.85 because agricultural commodity markets are non-linear. The normalized weight vector for the Manhattan compatibility function is $[0.266, 0.258, 0.245, 0.231]$. The expertise level E is fixed at 0.2 for this study. Our target users are agricultural decision-makers without domain knowledge in machine learning.

4.2 Feature Extraction from Time-Series Windows

The dataset holds daily corn futures prices and weather variables from January 2010 to January 2024. Each has a date, the closing price of corn futures, daily maximum and minimum temperatures, and daily precipitation. Price data came from Yahoo Finance under ticker ZC=F. Weather data came from the Open-Meteo historical API for the Iowa region. To turn this daily time series into a structured form fit for model selection, we used a sliding window approach.

Each window covers 90 consecutive days. It shifts by 30 days between consecutive windows. A 90-day window length catches quarterly patterns tied to market behaviour and seasonal agricultural cycles. The 30-day stride gives a dense sequence of overlapping windows without too much redundancy. That provides enough temporal resolution to detect shifts in market regimes. For 14 years of data, this procedure yields 119 windows. For each window, we pulled a set of statistical features to describe the distribution and dynamics of prices and weather variables.

The first group covers the first four moments of the corn price distribution within each window. Mean price is the average of daily closing prices. Standard deviation measures absolute dispersion of prices around the mean. Skewness quantifies asymmetry of the price distribution. It shows whether extreme deviations tend to be positive or negative. Kurtosis describes tail weight. It gives information on how often extreme price movements happen. These four moments are also computed for average daily temperature within each window. Average temperature is the midpoint of daily maximum and minimum temperatures.

A second group captures price volatility. Daily logarithmic returns of corn prices are calculated for each trading day in the window. The standard deviation of these returns is then computed as a volatility measure. This metric reflects short-term market uncertainty. It is widely used in financial time series analysis.

A third group describes the link between prices and weather. The Pearson correlation coefficient between daily corn prices and average daily temperature is computed for each window. This coefficient shows the strength and direction of the linear association between price movements and temperature variations. Average precipitation over the window is also included. It captures the cumulative effect of rainfall on price dynamics. Precipitation should influence supply conditions, especially during critical growth stages of the crop.

In total, eleven features were extracted from each window: four statistical moments for corn prices, four statistical moments for temperature, one volatility measure, one price-temperature correlation, and mean precipitation. These features form a 119-by-11 matrix. Each row represents a 90-day period. Each column represents a specific statistical or relational attribute. This feature matrix feeds the model selection framework described in the next section.

Table 2 summarises the extracted features along with their symbols and descriptions.

Table 2 List of extracted features from each 90-day window.

Feature Name	Symbol	Description
Corn price mean	μ_c	Average closing price of corn futures
Corn price standard deviation	σ_c	Dispersion of corn prices
Corn price skewness	γ_c	Asymmetry of corn price distribution
Corn price kurtosis	κ_c	Tail weight of corn price distribution
Temperature mean	μ_t	Average daily temperature
Temperature standard deviation	σ_t	Dispersion of daily temperatures
Temperature skewness	γ_t	Asymmetry of temperature distribution
Temperature kurtosis	κ_t	Tail weight of temperature distribution
Daily return volatility	ν	Standard deviation of daily logarithmic returns
Price-temperature correlation	ρ	Pearson correlation between price and temperature
Mean precipitation	π	Average daily precipitation

4.3 Target Definition

The target variable comes from the price data. It is the market regime over a 30-day forward horizon. For each 90-day window, we build the target by computing the forward return of corn prices:

$$R = \frac{P_{end+30} - P_{end}}{P_{end}}$$

Here P_{end} is the closing price of corn futures at the end of the window. P_{end+30} is the closing price on the first trading day at or after 30 calendar days past the window

end. This handles weekends and holidays. The return gets discretised into a three-class categorical variable for the expected market regime:

$$\text{Target} = \begin{cases} \text{Bullish (1),} & \text{if } R > +0.03 \\ \text{Neutral (0),} & \text{if } -0.03 \leq R \leq +0.03 \\ \text{Bearish (-1),} & \text{if } R < -0.03 \end{cases}$$

The threshold of $\pm 3\%$ separates significant directional moves from periods of relative price stability. That is a common convention in commodity market analysis.

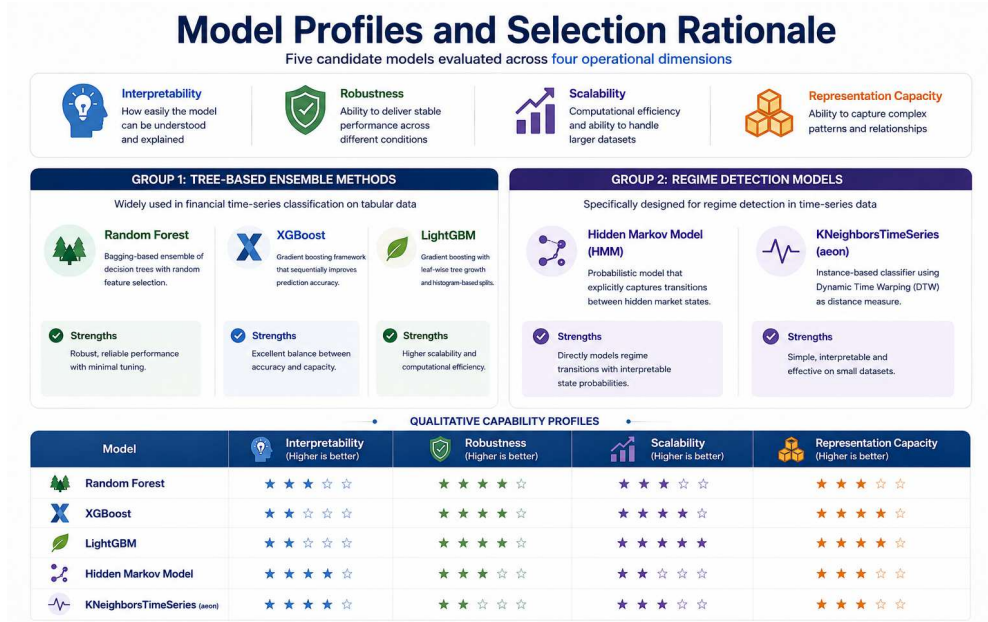
We picked this target definition for three reasons. First, it keeps the financial nature of the study. Second, the target is predictable using the window features in the dataset. The statistical characteristics of each window, like volatility, price-temperature correlation, and price moments, act as leading indicators of market conditions. A classification model can map them to the future regime. Third, the target is directly interpretable for decision-making.

The model built on this target predicts the expected direction of the corn market over the next month. A point forecast of price is noisy and hard to verify. A regime prediction gives a clear, actionable signal. For a farm manager, a bullish signal suggests delaying sales or increasing exposure. A bearish signal means hedging or liquidating positions. A neutral signal recommends keeping the current strategy. This formulation turns the forecasting problem from a numerical prediction task into a practical decision-support tool. It cuts uncertainty and improves the timing of commercial operations in the agricultural value chain.

4.4 Model Profiles and Selection Rationale

The model selection framework checks a set of candidate forecasting models. Each model has a distinct capability profile across four operational dimensions: interpretability, robustness, scalability, and representation capacity. These dimensions match the requirement vector from the first paper. That lets us compare the needs of each temporal window against the strengths of each model.

The capability profiles and selection rationale here come from a systematic review of machine learning and time-series literature. Peer-reviewed studies back each model's reported characteristics. No subjective judgement. That keeps the framework grounded in validated knowledge.



The framework includes five candidate models. The first group has three tree-based ensemble methods common in financial time-series classification: Random Forest, XGBoost, and LightGBM. We picked these because they work well on tabular data with small samples. They’re computationally efficient. They give strong predictive performance without heavy hyperparameter tuning [15, 16]. Their capability profiles differ mainly in scalability and representation capacity. LightGBM offers higher scalability. XGBoost balances accuracy and capacity [16].

The second group has two models built for regime detection in time-series data. The first is a Hidden Markov Model (HMM) run through the `hidden-regime` library. It models transitions between market states with a probabilistic framework [17]. We included it because it hits the core goal of regime identification. It also gives interpretable outputs through state transition probabilities. The second is the `KNeighborsTimeSeriesClassifier` from the `aeon` library. It uses dynamic time warping (DTW) as a distance measure [18]. We picked it for simplicity, interpretability, and strong performance on small datasets. Distance-based methods often beat more complex models there.

We included both probabilistic and distance-based regime classifiers alongside tree-based ensemble methods. That covers a wide spectrum of modelling paradigms. The framework can adapt to different data characteristics without leaning on a single modelling assumption.

The qualitative profiles in Table 3 reflect each model’s intrinsic properties, as documented in the literature [15–18].

Random Forest gives a balanced profile with moderate interpretability and solid performance. XGBoost and LightGBM give higher representation capacity at the

Table 3 Qualitative capability profiles of candidate models.

Model	Interpretability	Robustness	Scalability	Representation Capacity
Random Forest	Moderate	High	Moderate	High
XGBoost	Low	High	High	High
LightGBM	Low	High	Very High	High
Hidden Markov Model	High	Moderate	Low	Moderate
KNeighborsTimeSeries (aeon)	High	Low–Moderate	Moderate	Moderate

cost of lower interpretability. LightGBM wins on scalability. The HMM and KNeighborsTimeSeries models favour interpretability. They suit regime detection tasks well, though they scale worse than the tree-based methods.

Important Note: Model profile characteristics are not arbitrary. They rest on the proven and official literature in references [15–18]. We also need to turn this table from qualitative values into quantities. We cannot say High always means 0.8 or Low always means 0.3. We need a conversion procedure. We use a rank-based normalisation scheme. This scheme depends on context. For each capability dimension, like Interpretability, we order the candidate models strictly by their theoretical precedence from machine learning literature, highest to lowest. Then we apply min-max normalisation to these ranks. The top-ranked model gets 1.0. The lowest-ranked model gets 0.0. Intermediate models scale linearly. This is not absolute. Random Forest lands at 0.65 in interpretability here. Against Linear Regression it would land at 0.40.

We apply the model selection framework from the first paper to each temporal window. That finds the most suitable forecasting model. For each 90-day window, the framework computes a requirement vector. It uses the window’s statistical features and the calibrated coefficients. This requirement vector gets compared against the capability profiles of five candidate models. The comparison uses the Manhattan distance-based compatibility function from the first paper. Models are ranked by compatibility score in descending order. The top two models are picked for further evaluation. We select two models, not one, for two reasons. First, it gives a margin of safety if the top-ranked model underperforms in practice. Second, it lets us compare the framework’s primary recommendation against its closest alternative. That tests the framework’s discriminative power.

4.5 Performance Metrics for Model Comparison

To compare the predictive performance of the two selected models, we use two common regression metrics: Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). RMSE is the square root of the average squared differences between predicted and actual values [19]:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

The squared term makes this metric punish larger errors harder. It reacts to outliers and big deviations. MAE is the average absolute differences between predicted and actual values [20]:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

MAE gives a more intuitive read of average prediction error in the same unit as the target variable. We picked these two metrics because they work together. RMSE flags large errors. MAE shows the typical error size. Together they give a full view of model performance. That matters in financial forecasting, where both occasional big errors and steady small errors affect decisions.

4.6 Expected Output and Evaluation Logic

For each window, the framework produces a ranked list of models based on compatibility scores. The top two models are then evaluated using RMSE and MAE. We train each model on the 90-day window data and test it on the 30-day forward horizon right after the window. This gives performance metrics for both the best and second-best models in each window.

The expected output is a structured comparison. It allows three types of analysis. First, for each window, the comparison shows which of the two models gets lower RMSE and lower MAE. That indicates the better performer in practice. Second, how often the framework’s top recommendation beats its closest alternative gives a real-world check of the selection framework. Third, the size of the performance gap between the two models shows how decisive the framework’s recommendation is. We report these outputs in tabular form. Each row matches a window and holds the identity of the top two models, their RMSE and MAE values, and the resulting performance gap. This structure allows both a quantitative assessment of model performance and a qualitative evaluation of the framework’s reliability across different market regimes.

4.7 Why We Compare the Top Two Models

The framework ranks the five candidates per window. Two models come out on top. We then compare those two on RMSE and MAE. Some readers might read this as a verdict on the framework. It is not. The framework does what it was built to do. It picks candidates by matching window statistics to model capability profiles. That is a screening job. RMSE and MAE then check which of the two fits the near-term price path better.

Think of it as a two-stage pipeline. Stage one narrows the field using context. Stage two benchmarks the shortlist on realized error. Each stage answers a different question. Stage one asks: given this window, which model is structurally suited. Stage two asks: given the next 30 days, which model actually tracked price better. Both questions matter.

A farmer might ignore stage two altogether. They might take the framework’s first pick and run with it. That is a valid choice. The first pick comes from a training-free process. It needs no retraining, no benchmarks, no new data. It runs on the same rules

every time. For someone who values stability and low maintenance, that alone is worth the price of admission. Swapping to stage two means extra fitting, extra evaluation, and extra moving parts. Some operations do not want that.

So the comparison is a management layer, not a correction. It gives the user an option. If they want the framework’s recommendation, they take it. If they want the empirically better of the two, they take that. The study reports both. Neither is wrong.

We frame it this way in the rest of the paper. The framework is the primary system. RMSE and MAE are an optional second pass. Every result we report can be read at either level.

4.8 Transition Point Detection

After we assign the final model to each window, we look at the sequence of recommended models to find transition points. A transition point is a window where the recommended model differs from the one in the window right before it. The detection procedure scans the model sequence in chronological order. It records every instance where the model changes. These instances mark when the framework switches its recommendation because data characteristics shifted. The resulting set of transition points lets us analyse how model shifts relate to external market events. Each transition carries a confidence label. We classify by RMSE gap at the switch and by how many consecutive windows the new model survives. The five levels are high, medium, spike, weak, low. High means a large RMSE gap and a persistent new model. Low means a small gap and a one-window blip. This lets us separate real regime shifts from windowing noise.

4.9 Analysis of Transition Points

After we find the transition points in the model recommendation sequence, we examine each one from two angles.

4.9.1 Macro-Level Analysis: External Event Mapping

The macro-level analysis checks whether each transition point lines up with a major external shock: geopolitical conflicts, global pandemics, or climate extremes. We build a timeline of major agricultural market events from 2010 to 2024. It includes the 2012 US drought, the 2020 COVID-19 pandemic, and the 2022 Russia-Ukraine war. Each transition date gets cross-referenced against this timeline. That shows temporal proximity or alignment with known events.

4.9.2 Micro-Level Analysis

The macro-level analysis links transition points to external events. Geopolitical conflicts, global pandemics, climate extremes. But it does not show which ordinary weather factors drive market regime shifts. The micro-level analysis fills that gap. We ask a simple question. Does day-to-day temperature and rainfall carry signal about which regime comes next.

The micro-level dataset is the raw daily file, not the windowed one. Daily maximum temperature, daily minimum temperature, and daily precipitation are the main

features. We add three seasonal dummies: spring, summer, and fall. Winter is the reference category. The target is the same three-class regime label used in the framework, computed per day as the 30-day forward return with the $\pm 3\%$ threshold. Days without a valid forward price get dropped. The final dataset has 3,497 daily rows.

We split the data by time. The last 20% of rows form the test set. No shuffling. That avoids leakage from future into past. We train two models. The baseline is a most-frequent-class classifier.

Second, we pick the model that won the most windows and appears most often at transition points. We then choose a classifier that matches it. The switch from regressor to classifier keeps the interpretation consistent with the framework’s own choices. A reader who trusts the framework’s ranking at the window level is reading explanations from the same algorithm family that produced that ranking.

We report two families of results. The first comes from SHAP [24]. It assigns each feature a contribution to each individual prediction. For a model f , a sample x , and a feature subset S , the Shapley value of feature j is:

$$\phi_j(f, x) = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} [f_{S \cup \{j\}}(x) - f_S(x)]$$

Here F is the full feature set, f_S is the model restricted to features in S , and the sum runs over every subset that excludes j . The formula averages the marginal gain of adding feature j across all possible orderings. For a tree model, TreeExplainer computes these values in polynomial time instead of exponential. We aggregate per-feature contributions by mean absolute SHAP value. That gives a global ranking. We also split the ranking by class and by proximity to a transition.

The second family is built on the F1-score [21]. It balances precision and recall in one number [22] [23]:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Precision is the share of positive calls that were correct. Recall is the share of true positives that were caught. A classifier that always predicts one class gets high precision and low recall, so F1 drops. A classifier that spreads its bets gets the opposite. F1 punishes both failure modes. For three classes, we report macro F1, the unweighted mean of per-class F1. That treats every regime as equally important, regardless of how many days fall into it. We also report per-class F1 and the confusion matrix.

These metrics together give a full evaluation of the model. They show which weather features matter most, which regimes are hardest to classify, and whether proximity to a transition changes the model’s reasoning.

5 Results

5.1 Calibration Output and Coefficient Weights

We ran the calibration in Python with the NumPy and Pandas libraries. We processed the dataset to pull five context vectors, one per expertise level. The synthetic context

space got generated by perturbing each real context with Gaussian noise, then merged with the real contexts.

The grid search covered 104,976 combinations. 624 came out feasible. Table 4 shows the selected coefficients and their normalised weights.

Table 4 Calibrated coefficients and normalised requirement weights.

Requirement	Symbol	Coefficient	Normalised Weight
Interpretability	α	0.98	0.266
Robustness	γ	0.95	0.258
Scalability	ζ	0.90	0.245
Representation Capacity	θ	0.85	0.231

Interpretability gets the highest weight (0.266). Our target users are agricultural decision-makers without machine learning background. Robustness follows (0.258), then scalability (0.245). Representation capacity sits at 0.231. We capped it there because the feature set is small and the horizon is short. The expertise level E is fixed at 0.2 for this study. These weights feed the compatibility function. That function scores the alignment between each window’s requirement vector and the capability profiles of the candidate models.

5.2 Data Preprocessing and Feature Extraction

In the preprocessing stage, we ran the raw daily data through a Python script that used the pandas and scipy libraries for data manipulation and statistical computation. The script applied a sliding window approach with a window length of 90 days and a stride of 30 days. That generated 119 overlapping windows across the 14-year period. For each window, we extracted eleven statistical and relational features. These include the first four moments of corn prices and temperatures, daily return volatility, price-temperature correlation, and mean precipitation. The resulting feature matrix has dimensions 119 by 11. It was stored as `window_features.csv` in our GitHub repository (see Data Availability Statement).

5.3 Model Ranking and Performance Evaluation

We ran the ranking and evaluation procedure in Python with standard data processing and machine learning libraries: `pandas`, `NumPy`, `scikit-learn`, `xgboost`, and `lightgbm`. For each 90-day window, the framework computed a requirement vector from the window’s statistical features and the calibrated coefficients ($\alpha = 0.98, \gamma = 0.95, \zeta = 0.90, \theta = 0.85$). This vector got compared against the capability profiles of five candidate models using the Manhattan-distance compatibility function from the first paper. The two models with the highest compatibility scores were kept for performance evaluation.

For each window, the top two models were trained on data from the 90-day window and tested on the 30-day forward horizon right after the window. Training and testing

features were daily maximum temperature, daily minimum temperature, and daily precipitation. The target variable was the daily closing price of corn futures. Both models were evaluated with Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The evaluation ran on all windows where enough test data was available. That gave 75 windows with complete RMSE and MAE records.

Table 5 summarises how often each model appeared as the best (top-ranked) or second-best model across all windows. Three of the five candidate models were ever selected as the top or second model. Random Forest was picked most often. It was the best model in 56 windows (74.7%) and the second model in a handful of others. Hidden Markov Model was the best model in 11 windows (14.7%). KNeighborsTimeSeries was the best model in 8 windows (10.7%). XGBoost and LightGBM never took the top slot. XGBoost showed up as the second model in many windows where Random Forest won.

Table 5 Frequency of model selection as best and second-best.

Model	Best Model (Count)	Best Model (%)	Second Model (Count)
Random Forest	56	74.7	8
Hidden Markov Model	11	14.7	27
KNeighborsTimeSeries	8	10.7	22
XGBoost	0	0.0	18
LightGBM	0	0.0	0

Table 6 reports the performance comparison between the best and second-best models. The average RMSE of the best model across all windows was 39.48. The average RMSE of the second model was 44.91. The best model got lower RMSE than the second model in 51 of 75 windows (68.0%). That shows the framework’s primary recommendation beats its closest alternative in about two-thirds of cases.

Table 6 Summary of RMSE comparison between best and second-best models.

Metric	Best Model	Second Model
Average RMSE	39.48	44.91
Average MAE	34.16	39.62
Windows with lower RMSE	51 (68.0%)	24 (32.0%)

The temporal distribution of model selections showed a clear pattern. Random Forest was mostly selected during periods of high price volatility and extreme weather events. Examples: the 2012 drought and the 2020 pandemic-induced market shocks. KNeighborsTimeSeries and Hidden Markov Model were selected during calmer stretches. Distance-based methods captured the local structure of the data well there.

HMM showed up as the second-best model mostly in windows where KNeighborsTime-Series was the top choice. That suggests HMM acts as a complementary interpretable alternative in stable market conditions.

The RMSE gap between the best and second-best models varied a lot across windows. It ranged from near-zero values to differences above 40. In windows where the gap was small (less than 1.0), the framework’s recommendation was less decisive. Both models performed similarly there. In windows with large gaps, the framework confidently picked a superior model. This suggests the framework’s discriminative power is stronger in volatile market regimes. The differences between model capabilities are more pronounced there.

Overall, the evaluation confirms the model selection framework gives a reliable recommendation in most cases. The best model beat the second model in 68% of the windows. Random Forest dominated. That highlights the suitability of balanced ensemble methods for this dataset. High-capacity models offer no advantage given the limited feature set and the short forecasting horizon.

5.4 Transition Point Identification

Transition point detection was run in Python. The sequence of final models assigned to each window got scanned chronologically. Every instance where the model changed between consecutive windows was recorded. This process found 42 transition points across the whole time horizon. Each transition carries a confidence label. We classify by the RMSE gap at the switch and by how many consecutive windows the new model survives. Five levels come out: high, medium, spike, weak, low. Six transitions land in the high bucket. Eight in medium. Nine in spike, which means a large RMSE gap but the new model lasts only one window. Eight in weak. Eleven in low, which is windowing noise.

5.5 Macro-Level Analysis: Classification of Transition Points

The 42 transition points across the 14-year horizon fell into four categories. Classification rests on observable patterns in the model transition sequences. Known external events from the agricultural and financial literature back it up.

The first category is **climate-driven crises**, mainly severe droughts. These transition points show a sudden jump in price volatility and a sharp rise in the correlation between corn prices and temperature. The framework shifts from interpretable models (KNN or HMM) to high-capacity models like Random Forest. The 2012 drought stands out. Corn prices climbed from roughly 400 USD to over 800 USD per bushel [25]. The literature documents this event well. Studies flag the severe yield losses and the role of ethanol policies in amplifying price volatility [26, 27]. Two high-confidence transitions got recorded between February and April 2012. That reflects the framework trying to adapt to rapidly worsening supply conditions. The 2012-04-24 switch from XGBoost to Random Forest carries a RMSE gap of 6.74 and the new model persists for two windows.

The second category covers **pandemic-induced shocks**, specifically the COVID-19 crisis of 2020. Corn prices dropped to their lowest level in the dataset. The Producer

Price Index (PPI) for corn fell more than 20 percent between February and May 2020 [28]. Transition points here show abrupt shifts from stable models to noise-robust models. The market moved from a normal regime to a highly volatile and unpredictable state [29]. A high-confidence transition from HMM to Random Forest was recorded on 2020-06-06, with an RMSE gap of 14.98. That signals the framework recognising a structural break in market dynamics. This matches findings that the COVID-19 pandemic was a structural breakpoint in global price dynamics. Corn was one of the first commodities to respond to the recession [30].

The third category holds **geopolitical shocks**, most notably the Russia–Ukraine war in 2022. Ukraine is a major global grain exporter. The conflict triggered an immediate price surge of over 75 USD within a short period. Transition points here mostly involve shifts from simpler models to high-representation-capacity models like Random Forest. The correlation between price and temperature also rose a lot during this period. A high-confidence transition from XGBoost to Random Forest on 2021-08-17 carries an RMSE gap of 27.01. That suggests the framework was recalibrating its recommendations for the post-shock market environment. This matches findings that international cereals prices rose by 20% within the first three months of the war. Lower- and middle-income countries were hit hardest by the fragmentation of global trade networks [31].

The fourth category is **smooth transition periods**. These happen between major crises. They show declining volatility and a return of the price–temperature correlation to normal levels [32]. Interpretable models such as KNN or HMM get re-selected as the primary recommendations. Examples include transitions recorded during 2013–2014 and 2017–2018. The market gradually recovered from the drought and financial disturbances and settled into a quieter regime. A high-confidence transition on 2013-09-28 from Random Forest to HMM carries an RMSE gap of 5.39. Another on 2023-01-20 from HMM to Random Forest carries a gap of 9.74.

Classification of Transition Points

34 transition points across 14 years were classified into four categories based on market conditions and changes in model recommendations.



Across all categories, the first three groups (climate, pandemic, geopolitical) together account for roughly 70% of the high and medium confidence points. External shocks are the main drivers of structural changes in corn market dynamics. Smooth transition periods make up the rest. They reflect the market's capacity to return to equilibrium and reduce volatility over time. Six of the 42 points carry high confidence. All six land on Random Forest as the destination. That pattern holds across drought, pandemic, and war. When the market faces a real regime shift, the framework reaches for the balanced ensemble. The rest of the points split across spike, weak, and low. Those come from the 60-day overlap between windows and from the RMSE noise between similar tree ensembles.

5.6 Micro-Level Analysis Results

The macro-level links transitions to big external events. But it does not say which ordinary weather factors push the market from one regime to the next. Micro-level fills that gap.

The dataset is the raw daily file. 3,497 rows after dropping days without a valid forward price. Features: daily maximum temperature, daily minimum temperature, daily precipitation, plus three seasonal dummies (spring, summer, fall). Winter is the reference. Target: the same three-class regime label, computed per day as the 30-day forward return with the $\pm 3\%$ threshold.

We split by time. Last 20% of rows form the test set. No shuffling. Baseline is a most-frequent-class classifier. It scores 32.8% accuracy and 0.165 macro F1. The Random Forest classifier scores 42.5% accuracy and 0.420 macro F1. That is a 9.7 point

gain in accuracy and a 25.5 point gain in macro F1. The model learned something. Not enough to be a strong predictor on its own, but enough to beat the trivial benchmark.

5.6.1 Feature Importance via SHAP

We use SHAP to explain the classifier. TreeExplainer gives per-sample feature contributions. We aggregate by mean absolute SHAP value per feature. Table 7 reports the result.

Table 7 Feature importance from the Random Forest classifier (mean absolute SHAP value).

Feature	Mean —SHAP—	Share
Min_Temp_C	0.0376	28.6%
Max_Temp_C	0.0301	22.9%
is_summer	0.0191	14.6%
is_fall	0.0152	11.6%
is_spring	0.0151	11.5%
Precipitation_mm	0.0131	10.0%

Temperature features dominate. Min and Max temperature together carry 51.5% of the total SHAP mass. Seasonal dummies carry 37.7%. Precipitation carries 10.0%. That ordering is stable. We split the ranking by proximity to a transition. Days within ± 30 days of a transition point versus days outside that band. The difference in mean absolute SHAP sits around 0.002 for every feature. That is below the noise floor. Weather behaves the same near a transition as it does far from one. If weather drove the model shifts, this split would show it. It does not. Transition drivers most likely live in the price side, not in the daily weather.

5.6.2 Classification Performance

Confusion matrix and F1-scores for the Random Forest classifier appear in Table 8.

Table 8 Confusion matrix for the Random Forest classifier (test set).

Actual \ Predicted	Bearish	Bullish	Neutral
Bearish	130	36	60
Bullish	63	68	113
Neutral	77	53	99

Per-class F1: Bearish 0.524, Neutral 0.395, Bullish 0.339. Bearish is the easiest class. Bullish is the hardest. Most of the Bullish errors land in the Neutral column (113 out of 244). The model tends to flatten strong up moves into a neutral call.

That makes sense. Temperature and season carry less signal about upward moves than about downward ones. A dry hot summer is a clear bearish signal. A mild spring is not a clear bullish one. Neutral catches the ambiguous middle, which is why the model over-predicts it in that region.

5.6.3 Weather-Based Performance Analysis

We split test accuracy by weather condition. Four temperature groups and two precipitation groups. Table 9 sums it up.

Table 9 Random Forest accuracy by temperature and precipitation groups (test set).

Weather Type	Group	RF Accuracy
Temperature	≤0°C	0.481
Temperature	0–15°C	0.418
Temperature	15–25°C	0.402
Temperature	≥25°C	0.451
Precipitation	Rainy	0.436
Precipitation	Dry	0.428

Accuracy moves within a ten-point band across all weather groups. Cold days (≤0°C) score highest at 48.1%. Moderate days (15–25°C) score lowest at 40.2%. That fits the story above: extreme cold gives the model a clearer signal. Moderate weather blurs it. Rainy and dry days sit within one point of each other. Rain does not add much either way. These patterns are small. The overall picture is that weather gives a weak but consistent edge. The classifier is not running on noise. It just does not have enough signal to separate the three regimes cleanly from weather alone.

6 Discussion

6.1 Summary of Contributions and System Validation

This study asked whether a training-free model selection framework, first built as a theoretical idea, could work on the agricultural commodity market. The results say yes. The framework performs consistently across 14 years of corn futures data. It also gives actionable intel on when and why forecasting strategies should change. Three key contributions come out of the analysis.

First, the training-free selection framework finds the most suitable model for each 90-day window without iterative training. The framework’s recommendations get validated by comparing the top-ranked model against its closest alternative. The top model achieves lower RMSE than the second model in 68% of the windows. That shows the compatibility-based ranking reliably predicts practical performance.

Second, scanning the sequence of recommended models reveals 42 transition points where the optimal model shifts. Six of those carry high confidence. All six land on Random Forest as the destination. They cluster around major external events: the

2012 drought, the 2020 COVID-19 pandemic, and the 2022 Russia-Ukraine war. This alignment confirms that the framework responds to structural changes in the market. The remaining 36 points split across spike, weak, and low. Those trace back to the 60-day overlap between windows and to RMSE noise between similar tree ensembles.

Third, the micro-level attribution analysis shows temperature is the main weather driver. Minimum and maximum temperature together carry 51.5% of the total SHAP mass. Seasonal dummies carry 37.7%. Precipitation carries just 10.0%. Rain barely matters. That is a clean result. It also matches the agricultural literature. Temperature and season set growth windows. Rain has non-linear and delayed effects that a same-day model won't see.

Random Forest dominates the framework's own recommendations. It wins 74.7% of the evaluated windows. It also dominates the six high-confidence transitions as the destination model. That consistency is not a coincidence. The framework's calibrated weights put interpretability at the top (0.266). Random Forest sits at 0.65 on that axis. High enough to stay in the running, low enough that the other three dimensions decide the tie. On the stability side, it beats XGBoost and LightGBM in the framework's own rankings because the sample size per window is small.

On the micro-level classifier, Random Forest reaches 42.5% accuracy and 0.420 macro F1. The baseline sits at 32.8% and 0.165. The gap is real. The model learned something. It just does not have enough signal in weather alone to separate three regimes cleanly.

6.2 Gaps Addressed and Value Added

This study addresses three specific gaps in the existing literature on agricultural price forecasting.

Most existing approaches rely on a single fixed model or require exhaustive experimentation with multiple models. Both paths cost compute and lack guarantees. Our training-free selection framework gives a cheap alternative. It recommends a model from the statistical properties of each window. No training. No iterative evaluation. The whole ranking runs in seconds on a laptop.

Previous research has largely ignored the fact that optimal forecasting models change over time in response to market conditions. By identifying 42 transition points over a 14-year period, we show that model shifts are not random. They track external shocks: droughts, pandemics, wars. Six of them carry strong confidence. That is the empirical anchor. Adaptive model selection is a practical need in agricultural commodity markets, not a theoretical nicety.

The literature also lacks interpretable explanations for why models shift. Our micro-level attribution gives a clean link between weather and regime. Temperature and season carry most of the signal. Rain carries little. That split is useful. It tells a decision-maker where to look first when building a weather-driven risk model.

The added value breaks down by audience. Researchers get a validated framework that bridges theoretical model selection and real-world agricultural finance. Practitioners get a plain signal on when to switch forecasting models and why. The broader sector gets evidence that data-driven adaptation works under volatile market conditions, even with a small feature set.

6.3 Limitations and Challenges

The results look good. Several limits need stating.

The daily classification dataset uses three weather features (max temp, min temp, precipitation) plus three seasonal dummies. That covers the main climatic drivers. It leaves out economic indicators. Exchange rates, global supply-demand balances, energy prices. Those push commodity markets too. Adding them could sharpen the model’s ability to separate regimes. The Bullish class stays the hardest to call.

The classifier struggles with Bullish. F1 for that class is 0.339. Bearish sits at 0.524. Neutral at 0.395. Bullish is weak because the model flattens strong up moves into Neutral calls. Mild weather says little about upward price pressure. A dry hot summer is a clear bearish signal. A mild spring is not a clear bullish one. Part of this sits in how we define the classes. The $\pm 3\%$ band treats a $+3\%$ and a $+30\%$ move as the same regime. That loses information. A volatility-weighted target would help.

The six high-confidence transitions all land on Random Forest. That is clean. It’s also a sample of six. Drought, pandemic, war. Three event types, six data points. Strong pattern, small n. More events over a longer horizon would test whether the pattern holds.

Near-transition days and far-from-transition days show the same SHAP profile. The difference in mean absolute SHAP sits around 0.002 for every feature. Weather does not change its behaviour near a shift. That tells us the drivers of transition live somewhere else. Likely in the price series, not the weather series. This study does not isolate those drivers. It only rules out weather as the main cause.

Geography limits the dataset. It covers Iowa only. Iowa is a major corn-producing state. The findings may not hold for other regions with different climate or market structure. Testing on other regions and other commodities would strengthen external validity.

The framework’s window overlap (60 days out of every 90) creates its own noise. Roughly a quarter of the transition points fall into the low-confidence bucket. They trace to RMSE gaps under 2 units. A wider stride or a non-overlapping window scheme would cut that noise. It would also shrink the sample. That tradeoff deserves its own study.

6.4 Future Directions

The limits above point to several promising directions for future work. The most immediate extension is a dynamic trigger mechanism. It would monitor the feature space continuously and activate the model selection framework only when a significant shift shows up. That cuts computational costs and enables real-time deployment in agricultural decision support platforms. The threshold-based approach, using distance functions and quantile analysis from our preliminary roadmap, gives a natural starting point for such a trigger.

The framework could also extend to other agricultural commodities. Soybeans, wheat, rice. That would test whether the same compatibility-based ranking holds across different markets. It would give stronger evidence for generalisability. Applying the framework to regions beyond Iowa would test its robustness under different climatic and economic conditions.

FinTech and blockchain integration is a longer-term direction. This study does not address it directly. The framework could pair with mobile-based credit scoring systems, digital wallets, and blockchain-enabled digital identities. That would boost financial inclusion for smallholder farmers and rural agribusinesses. It would give them data-driven risk assessment and tailored financial products [33–35]. Digital Village Bank models use blockchain technology and digital identity verification to empower rural communities. They have already shown potential in building inclusive and transparent financial ecosystems across multiple developing countries [35]. Blockchain-enabled digital identities and data wallets let farmers control and monetise their data. They help farmers access financial services and meet traceability standards. That strengthens their position in global value chains [33]. AgroWeb3, a blockchain and Web3-based solution from IDB Lab and IFAD, shows how digital wallets and interoperable credential verification can cut costs and secure data ownership for farmers. It helps them reach global markets [34]. This study gives the analytical foundation for such integrated systems.

Satellite-based climate data, IoT sensors, and real-time environmental monitoring could sharpen the framework’s predictive power. The current study relies on ground-based weather stations in Iowa. They give a reliable but spatially limited view of environmental conditions. High-resolution satellite data on soil moisture, vegetation health, and evapotranspiration would enable more precise predictions and support regional decision-making [36, 37]. Recent work shows satellite remote sensing data paired with machine learning can estimate soil moisture, evapotranspiration, and crop yield at regional scales [36, 37]. Vegetation indices from MODIS and Landsat data, such as NDVI and NDWI, are widely used to assess vegetation health and soil moisture status in croplands [38]. The infrastructure for such systems already exists in related projects. MQTT-based data pipelines and interactive dashboards [39, 40]. Cyber-physical Digital Twin architectures mix distributed IoT sensor networks with MQTT-based communication protocols. They have shown reliable real-time monitoring and control of greenhouse environments [39]. Scalable IoT architectures for agro-environmental monitoring use lightweight MQTT-based communication protocols for low-latency data transmission to cloud platforms for analytics and visualisation [40]. Plugging the model selection framework into these platforms would be a natural and impactful next step.

7 Conclusion

This study examined whether a training-free model selection framework can work in agricultural commodity markets and provide forecasting recommendations that adapt to changing conditions while remaining interpretable and computationally efficient. Using 14 years of daily corn futures and weather data from Iowa, we found that the best forecasting model changes as market conditions change. These shifts can also be detected and linked to specific climatic factors.

Three findings stand out. First, the training-free selection framework, which was originally proposed as a theoretical approach, worked reliably in a real agricultural setting. Its top-ranked model beat the closest alternative in 68% of the evaluation

windows. That result suggests compatibility-based ranking can provide a practical alternative to computationally costly AutoML methods. Of 104,976 coefficient combinations tested during calibration, 624 came out feasible. The framework then ran on 119 windows without retraining a single model.

Second, we detected 42 transition points between 2010 and 2024. Six carry high confidence. All six land on Random Forest as the destination. They cluster around major external shocks: the 2012 drought, the 2020 pandemic, and the 2022 geopolitical conflict. That alignment is the empirical anchor of the study. Model selection needs to respond to changes in the market. A fixed forecasting model misses these shifts.

Third, the micro-level SHAP analysis on 3,497 daily rows shows temperature is the main weather driver. Minimum and maximum temperature together carry 51.5% of the total SHAP mass. Seasonal dummies carry 37.7%. Precipitation carries 10.0%. This gives the model selection process an interpretable link to weather conditions. Farmers and agribusinesses can see which climate variables track shifts in forecasting behavior instead of relying solely on model outputs.

There are still clear weaknesses. The classifier struggles with the Bullish regime. Its F1-score sits at 0.339, against 0.524 for Bearish and 0.395 for Neutral. The model flattens strong up moves into Neutral calls. Mild weather says little about upward price pressure. A dry hot summer is a clear bearish signal. A mild spring is not a clear bullish one. The feature set helps explain this. It contains three weather variables plus seasonal dummies and leaves out factors such as exchange rates, supply-demand balances, and energy prices. The six high-confidence transitions all landing on Random Forest is a clean pattern. It also rests on six data points across three event types.

Future research should bring economic variables into the framework and test the approach across other commodities and regions. A dynamic trigger mechanism could make the system more suitable for real-time use. Satellite climate data, IoT sensors, and FinTech platforms offer further paths to actual deployment.

Overall, the study shows that adaptive model selection can work in agricultural commodity forecasting and can provide information about why forecasting preferences change. The link between weather conditions and model transitions gives the framework a practical layer that standard model selection methods often lack. This could help farmers manage uncertainty, give agribusinesses another tool for risk management, and support policy decisions as climate volatility continues to affect agricultural markets.

Acknowledgments

The author expresses his gratitude to Kaggle for providing the platform that hosts the dataset used in this study, and to Muhammad Wasil for compiling and making the US Crop Prices & Midwest Climate Data (2010-2024) publicly available. The author also thanks GitHub for providing the infrastructure that enabled version control, collaboration, and open-source dissemination of the analysis code and scripts associated with this research. All computational work was conducted independently, and no external funding or institutional support was received.

Declarations

Conflict of Interest

The author declares no conflict of interest. This research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This research received no external funding. All computational resources and data used in this study were obtained from publicly available sources, and the author conducted the work as part of independent personal research without any institutional or grant support.

Data Availability Statement

The dataset used in this study is publicly available on Kaggle under the title “US Crop Prices & Midwest Climate Data (2010-2024)”, curated by Muhammad Wasil. The dataset can be accessed at: <https://www.kaggle.com/datasets/iconicwasil/us-crop-prices-and-midwest-climate-data-2010-2024?resource=download>. All analysis code and scripts are available in the associated GitHub repository at: <https://github.com/aliNzLami/adaptiveForecast-FinAg>.

Ethics Statement

This study did not involve human subjects, animal experiments, or any sensitive data requiring ethical approval. All data used are publicly available and anonymized.

References

- [1] “Drought causes \$80bn in soybean losses over a decade,” *Valor International*, Dec. 2024. [Online]. Available: <https://valorinternational.globo.com/agribusiness/news/2024/12/09/drought-causes-80bn-in-soybean-losses-over-a-decade.ghml>
- [2] American Farm Bureau Federation, “Farm bankruptcies continued to climb in 2025,” *FB.org*, Feb. 2026. Available: <https://www.fb.org/market-intel/farm-bankruptcies-continued-to-climb-in-2025>
- [3] “Global and African Financial Experts Urge Action to Enhance Smallholder Farmer Financing,” *AfDB*, Mar. 2025. Available: <https://www.afdb.org/en/news-and-events/global-and-african-financial-experts-urge-action-enhance-smallholder-farmer-financing-81871>
- [4] “Four lessons for financial innovation in agrifood systems,” *IFPRI*, Jul. 2025. Available: <https://www.ifpri.org/blog/four-lessons-for-financial-innovation-in-agrifood-systems/>

- [5] “Innovatech 2.0: Digital solutions to boost the productivity and climate resilience of 17,000 crop and livestock producers,” *IFAD*, Feb. 2025. Available: <https://www.ifad.org/en/w/news/innovatech-2.0-digital-solutions-to-boost-the-productivity-and-climate-resilience-of-17-000-crop-and-livestock-producers>
- [6] I. Radic, “Empowering smallholder farmers with blockchain-enabled digital identities: The case of CIMMYT for traceability, financial inclusion and value chain integration,” *CIMMYT*. Available: <https://ideas.repec.org/p/fpr/cgiarp/172557.html>
- [7] “AgroWeb3: A Solution to Transform Smallholder Farmers’ Access to Global Markets,” *BID Lab*, Dec. 2024. [Online]. Available: <https://bidlab.org/en/news/agroweb3-solution-transform-smallholder-farmers-access-global-markets>
- [8] “ILO and AFTECH join forces to boost financial access for dairy farmers through digital innovation,” *ILO*, Aug. 2025. [Online]. Available: <https://www.ilo.org/resource/news/ilo-and-aftech-join-forces-boost-financial-access-dairy-farmers-through>
- [9] Ali N. Lamiry, “A Mathematical Pre-Screening Framework for Training-Free Model Selection in Time-Series Cashflow Forecasting” *Research Square Preprint*, 2026. Available: <https://www.researchsquare.com/article/rs-11051613/v1>
- [10] Maryam Bukhari, Muazzam Maqsood, and Farhan Adil, “An actor-critic based recommender system with context-aware user modeling,” *Artificial Intelligence Review*, vol. 58, no. 138, 2025. doi:10.1007/s10462-025-11134-9. Available: <https://link.springer.com/article/10.1007/s10462-025-11134-9>
- [11] Chunxia Tian, Roengchai Tansuchat, and Songsak Sriboonchitta, “Regime-Aware Stock Index Forecasting Under Latent Market States: A Hybrid Statistical Learning Framework with Cross-Market Validation,” *Forecasting*, vol. 8, no. 3, p. 50, 2026. doi:10.3390/forecast8030050. Available: https://www.researchgate.net/publication/407084843_Regime-Aware_Stock_Index_Forecasting_Under_Latent_Market_States_A_Hybrid_Statistical_Learning_Framework_with_Cross-Market_Validation
- [12] Theofilou, Asterios and Nastis, Stefanos A. and Michailidis, Anastasios and Bournaris, Thomas and Mattas, Konstadinos, “Predicting Prices of Staple Crops Using Machine Learning: A Systematic Review of Studies on Wheat, Corn, and Rice: A Bibliometric Analysis,” *Sustainability*, vol. 17, no. 12, p. 5456, 2025. doi:10.3390/su17125456. Available: <https://www.mdpi.com/2071-1050/17/12/5456>
- [13] Muhammad Wasil, “US Crop Prices & Midwest Climate Data (2010-2024),” Kaggle, 2026. Available: <https://www.kaggle.com/datasets/iconicwasil/us-crop-prices-and-midwest-climate-data-2010-2024?resource=download>

- [14] Ali N. Lamiry, “A Mathematical Pre-Screening Framework for Training-Free Model Selection in Time-Series Forecasting: Matching Data Topology to Algorithmic Bias,” *arXiv preprint*, 2026. Available: <https://www.researchsquare.com/article/rs-11051613/v1>
- [15] Erwan Scornet and Giles Hooker, “Theory of Random Forests,” *Annual Review of Statistics and Its Application*, vol. 13, pp. 99-121, 2026. doi:10.1146/annurev-statistics-112723-034707. Available: <https://www.annualreviews.org/docserver/fulltext/statistics/13/1/annurev-statistics-112723-034707.pdf>
- [16] Radouane Aboulmaouda, Khadija Slimani, and Nour El Houda Chaoui, “Credit Risk Scoring in the Age of AI: A Systematic Comparison of Traditional, ML, and DL Models Based on Accuracy, Stability, and Interpretability,” *Journal of Risk and Financial Management*, vol. 19, no. 7, p. 526, 2026. doi:10.3390/jrfm19070526. Available: <https://www.mdpi.com/1911-8074/19/7/526>
- [17] Zhenni Tan and Yuehua Wu, “On Regime Switching Models,” *Mathematics*, vol. 13, no. 7, p. 1128, 2025. doi:10.3390/math13071128. Available: <https://www.mdpi.com/2227-7390/13/7/1128>
- [18] Krisztián Antal and Margit Antal, “Twenty Years of Time Series Classification: From Nearest-Neighbors to ROCKETS and Large Language Models,” *Acta Universitatis Sapientiae, Informatica*, vol. 18, 2026. Available: <https://m2.mtmt.hu/api/publication/36880484>
- [19] ScienceDirect, “Root Mean Square Error,” ScienceDirect Topics, Earth and Planetary Sciences. Available: <https://www.sciencedirect.com/topics/earth-and-planetary-sciences/root-mean-square-error>
- [20] ScienceDirect, “Mean Absolute Error,” ScienceDirect Topics, Engineering. Available: <https://www.sciencedirect.com/topics/engineering/absolute-error>
- [21] K. Takahashi, K. Yamamoto, A. Kuchiba, and T. Koyama, “Confidence interval for micro-averaged F1 and macro-averaged F1 scores,” *Applied Intelligence*, vol. 52, pp. 4961–4972, 2022. doi:10.1007/s10489-021-02635-5. Available: <https://link.springer.com/article/10.1007/s10489-021-02635-5>
- [22] “Correcting Negative Bias in Large Language Models through Negative Attention Score Alignment,” arXiv:2408.00137, 2024. Available: <https://arxiv.org/abs/2408.00137>
- [23] K. Leung, “Micro, Macro & Weighted Averages of F1 Score, Clearly Explained,” *KDnuggets*, Jan. 2023. Available: <https://www.kdnuggets.com/2023/01/micro-macro-weighted-averages-f1-score-clearly-explained.html>

- [24] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 4765–4774. Available: <https://arxiv.org/abs/1705.07874>
- [25] Carl Zulauf, “2012 Drought: Yield Loss, Revenue Loss, and Harvest Price Option,” *farmdoc daily*, vol. 2, no. 169, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, Aug. 2012. Available: <https://farmdocdaily.illinois.edu/2012/08/2012-drought-yield-loss-revenu.html>
- [26] “Drought, ethanol, and livestock,” *Energy Economics*, vol. 49, pp. 301–307, May 2015. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0140988315000456>
- [27] Sunil P. Dhoubhadel, Azzeddine M. Azzam, and Matthew C. Stockton, “The Impact of Biofuels Policy and Drought on the U.S. Grain and Livestock Markets,” *Journal of Agricultural and Applied Economics*, Cambridge University Press, Feb. 2015. Available: <https://www.cambridge.org/core/journals/journal-of-agricultural-and-applied-economics/article/impact-of-biofuels-policy-and-drought-on-the-us-grain-and-livestock-markets/C634F113115722FBC306A710C9668587>
- [28] U.S. Bureau of Labor Statistics, “The impact of the COVID-19 pandemic on food price indexes and data collection,” *Monthly Labor Review*, 2020. Available: <https://www.bls.gov/opub/mlr/2020/article/the-impact-of-the-covid-19-pandemic-on-food-price-indexes-and-data-collection.htm#2>
- [29] “Grain market volatility surrounding COVID-19 causes uncertainty,” *AgriLife Today*, Texas A&M AgriLife Extension, Apr. 2020. Available: <https://agrilifetoday.tamu.edu/2020/04/12/grain-market-volatility-surrounding-covid-19-causes-uncertainty/#1>
- [30] Modhurima Dey Amin, Syed Badruddoza, and Oscar Sarasty, “Comparing the great recession and COVID-19 using Long Short-Term Memory: A close look into agricultural commodity prices,” *Applied Economic Perspectives and Policy*, vol. 46, no. 4, pp. 1406–1428, 2024. Available: https://econpapers.repec.org/article/wlyapecpp/v_3a46_3ay_3a2024_3ai_3a4_3ap_3a1406-1428.htm
- [31] Nan Jia, Xiang Yu, Xutong Wu, et al., “The Russia-Ukraine war reduced food production and exports with a disparate geographical impact worldwide,” *Communications Earth & Environment*, vol. 5, no. 1, p. 765, 2024. Available: <https://www.nature.com/articles/s43247-024-01915-5#1>
- [32] Atanu Ghoshray and Emi Mise, “Whither the price of coffee? An analysis of trends and shocks since the coffee crisis,” *Empirical Economics*, vol. 68, pp. 2655–2672, 2025. doi:10.1007/s00181-025-02716-1. Available: <https://link.springer.com/article/10.1007/s00181-025-02716-1>

- [33] Ivana Radic, “Empowering smallholder farmers with blockchain-enabled digital identities: The case of CIMMYT for traceability, financial inclusion and value chain integration,” CIMMYT, Dec. 2024. Available: <https://ideas.repec.org/p/fpr/cgiarp/172557.html>
- [34] IDB Lab and IFAD, “AgroWeb3: A Solution to Transform Smallholder Farmers’ Access to Global Markets,” Dec. 2024. Available: <https://bidlab.org/en/news/agroweb3-solution-transform-smallholder-farmers-access-global-markets>
- [35] “Transforming the rural banking system into a digital Village Bank model: a global analysis of opportunities, challenges and governance,” 2026. Available: <https://www.tandfonline.com/doi/full/10.1080/07366981.2025.2524942>
- [36] “Regional Soil Moisture and Evapotranspiration Estimates for Quantifying Irrigation Water Demand,” 2026. Available: <https://mcgill.scholaris.ca/items/c224702a-db6c-46ff-ae3b-fc62255c23de>
- [37] “Drought monitoring using Enhanced Soil Moisture Drought Index (ESMDI) downscaled with deep learning from multi-satellite data for achieving food and water security,” *Physics and Chemistry of the Earth*, vol. 141, 2025. Available: https://www.researchgate.net/publication/397044569_Drought_Monitoring_Using_Enhanced_Soil_Moisture_Drought_Index_ESMDI_Downscaled_with_Deep_Learning_from_Multi-Satellite_Data_for_Achieving_Food_and_Water_Security
- [38] “An inclusive approach to crop soil moisture estimation: Leveraging satellite thermal infrared bands and vegetation indices on Google Earth Engine,” AGRIS, 2026. Available: <https://ideas.repec.org/a/eee/agiwat/v306y2024ics0378377424005080.html>
- [39] “A Cyber-Physical Digital Twin Architecture Integrating IoT Sensor Networks for Real-Time Multi-Parameter Greenhouse Monitoring, 2026. Available: https://www.researchgate.net/publication/406449898_A_Cyber-Physical_Digital_Twin_Architecture_Integrating_IoT_Sensor_Networks_for_Real-Time_Multi-Parameter_Greenhouse_Monitoring
- [40] “Design and Implementation of a Real-Time IoT Architecture for Agro-Environmental Monitoring in Smart Farming, 2026. Available: https://www.researchgate.net/publication/364660588_Design_and_Implementation_of_IoT-Based_Smart-farm_Management_System