

RESEARCH ARTICLE

Machine Learning Corn Price Predictions

Bingzi Jin¹ and Xiaojie Xu^{2,*} ¹Advanced Micro Devices Co., Ltd., China²North Carolina State University, USA

Abstract: For a substantial period of time, a large number of market players have given price projections for the main agricultural commodities a high degree of significance. In this research, we look at the daily published price of corn in order to address the issue. The sample under analysis runs across a period of more than 10 years, from January 2, 2014, to April 10, 2024. The price series under investigation has significant effects on the business sector. In particular, Gaussian process regression models are created for this situation by using Bayesian optimization techniques and cross-validation (CV) methods. As a consequence, the circumstance leads to the development of methodologies that are used for price forecasting. Our empirical forecasting technique produces relatively accurate price projections for the out-of-sample assessment period, which runs from April 6, 2022, to April 10, 2024. For the price of corn, the relative root mean square error was found to be 1.8537%. Because price forecasting models are readily available, investors and governments may make informed decisions about the corn market as they have access to the necessary information.

Keywords: corn price, time-series prediction, GPR, Bayesian optimization, CV

1. Introduction

Projections of price time series of major agricultural commodities remain a major concern for the business and government sectors for the time being [1]. It will be necessary to obtain estimates of the price time series of certain commodities, like corn, in order to facilitate the construction of informed strategies and reduce the level of uncertainty that now exists [2]. This is due to the fact that many countries and areas mostly depend on agricultural products to meet their strategic objectives. For the vast majority of businesses involved in the commodities markets, being able to get information about price expectations is essential to making informed decisions [3]. One facet of this is the ability of predictive analytics to provide valuable insights into potential future pricing strategies that commodities processors may choose to use [4]. Price estimates may be used to provide trade partners with pricing information, assisting them in selecting the terms of the contract [5]. Participants in the cash and futures markets, including traders and investors, may benefit from the use of predictive data [6]. As shown by the conclusions of risk analysts, forecast data are a crucial part of effective risk management strategies [7]. A large and sustained shift in commodity prices [8] might have a substantial impact on society's general well-being and productivity [9]. It seems that little incentive is needed to understand the significance of time series projections for commodity prices, especially for corn [10].

In the published papers, a variety of time series strategies for predicting price series of various commodities were studied and assessed in order to achieve this goal [11]. The creation of vector

autoregressive (VAR) models, autoregressive integrated moving average (ARIMA) models, and price series forecasting methods akin to vector error correction models (VECMs) was largely the result of earlier research. While several approaches provide further customizations and improvements [12–14], the majority of these procedures may be more challenging. Forecasting has been made possible since its creation by the univariate ARIMA model, which makes use of historical data obtained from previously gathered values. In terms of prediction information sets, VAR approaches outperform ARIMA models with respect to the relationships between several integrated economic components. VECM models perform better than VAR models in illuminating the long-term relationships between a number of economic parameters. To this end, the model incorporates the notion of cointegration explicitly. Stated differently, a major driving force for the creation of the VECM was the inclusion of projections for the furthest future.

The various technical breakthroughs over the previous many years have enabled us to have access to far more formidable processing power than we did even a few years ago [15]. In line with this trend, the domains of finance and economics are paying more attention to artificial intelligence (AI) and machine learning (ML) technologies [16, 17]. Moreover, recent research has shown the significant potential for anticipating trends in time series of commodity prices via the use of ML and AI approaches [18, 19]. Using the agriculture industry sector as an example, these ML-based price forecast exercises have covered corn, soybeans, cotton, wheat, canola, oats, soybean oil, coffee, and sugar [20–27]. In relation to the energy sector, these ML-based price prediction exercises have considered crudeoil, electricity, energy securities, heating oil, coal, carbon emission allowances, and natural gas [28–31]. Concerning the metals sector, these ML-based price

*Corresponding author: Xiaojie Xu, North Carolina State University, USA. Email: xxu6@alumni.ncsu.edu

forecast studies have explored palladium, platinum, steel, silver, lead, copper, and gold [32–37]. Previous studies, as reported in the literature, indicate that academics frequently employ neural networks as modeling tools for tackling issues pertaining to commodity price projections. Even though we only looked at a limited sample size, our results generally support previous studies showing the advantages of using AI/ML approaches in financial and economic forecasting. Past research on the price forecasts generated by Gaussian process regressions (GPRs) employing time-series data gathered for various commodities, such as corn prices, has not been sufficiently conducted. In the present work, an attempt is made to tackle this prediction challenge.

Neal [38] created a novel regression technique by combining the construction of neural networks with Bayesian learning. This approach is mostly focused on using priors to analyze noisy data over Gaussian processes (GP). There is evidence that certain Bayesian regression approaches that are based on neural networks have a tendency to converge to GP in the domain of infinite network borders [38]. An illustration of how to use GPRs in situations that are both chaotic and noise-free [39] is provided by this modeling technique. A demonstration of how Bayesian learning may be used in the context of stationary time-series data is carried out by Brahim-Belhouari and Vesin [40] via the utilization of a study of the performance of radial basis function neural networks with GP. It is essential that a broad range of distinct covariance functions be used in this study in light of the findings of the research conducted by Brahim-Belhouari and Bermak [41]. The use of GP approaches may eventually prove to be quite helpful in the process of producing predictions from time series that already exhibit nonstationary patterns. Based on the research results that Brahim-Belhouari and Bermak [41] conducted, there is a high likelihood that GPR techniques will provide better results than radial basis function neural networks. Furthermore, the GP formulations have the benefit of using the precise matrix operations needed for the integration of the prior and noise models [41]. Model-averaging approaches that are similar to the GP predictors used by Brahim-Belhouari and Bermak [41] in their multi-model predictions would be applied to our situation.

The approaches connected to predictions that were developed in conjunction with this investigation were based on the GPRs. We use the information gathered between January 2, 2014, and April 10, 2024, to examine the corn prices' daily time series. We integrate the model estimation and training processes, using these techniques in the framework of CV and Bayesian optimization techniques [42]. For the out-of-sample period from April 6, 2022, to April 10, 2024, the models under examination show a good degree of projected accuracy, with a relative root mean square error (RRMSE) of 1.8537%. The empirical data may be used to provide technical predictions for relevant policy research, either by itself or in combination with other expected results. Other economic sectors dealing with comparable prediction challenges may find similar forecasting tools useful. In general, our work demonstrates the systematic configuration and evaluation of the GPR method for forecasting daily corn prices. The value of the study lies in combining the selection of kernels and basis functions with Bayesian optimization, CV, and forecast evaluation using a long time series.

Based on the identified research gaps and the promising but underexplored potential of GPRs in commodity price forecasting, the present study is guided by three primary research questions. First, can GPR models, utilizing a 10-day lag structure, generate accurate one-day-ahead out-of-sample forecasts for daily corn prices? Second, how does the choice of covariance

function and basis function within the GPR framework impact forecast accuracy for this particular price series? Third, what is the quantifiable level of predictive accuracy, as measured by the RRMSE, that can be achieved by GPR models in this context? Furthermore, the assertion of novelty in this work is not predicated on the mere application of GPR but on the rigorous and systematic justification for its use as a potentially superior tool for this specific forecasting problem. While numerous ML models, such as neural networks and support vector machines, have been applied to commodity price forecasting, they are often treated as “black boxes,” and their performance can be sensitive to hyperparameter tuning and the risk of overfitting. In contrast, GPR offers a principled, Bayesian framework that provides price forecasts of high accuracy, which are invaluable for quantifying volatile market data. The nature of GPR enables policymakers and market participants to make more robust, risk-aware decisions. This is a significant advantage over many other deterministic ML models that are often employed. Moreover, the flexibility of GPRs, achieved through the selection of various covariance (kernel) functions, provides a powerful mechanism to model the complex, nonlinear patterns that are characteristic of agricultural price time series. As the Brock–Dechert–Scheinkman (BDS) test [43] in this study confirms the presence of nonlinearities in the corn price data, the ability of GPR to flexibly model such structures through its kernel is a key advantage. Unlike some other ML methods that may require extensive feature engineering, GPRs can capture complex dependencies directly through the kernel function, which can be tailored to the underlying data patterns. This is a more elegant and data-efficient approach compared to the arbitrary architectures of deep learning (DL) models. While DL models have demonstrated success, they typically require vast amounts of data and can be computationally demanding, which may not be available or feasible for all time-series forecasting tasks. GPR, by contrast, is a nonparametric model that becomes more expressive as data increases while maintaining its interpretation. The Bayesian nature of GPR also inherently provides a form of regularization, mitigating overfitting and enhancing generalization to unseen data, a property particularly advantageous for out-of-sample forecasting. Therefore, the novelty of this research lies in the comprehensive evaluation of GPR for corn price forecasting, employing a systematic approach to kernel and basis function selection within a rigorous CV and Bayesian optimization framework. This represents a significant and methodologically sound step toward establishing GPR as a dependable, interpretable, and highly accurate tool for agricultural commodity price prediction, addressing the limitations of previous models and setting a new benchmark for future work in this domain.

The remainder of this paper is organized as follows. Section 2 provides a review of the relevant literature, covering both traditional econometric approaches—such as ARIMA, VAR, and VECM models—and recent ML applications for commodity price forecasting, thereby situating our proposed GPR framework within the existing body of knowledge. Section 3 describes the daily corn price dataset utilized in this study, presents summary statistics and distributional properties, and reports the results of the BDS nonlinearity test, which motivates the adoption of a flexible nonlinear modeling approach. Section 4 details the methodological framework, including the specification of GPRs, the alternative isotropic and nonisotropic kernel functions and basis functions under consideration, the Bayesian optimization and CV procedures used for model training and parameter selection, and the forecast evaluation

metrics employed. Section 5 presents the empirical results, encompassing the optimal model selection, out-of-sample forecast performance, comparative benchmark analysis against autoregressive, support vector regression (SVR), and regression tree (RT) models, and a discussion of the economic implications and interpretability of the findings. Finally, Section 6 concludes the paper by summarizing the key contributions, acknowledging the limitations of the current univariate framework, and offering directions for future research, particularly the extension to multivariate predictors that incorporate macroeconomic, energy, and weather variables.

2. Literature Review

For a variety of businesses that rely on commodities, economists have had to devote a great deal of time and effort to overcoming the difficulty of developing accurate and dependable commodity price forecasts [44]. For example, despite their widespread usage in earlier research, ARIMA models remain in great demand for a variety of prediction tasks, such as price data extraction from time series databases. For example, Lv and Wu [45] built a seasonal ARIMA model for predicting corn futures contracts' settlement prices for Dalian Commodity Exchange and found that the Seasonal Autoregressive Integrated Moving Average (SARIMA) (2, 1, 0)(3, 1, 1)12 structure leads to decent prediction accuracy. Similarly, Satapathy et al. [46] constructed the ARIMA(3, 1, 2) model for corn price forecasts for Nabarangur District of Odisha and found that this model results in a good level of forecast accuracy. The VAR technique is an econometric tool that is frequently employed for price series forecasts. It utilizes the relationships that exist between significant economic aspects. For instance, Pierre and Kaminski [47] utilized a global VAR model in analyzing price shock propagation across 27 domestic corn markets of sub-Saharan Africa and found that the majority of price time series are more responsive to those of regional neighbors rather than global shocks, suggesting that it could be helpful to improve price forecast accuracy by effectively exploring local price information. de Albuquerque et al. [48] considered corn price forecasting issues based upon threshold autoregressive models and concluded that these models could help the incorporation of transition regimes and potential nonlinearities. Cointegration enables long-term correlations or linkages between several economic variables to be included in VAR-based models, often referred to as VECMs. This might have a major effect on many price time series' long-term estimates. For example, Guo and Tanaka [49] applied the VECM with the Generalized Autoregressive Conditional Heteroskedasticity-Baba, Engle, Kraft, and Kroner (BEKK-GARCH) and Generalized Autoregressive Conditional Heteroskedasticity-Dynamic Conditional Correlation (DCC-GARCH) specifications to investigate price dynamics between global and US prices of corn, soybeans, and wheat and determined that global prices of corn and wheat lead those of the USA, shedding light on price predictive information running from the global to US market. Jyothi Shiva Kumar and Kotreshwar [50] assessed the Indian corn spot and futures prices through the VECM and found that the predictive information runs unidirectionally from the futures to the spot market.

The number of studies on the use of ML approaches to commodity price forecasting that have been carried out in recent years seems to have increased [51]. For instance, Singh and Sindhu [52] built an RT model for forecasting corn prices. Mayabi [53] found

that the neural network model could also generate decent corn price forecasts. Teste et al. [54] used satellite data with variable dimensions reduced for predicting corn prices and yields through the construction of the convolutional neural network (CNN) model. Jain et al. [55] assessed the importance of data quality when forecasting corn and tomato prices through the use of the long short-term memory (LSTM) neural network model. Sanusi et al. [56] compared the ARIMA and LSTM models for forecasting prices of several agricultural commodities, such as corn, wheat, and beans, and determined that the optimal model in terms of forecast accuracy depends on the specific price series and commodity under consideration. Choudhary et al. [57] combined the variational mode decomposition with the time delay neural network model for forecasting prices of corn, soybeans, and palm oil and found that the hybrid strategy could help improve forecast accuracy. Wu et al. [58] proposed a hybrid interpretable framework constituted of the ensemble empirical mode decomposition, CNN model, adaptive differential evolution with optional external archiving, and temporal fusion transformers for effectively forecasting corn prices by utilizing online data. Murugesan and Radha [59] propose a hybrid ensemble model for crop price prediction, combining the ARIMA statistical model with the SVR. The methodology is implemented in three distinct configurations: The first one applies SVR to ARIMA residuals, summing the results with ARIMA predictions; the second one uses ARIMA's forecasted values as an additional input feature to SVR; and the third one incorporates both the forecasted values and residuals from ARIMA as inputs to SVR, with the final prediction being the sum of SVR's forecasted value and residual. The models were trained and tested on a 10-year coconut price dataset from Tamil Nadu, India, with 80% training and 20% testing splits. The key findings reveal that the third configuration outperforms the others, achieving the lowest prediction errors (mean absolute error (MAE) = 1.938, root mean square error (RMSE) = 2.395) and a deviation of 13.37% from actual observations, compared to 14.68% and 16.48% for the first and second configurations. However, potential gaps remain: The ARIMA foundation imposes a stationarity prerequisite, rendering the framework vulnerable to abrupt market shocks, policy interventions, or seasonal anomalies that violate assumptions. Moreover, the architecture demands extensive optimization for SVR, possibly hampering cross-market generalizability. GPR offers a compelling alternative to circumvent these constraints by leveraging custom composite kernels that explicitly encode covariates directly into the covariance structure, thereby circumventing ARIMA's rigid stationarity requirements. Unlike SVR's isolated tuning, GPR's Bayesian nonparametric learning adapts its complexity to the available samples, facilitating smoother modeling and enabling efficient parameter estimates, which alleviates the small-sample and generalizability bottlenecks without necessitating exhaustive manual recalibration for each new market scenario. For the purpose of addressing the problem that is associated with corn price forecasting, the goal of this research project is to construct GPR models and employ the constructed models for performing forecasting exercises. When it comes to acquiring knowledge of the cost of agricultural commodities, the price of corn is an essential component that is important for both policymakers and market participants to consider. Furthermore, the degree to which the price of corn affects the economy is significant. For the goal of aiding decision-making and the study of price patterns, these models need to produce results that are reliable and accurate on the basis of their predictions.

3. Data

In this specific inquiry, Figure 1 displays the price of corn on a daily basis for the period from January 2, 2014, to April 10, 2024, which is sourced from Macrotrends¹. There are no data reported for weekends or non-trading days. The sourced data do not contain missing or duplicate observations. The prices presented are futures prices, all referencing front-month contracts. They are adjusted using a calendar-weighted method, rolling contracts on the first of each month to build a continuous price history. This calendar-weighted approach achieves a smooth handoff from one contract to the next by employing blended or weighted-average combined prices over a predetermined transition window centered on the roll date. The technique serves as a compromise between the forward Panama Canal method and the backward Panama Canal method. Using the first day of the month as the roll date and calendar-weighted rolling as the adjustment rule makes the process fully deterministic, predictable, and seamless, without contaminating the economic content of the price history, so the resulting series is well-suited for economic forecasting. Corn prices are measured in dollars per bushel. Another thing to keep in mind is that the first price differences are shown in this chart. Furthermore, Figure 1 presents 50-bin histograms, kernel estimates, and quantile–quantile plots for the prices and first differences. An overview of the price data's time series is shown in Table 1. It is evident from the findings of the Jarque–Bera and Anderson–Darling tests that the prices do not correspond to a normal distribution. Considering the features of many financial and economic series, this outcome wouldn't be surprising. The right skewness and leptokurtic structure of the price time series could be identified.

The appearance of nonlinear features in a variety of time series related to the financial and economic realms has been the subject of extensive research. The daily corn price data are

subjected to the BDS [43] test analysis to look for any potential nonlinearities. It will be necessary to apply the ϵ distance in addition to the embedding dimensions, which vary from 2 to 10, in order to conclude this inquiry. The standard deviation of the prices will be multiplied by the ϵ distance, which may be represented by the following numbers—0.5, 1.0, 1.5, 2.0, 2.5, and 3.0—to get the distances that separate the various data points. All test results showed the p -value to be extremely close to zero. It is important to note that these data support the hypothesis that there are nonlinearities in the price series. In this study, prices were estimated by using GPRs, which took into consideration the nonlinear aspects of the data.

As historical prices are used as predictors in our work, the data processing involves generating lagged variables, aligning dates, removing necessary initial observations after the lags were created, and dividing the series into training and testing periods.

4. Method

4.1. GPR

The primary instrument used in the prediction techniques in this study is Gaussian process regressions, or GPRs. GPR is a kind of probabilistic kernel structure that finds use in several scientific domains and may be utilized to forecast different nonlinear events. The equation $\{(x_i, y_i); i = 1, 2, \dots, T\}$ is used by the model in order to accurately represent the unknown distribution in relation to the training data. By using the equations $y_i \in R$ and $x_i \in R^d$, respectively, the target variable and the d -dimension predictor are expressed. For this scenario, the 10-lag prices will serve as the prediction components. Stated differently, the eleventh day's price would be predicted using the prices from the preceding 10 days. We experimented with several alternative numbers of price lags as predictors, including 5, 15, and 20 days, roughly corresponding to one week, three weeks, and one month. We found that using fewer lags (i.e., 5) speeds up model training and forecasting but comes

Figure 1
Daily corn prices from January 2, 2014, to April 10, 2024: plots of price time series and distributions

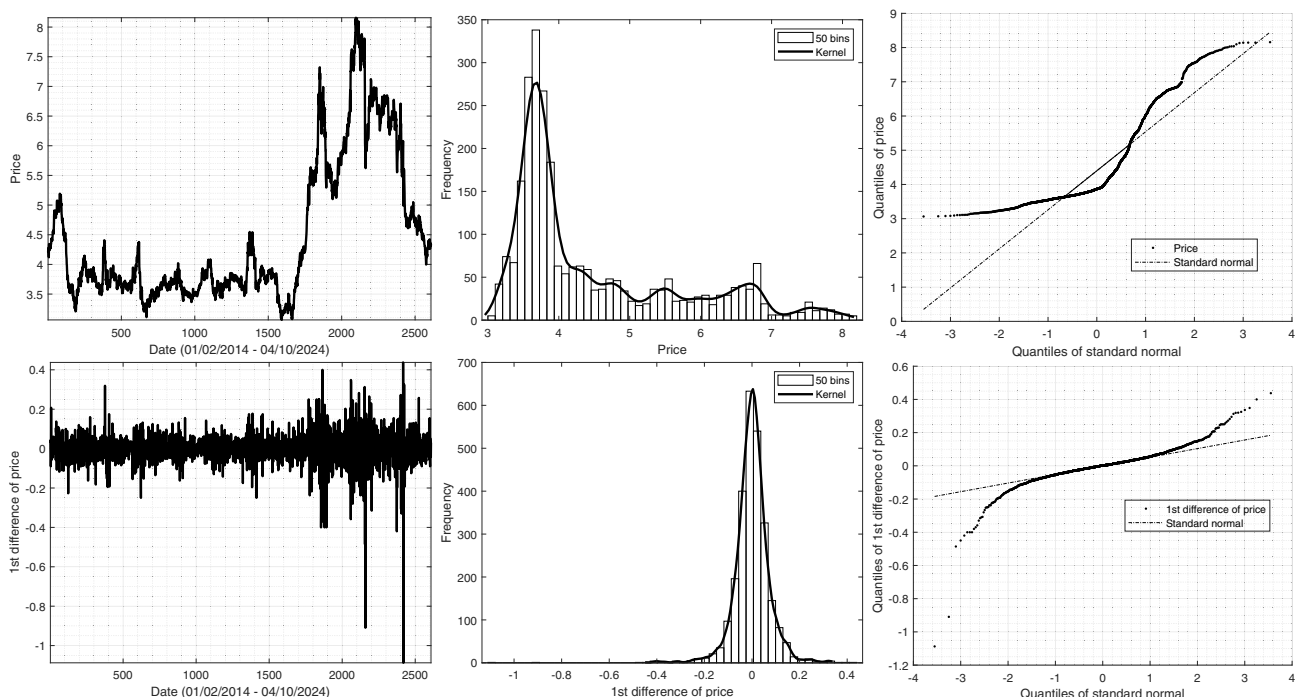


Table 1
Summary statistics of daily corn prices from January 2, 2014 to April 10, 2024

Series	Minimum	1st percentile	5th percentile	Mean	Median	Standard deviation	95th percentile	99th percentile	Maximum	Skewness	Kurtosis	Jarque-Bera	Ander-son-Darling
Price	3.064500	3.178370	3.319750	4.477344	3.867000	1.210883	6.850000	7.833550	8.157500	1.195282	3.269372	<0.001	<0.0005
First differences	-1.087500	-0.222500	-0.107500	0.000053	0.002000	0.077938	0.110000	0.206075	0.437500	-1.902115	28.670905	<0.001	<0.0005

at the cost of lower accuracy. Conversely, using more lags (i.e., 15 or 20) slows down both training and forecasting, yet it also results in reduced accuracy.

4.2. Kernel and basis function

These formulae, denoted $y = x^T \beta + \varepsilon$ and $\varepsilon \sim N(0, \sigma^2)$, respectively, indicate the linear regression and the error. However, basis functions and latent variables offer target variable estimations for GPRs. Basis functions are denoted with b , while latent variables satisfying the joint Gaussian criteria are denoted with $l(x_i)$. The target variable's smoothness may be assessed using the covariance functions of the latent variables. It is possible to use basis functions to map the predictors to feature space. The variance and mean metrics are often employed to create GP, which are represented by the notation *GPS*. $m(x) = E(l(x))$ should be used to display the mean, while $k(x, x') = \text{Cov}[l(x), l(x')]$ may be used to display covariance. Using the $y = b(x)^T \beta + l(x)$ formula, one may express GPRs, where $b(x) \in R^p$ and $l(x) \sim GP(0, k(x, x'))$. Using the $k(x, x' | \theta)$ hyperparameter is a useful way to make sure that the numerical value of θ is produced accurately. GPR training often makes use of estimates for β , σ^2 , and θ . Throughout training, the model obtains the required kernels (k) and basis functions (b). This study is mostly concerned with isotropic and nonisotropic kernels. For these two kinds of kernels, we look at five distinct kernel configurations. Equations (1) through (10) provide a thorough explanation of every single kernel that was examined throughout this inquiry. α , a positive number, represents the scale-mixture parameter. Isotropic kernel characteristic length scales are represented by the metric σ_l . The standard deviation of the signal is represented by the metric σ_f , and $r = \sqrt{(x_i - x_j)(x_i - x_j)}$. According to $\theta = (\theta_1, \theta_2) = (\log \sigma_l, \log \sigma_f)$, there is a guarantee that σ_f and σ_l will be greater than zero. One may denote the length scale of a nonisotropic kernel using the notation σ_m ($m = 1, 2, \dots, d$). Taking this into account, $\theta = (\theta_1, \theta_2, \dots, \theta_{d+1}) = (\log \sigma_1, \log \sigma_2, \dots, \log \sigma_d, \log \sigma_f)$ would reflect θ .

$$\text{Isotropic Exponential: } k(x_i, x_j | \theta) = \sigma_f^2 e^{-\frac{r}{\sigma_l}} \quad (1)$$

$$\begin{aligned} \text{Isotropic Squared Exponential: } k(x_i, x_j | \theta) \\ = \sigma_f^2 e^{-\frac{1}{2} \frac{(x_i - x_j)^T (x_i - x_j)}{\sigma_l^2}} \end{aligned} \quad (2)$$

$$\begin{aligned} \text{Isotropic Matern } \frac{5}{2}: k(x_i, x_j | \theta) \\ = \sigma_f^2 \left(1 + \frac{\sqrt{5}r}{\sigma_l} + \frac{5r^2}{3\sigma_l^2} \right) e^{-\frac{\sqrt{5}r}{\sigma_l}} \end{aligned} \quad (3)$$

$$\text{Isotropic Rational Quadratic: } k(x_i, x_j | \theta) = \sigma_f^2 \left(1 + \frac{r^2}{2\alpha\sigma_l^2} \right)^{-\alpha} \quad (4)$$

$$\text{Isotropic Matern } \frac{3}{2}: k(x_i, x_j | \theta) = \sigma_f^2 \left(1 + \frac{\sqrt{3}r}{\sigma_l} \right) e^{-\frac{\sqrt{3}r}{\sigma_l}} \quad (5)$$

Nonisotropic Squared Exponential:

$$k(x_i, x_j | \theta) = \sigma_f^2 e^{-\sqrt{\sum_{m=1}^d \frac{(x_{im} - x_{jm})^2}{\sigma_m^2}}} \quad (6)$$

Nonisotropic Squared Exponential:

$$k(x_i, x_j | \theta) = \sigma_f^2 e^{-\frac{1}{2} \sum_{m=1}^d \frac{(x_{im} - x_{jm})^2}{\sigma_m^2}} \quad (7)$$

Nonisotropic Matern $\frac{5}{2}$: $k(x_i, x_j | \theta)$

$$\begin{aligned} = \sigma_f^2 \left(1 + \sqrt{5 \sum_{m=1}^d \frac{(x_{im} - x_{jm})^2}{\sigma_m^2}} + \frac{5}{3} \sum_{m=1}^d \frac{(x_{im} - x_{jm})}{\sigma_m^2} \right) \\ e^{-\sqrt{5 \sum_{m=1}^d \frac{(x_{im} - x_{jm})^2}{\sigma_m^2}}} \end{aligned} \quad (8)$$

Nonisotropic Rational Quadratic: $k(x_i, x_j | \theta)$

$$= \sigma_f^2 \left(1 + \frac{1}{2\alpha} \sum_{m=1}^d \frac{(x_{im} - x_{jm})^2}{\sigma_m^2} \right)^{-\alpha} \quad (9)$$

Nonisotropic Matern $\frac{3}{2}$: $k(x_i, x_j | \theta)$

$$= \sigma_f^2 \left(1 + \sqrt{3 \sum_{m=1}^d \frac{(x_{im} - x_{jm})^2}{\sigma_m^2}} \right) e^{-\sqrt{3 \sum_{m=1}^d \frac{(x_{im} - x_{jm})^2}{\sigma_m^2}}} \quad (10)$$

Equations (11)–(14) display the four distinct basis functions, which are the focus of this investigation. The study of these functions is comparable to the scenario in which several kernels are taken into account. With reference to the four basis functions that are being

examined, $X = (x_1, x_2, \dots, x_n)'$, $X^2 = \begin{pmatrix} x_{11}^2 & x_{12}^2 & \dots & x_{1d}^2 \\ x_{21}^2 & x_{22}^2 & \dots & x_{2d}^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1}^2 & x_{n2}^2 & \dots & x_{nd}^2 \end{pmatrix}$, and

$B = (b(x_1), b(x_2), \dots, b(x_n))'$.

Empty : $B = \text{Empty Matrix}$ (11)

Constant : $B = I_{n \times 1}$ (12)

$$\text{Linear : } B = [1, X] \quad (13)$$

$$\text{Pure QuadraticL } B = [1, X, X^2] \quad (14)$$

4.3. Bayesian optimization

The expected improvement per second plus (EIPSP) method and ten-fold (here the name of ten-fold is used to correspond to the 10 sub-models developed and reported in Table 2 in Section 5) CV procedures are used in Bayesian optimizations in order to determine the parameters of the model. During the training phase (January 02, 2014–April 5, 2022), 11 sequential non-overlapping time blocks are created following the ordering of time (i.e., sequential time blocks). There is no overlap across these different folds. More specifically, to ensure strict chronological integrity during model selection and hyperparameter tuning, we implemented a forward-expanding window CV scheme within the training phase. The training period was partitioned into 11 sequential, non-overlapping time blocks of equal size, denoted $B_1, B_2, \dots, B_{11}$, sorted from earliest to latest. In this CV procedure, for each fold $fd = 1, \dots, 10$, the training sample comprises the concatenated blocks $B_1 \cup B_2 \cup \dots \cup B_{fd}$, and the validation sample is the immediately subsequent block B_{fd+1} . This design guarantees that in every fold, all training observations chronologically precede all validation observations. The out-of-sample period ranges from April 6, 2022, to ending on April 10, 2024, which has no overlap with the training phase. There is no information leakage from future periods. We will utilize a GP model that uses the notation $f(x)$ to help offer some understanding of the situation. Following the random selection of the data points that fall within the variable boundaries, or x_i 's, Bayesian methods are used to evaluate the associated $y_i = f(x_i)$. In this case, the initial assessment would make use of the value of $N_s = 4$. The procedure will continue gathering data points in the event that an assessment includes faults until N_s examples of successful assessments are located. Then, in order to put the strategy into action, the next two phases will be completed. A posterior distribution over $Q(f|x_i, y_i \text{ for } i = 1, \dots, T)$ may be generated by utilizing the updates of the function $f(x)$. The primary objective of the second phase, which aims to minimize the acquisition function— $a(x)$, is to locate a new data point, denoted by x . $a(x)$ is one instrument that may be used to find the amount

that x is absorbed into Q . Instead of concentrating on any value that may raise it, the expected improvement acquisition functions would work to enhance the target function. The posterior mean generated at x_{best} is said to be the lowest that may be attained. $\mu_Q(x_{best})$ represents the value linking with x_{best} , which is the lowest. $EI(x, Q) = E_Q[\max(0, \mu_Q(x_{best}) - f(x))]$ might be used to represent the expected improvement, or EI . The acquisition functions of Bayesian processes may use time-weighting algorithms to maximize the benefits for each time unit, since the target assessment time changes based on the location. As optimization methods develop, one may use Bayesian models to calculate the goals' assessment time as a function of x . The EI per second, or $EIPS$, of the acquisition functions could be indicated by the value $EIPS(x) = EI_Q(x)/\mu_S(x)$. In terms of the extra-time GP, $\mu_S(x)$ stands for the posterior mean. As a result, these behavioral adjustments stop acquisition functions from enforcing local minimums or working on certain target regions. The $\sigma_Q^2(x) = \sigma F(x)^2 + \sigma PN^2$ requirement for additive noises would be satisfied by the posterior standard deviation. The standard deviation of the posterior goal with respect to x is denoted by $\sigma_F(x)$. The formula for the exploration ratio is $t_{\sigma PN} > 0$. At the end of each cycle, the acquisition function applies the *EIPSP* approach to confirm that the given data point, x , fits the $\sigma_F(x) < t_{\sigma PN} \sigma_{PN}$ condition. The kernel function expands by the number of iterations when this condition is met, indicating that x is being overexploited [60]. The value of σ_Q may significantly increase as data points migrate across observations when the EIPSP method is used. This suggests that the newly fitted kernel has to be used to produce a new data point. In the event that the subsequent experiment demonstrates that θ is exploitative, it will increase by a factor of 10. Using this method, finding an x that is not overexploited could be done up to five times. Finally, the updated x that will be the next exploration ratio is produced using the EIPSP technique. To get a more precise result, the algorithm makes advantage of both previously analyzed data points and later gathered data points located nearby. When doing Bayesian optimization, it is essential to assess the situation of basis functions, σ , kernels, and predictor standardization.

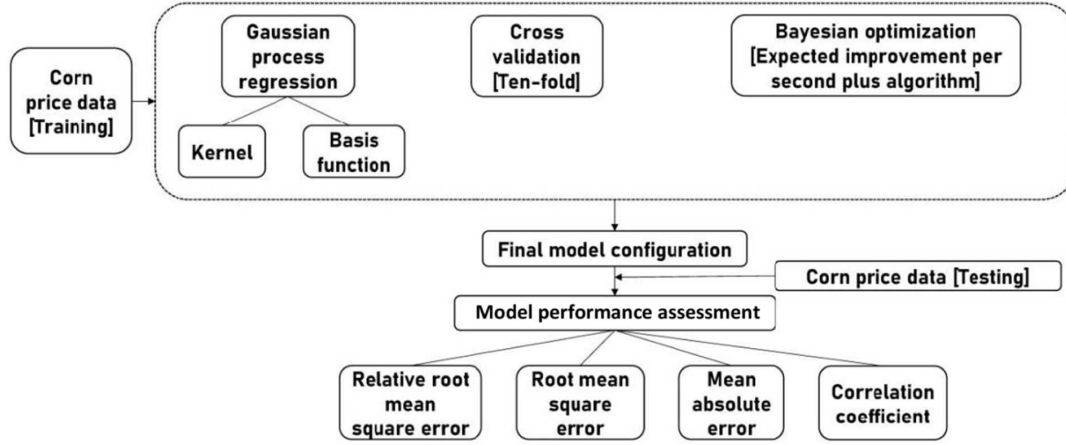
4.4. Performance measurement

The RRMSE is a frequently used statistic for assessing prediction accuracy. This error measure is often used to assess how

Table 2
Model parameter estimates

	CV1	CV2	CV3	CV4	CV5	CV6	CV7	CV8	CV9	CV10
σ	0.07735029332	0.07764086533	0.07679349742	0.07888140298	0.07240269964	0.07897881128	0.07157340188	0.07943484426	0.07518596640	0.07731925256
β_0	0.01045898424	0.00776902856	0.00684509252	0.00815390161	0.00537823456	0.01443527679	-0.00377609120	0.01466921558	-0.00036040910	0.01370607332
β_1	0.01694820168	-0.00270499388	0.01537396094	0.03635696524	0.04375367101	0.02752426932	-0.01243070186	0.00743690413	0.03698506722	0.04433079509
β_2	-0.08923356800	-0.09902650395	-0.08128417363	-0.12819123478	-0.10239598737	-0.12324307465	-0.04764126045	-0.10169800164	-0.11670056364	-0.14986945895
β_3	0.09657445238	0.13975074122	0.07070085033	0.10119081493	0.06322701974	0.10367019227	0.04299554012	0.07901835241	0.09152349959	0.13726538459
β_4	0.02393355044	-0.00341242019	0.05436943116	0.01078920533	0.03146204083	0.03219309671	0.05988809047	0.06202518112	0.03806335086	-0.01420909223
β_5	0.01739086635	0.03616405773	-0.02184277378	0.07806178278	0.00928153587	0.05204998976	0.00889670331	0.00901646402	-0.01537996508	0.03815550964
β_6	-0.02138169691	-0.04028966769	0.00187987292	-0.06223761249	0.00752683385	-0.05284188007	-0.00266261878	-0.01399734164	-0.00609944087	-0.02197603721
β_7	-0.17270139763	-0.13094114889	-0.14491386558	-0.16950181140	-0.15261586582	-0.16540630663	-0.12339916356	-0.19480289708	-0.14919351471	-0.17097581881
β_8	0.19229361744	0.13550357287	0.14482204240	0.14820612526	0.09902289253	0.18918474080	0.00894511052	0.19151249301	0.17647341825	0.19560955205
β_9	-0.07843809897	-0.03683371912	-0.00364398764	-0.00429743641	0.03112157699	-0.02869693501	0.08302344318	-0.06656087604	-0.01525833338	-0.07848022850
β_{10}	1.01229609034	0.9996338868	0.96307407693	0.98789779837	0.96846018753	0.96229831821	0.98347646510	1.02460537465	0.96010418711	1.01698584533
σ_f	0.00724412387	0.00724416344	0.00724416246	0.00724412481	0.00724415354	0.00724415159	0.00724412138	0.00724417645	0.00724412235	0.00724417909
σ_f	0.00133170614	0.00133768912	0.00132033057	0.00136377406	0.00123581567	0.00136585925	0.00122085025	0.00137567975	0.00128833753	0.00133106315

Figure 2
The workflow diagram of modeling processes



well various predictions made from various models and goals perform [61]. The following is the presentation of the RRMSE:

$$RRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i^{obs} - y_i^{for})^2}}{\frac{1}{n} \sum_{i=1}^n y_i^{obs}}$$

where the values of n , y_i^{obs} , and y_i^{for} indicate the number of observations needed to calculate prediction accuracy, together with the observed and predicted values of the target. Moreover, the $RMSE$ and MAE are used to evaluate the prediction accuracy. The $RMSE$ is expressed as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i^{obs} - y_i^{for})^2}$$

and this is how the MAE is expressed:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i^{obs} - y_i^{for}|$$

Lastly, another important statistic is the correlation coefficient (CC), represented as

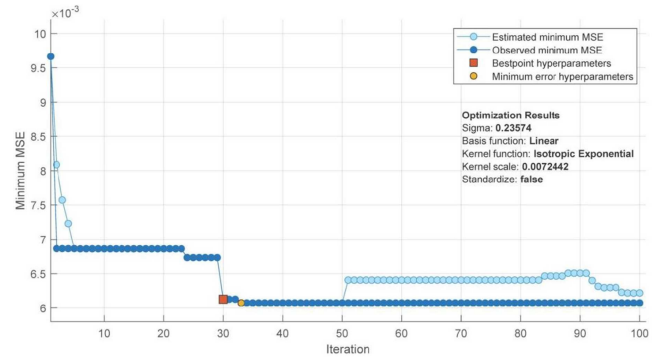
$$CC = \frac{\sum_{i=1}^n (y_i^{obs} - \overline{y_i^{obs}})(y_i^{for} - \overline{y_i^{for}})}{\sqrt{\sum_{i=1}^n (y_i^{obs} - \overline{y_i^{obs}})^2 \sum_{i=1}^n (y_i^{for} - \overline{y_i^{for}})^2}}$$

where $\overline{y_i^{obs}}$ denotes the average of y_i^{obs} and $\overline{y_i^{for}}$ signifies the average of y_i^{for} . The workflow diagram of our modeling processes is shown in Figure 2. Analysis in this work is conducted through the MATLAB platform. The model training time is close to 30 minutes on a desktop computer with 32GB RAM and i7-9700 CPU.

5. Result

The model was trained between January 02, 2014, and April 5, 2022. The projections are produced for the period starting on April 6, 2022, and ending on April 10, 2024, one day in advance. The *EIPSP* optimizations carried out during training are shown in Figure 3. After taking into account the optimization

Figure 3
Bayesian optimizations implemented through the *EIPSP* algorithm



outcomes shown in Figure 3, it was concluded that the isotropic exponential kernel expressed in Equation (1), the linear basis function expressed in Equation (13), and non-standardized predictors should be the best choice for the corn price series. 10-fold cross-validation was used to generate 10 different GPR models. The results of the parameter estimations for the models are shown in Table 2. Specifically, this table presents the parameter estimate results for each of the columns designated as “CV1” through “CV10” (where “CV” stands for “cross-validation”).

For the length of the out-of-sample evaluation period, which spans from April 6, 2022, to April 10, 2024, the 10 models offer daily price forecasts. Data collected between January 02, 2014, and April 5, 2022, were used to train these models. The final price prediction for each day of the testing period will be obtained by averaging the 10 estimations, which represents a model-averaging strategy. The consistency and reliability of future predictions are improved as a consequence of this, since it reduces the probability that any of the sub-models might possibly provide findings that are unexpected. The benefits of equal weighing systems have been explored in references [62, 63]. It is worth noting that we attempted the approach of training the model based on the whole training sample for our data, and we found that training the model using the whole training sample leads to worse out-of-sample performance as compared to using the CV framework. Figure 4 illustrates the differences in the time-series trends of the actual prices and the forecasted prices with regard to the trends. The

Figure 4
Out-of-sample forecast results from April 6, 2022, to April 10, 2024: plots of price forecasts against observations, as well as the upper and lower bounds of the 95% prediction bands

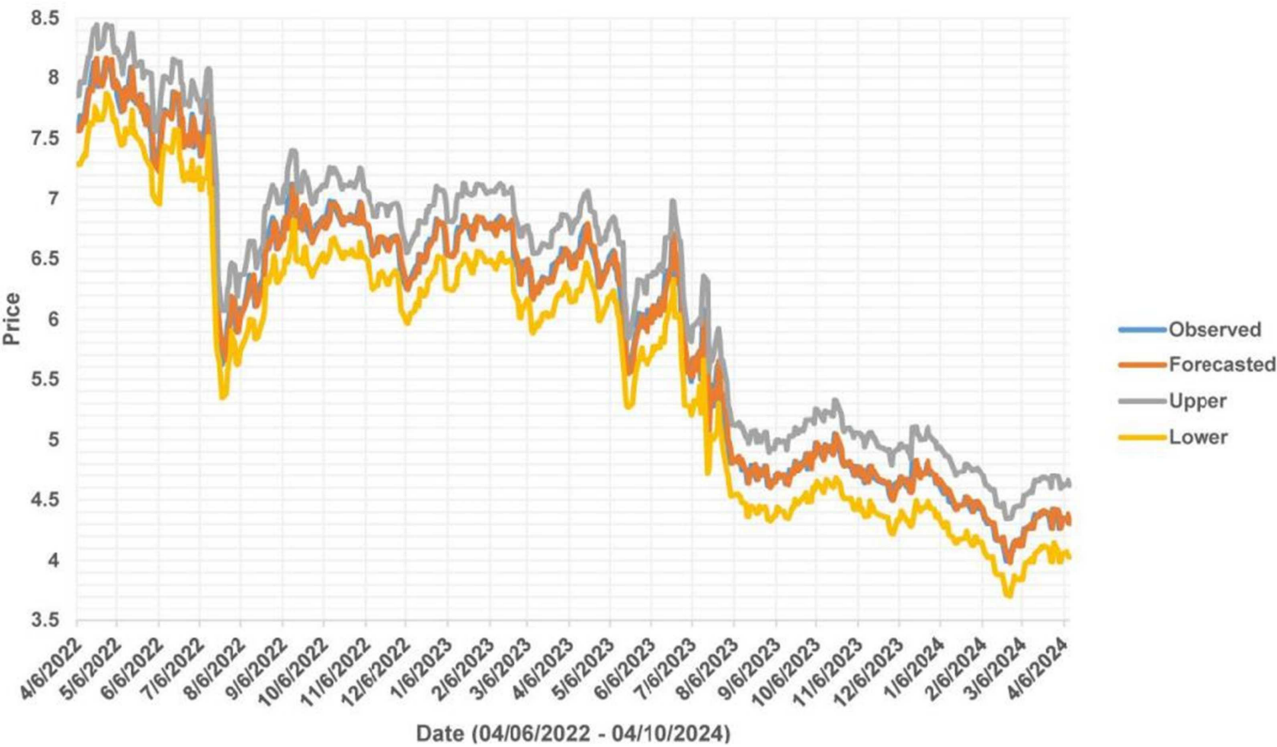
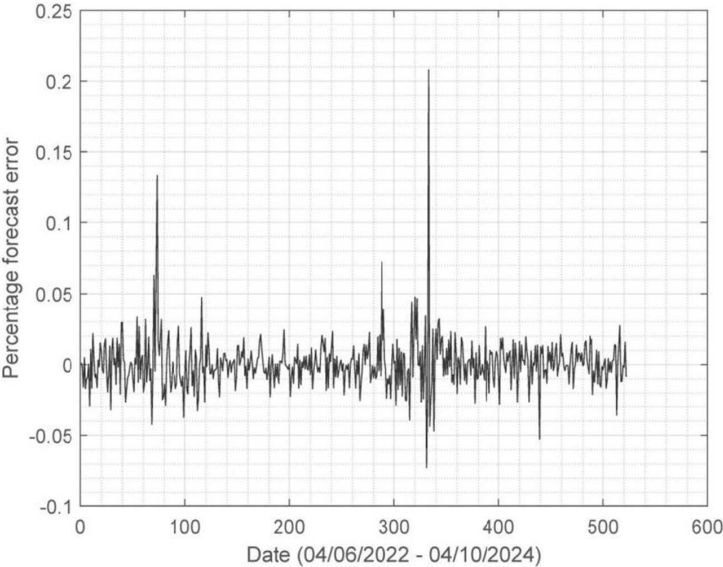


Figure 5
Out-of-sample percentage forecast errors from April 6, 2022, to April 10, 2024



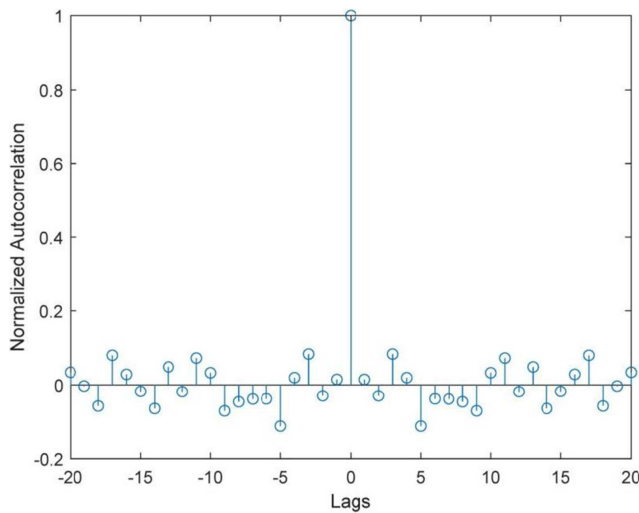
data shown in Figure 5, which illustrates the proportion of price prediction errors that are pertinent to our inquiry, lends credence to all of the conclusions that can be taken from Figure 4. The performance indicators of the estimated results are shown in Table 3. These indicators are derived from the integration of the RRMSE, RMSE, MAE, and CC. To be more accurate, Table 3 presents performance statistics as they are depicted in Figures 4 and 5. It is evident that the RRMSE of 1.8537% is achieved by the GPR models. In financial and economic forecasting contexts, potential

overfitting could be a concern. Thus, we also report the in-sample performance of the model, which shows the RRMSE of 1.6002%, RMSE of 0.0659, MAE of 0.0458, and CC of 99.753%. Therefore, comparing the in-sample and out-of-sample model performance does not reveal clear evidence of substantial overfitting for our case. Different criteria have been set by the research literature to examine the degree of accuracy that different models give [61]: When the RRMSE is less than 10%, it is thought that the projections have reached an excellent degree of accuracy; when the

Table 3
Out-of-sample forecast performance from April 6, 2022,
to April 10, 2024

RRMSE	RMSE	MAE	CC
1.8537%	0.1107	0.0703	99.528%

Figure 6
Error autocorrelation analysis: plots of normalized
autocorrelations



RRMSE is greater than 10% but smaller than 20%, it is thought that the projections have reached a good degree of accuracy; when the RRMSE is greater than 20% but smaller than 30%, it is thought that the projections have reached a fair degree of accuracy; and when the RRMSE is greater than 30%, it is thought that the projections have reached a poor degree of accuracy. Our models' excellent level of accuracy is achieved by following these grading guidelines.

The results of an error autocorrelation evaluation, which was carried out to evaluate the suitability of the generated models, are shown in Figure 6. More precisely, normalized autocorrelations for delays up to 20 are analyzed in this study. This study eliminates the possibility of major error autocorrelations and confirms the overall correctness of the models offered. It is crucial to remember, nevertheless, that empirical evidence from the literature suggests that autoregressive conditional heteroskedasticity effects might both enhance prediction accuracy and perhaps degrade the forecasting output of different time-series models.

We compare the GPR model with four alternative approaches: A naïve (no-change) model, an autoregressive (AR) model, SVR, and RT. All models use the same set of predictors,

except for the naïve model that uses the previous day's price as the forecast. Specifically, we examine the linear ϵ -insensitive SVR model, with its box constraint set to the interquartile range of the target variable divided by 1.349 and the half-width of its ϵ -insensitive band set to that same interquartile range divided by 13.49. The RT model is built on the classification analysis and regression tree algorithm, with the minimum number of observations at a branch node set to 10 and at a leaf node set to 4. The AR model is a linear time-series model and uses prices from the preceding 10 days as the predictors, which are the same as the GPR model. During the out-of-sample evaluation window from April 6, 2022, to April 10, 2024, the naïve, AR, SVR, and RT models achieve RMSE values of 0.1579, 0.3454, 0.2065, and 0.2635, respectively—each higher than the GPR model's RMSE of 0.1107. To determine whether these differences in accuracy are statistically meaningful, we apply the modified Diebold–Mariano (MDM) test [64, 65] to each pairwise comparison between GPR and the benchmark models. All resulting p -values fall below 0.01, indicating that the GPR model offers significantly superior predictive accuracy in our particular setting. Detailed numerical results regarding the comparisons are also reported in Table 4.

While our GPR models achieve strong statistical performance for corn price forecasting (1.8537% RRMSE), their economic value and interpretability are equally important. This level of accuracy has the potential to meaningfully improve procurement, inventory, and hedging for processors and inform planting, marketing, and sales timing for farmers. It is worth noting that a detailed separate assessment of economic effects, testing of trading or hedging strategies, consideration of transaction costs, and comparison of management outcomes would be necessary in future work to further evaluate the economic benefit of the forecasts. In contrast, the autoregressive, SVR, and RT models produce larger errors, with MDM tests ($p < 0.01$) confirming the GPR's superiority under the experimental conditions considered. From an interpretability standpoint, the GPR's probabilistic structure avoids black-box opacity. The selected isotropic exponential kernel with a linear basis function reveals a smooth, persistent temporal dependence, and the 10-day lag structure implies that roughly two weeks of price history optimally captures relevant information—suggesting market participants could focus on recent developments rather than longer, noisier histories under the GPR framework. Bayesian optimization yields parameter estimates that describe how past prices influence predictions, while

Table 4
Benchmark analysis: comparisons of the GPR model with the SVR, RT, AR, and Naïve (no-change) models during the
out-of-sample period from April 6, 2022, to April 10, 2024

Model	MAE	CC (%)	RMSE	RRMSE (%)	MDM (test statistic)	MDM test (p -value)
GPR	0.0703	99.528	0.1107	1.8537	—	—
SVR	0.3056	97.815	0.2065	3.4579	2.8062	0.0026
RT	0.3645	95.362	0.2635	4.4124	2.9074	0.0019
AR	0.5132	95.098	0.3454	5.7838	3.1726	0.0008
Naïve (no-change)	0.1981	98.963	0.1579	2.6441	2.5649	0.0053

CV model averaging confirms robustness. We note that while our results support the forecasting suitability of the selected model configuration, they do not, by themselves, prove the existence of specific market mechanisms such as momentum, mean reversion, or a particular rate at which information is absorbed by the market. Instead, such economic explanations could serve as possible hypotheses to be tested in future work. For policymakers, accurate and transparent forecasts enable proactive interventions in supply disruptions, price spikes, and programs like strategic reserves, with the model's clarity easing stakeholder communication. Financial market participants gain a tool for informed trading and risk management, and modest training time supports practical adoption. Forecast error analysis did not reveal evidence of substantial systematic bias within the evaluation period, which could be a crucial consideration for sound decision-making.

6. Conclusion

Previous research has shown that policymakers and decision-makers benefit from having a thorough grasp of commodity price projections. The same idea holds true for figuring out the price of corn. This study uses daily data from January 2, 2014, to April 10, 2024, to analyze the topic of corn price estimations. In this case, the model parameters are established using the CV approach and Bayesian optimization procedures, while the predictions are achieved via the implementation of GPRs. The RRMSE calculated for the out-of-sample testing period spanning April 6, 2022, to April 10, 2024, was 1.8537%. Technical prediction data could be employed alone or in combination with other findings for policy research involving pricing trends. The methods we used to evaluate these concerns may be useful in predicting challenges of a similar kind for other sectors of the economy. Because of its simplicity and versatility, this empirical framework may be used in a wide range of decision-making processes. While the GPR models developed in this study demonstrate strong predictive accuracy, we acknowledge the important observation regarding the limitation of relying on univariate lagged prices. The use of corn's own price history, though methodologically sound for establishing a baseline forecasting capability, does not capture the full range of economic forces that shape agricultural commodity markets. Future research should therefore extend the current framework by incorporating a broader set of explanatory variables that reflect the multifaceted determinants of corn prices. Specifically, macroeconomic indicators such as inflation rates, exchange rates, and interest rates could be integrated as additional predictors, given their potential influence on commodity price dynamics. Energy prices, particularly crude oil and natural gas, represent another critical dimension, as they could affect production costs through fertilizers, fuel, and transportation expenses. Weather variables, including temperature anomalies, precipitation patterns, and drought indices, constitute a third essential category, given the fundamental role of growing conditions in determining supply. Additionally, international market factors such as global stock levels, export–import flows, and price movements in related commodities (soybeans, wheat, and rice) could possibly help enhance the model's explanatory power. The GPR framework is well-suited to accommodate such expansions, as it can flexibly handle high-dimensional predictor spaces through its kernel-based approach, while maintaining its interpretation. This extension might not only improve forecast accuracy but would also possibly enhance the economic interpretability of the model, allowing policymakers and market participants to understand which external factors most strongly influence price movements.

Such multidimensional GPR models would help transform the forecasting exercise from a statistical endeavor into a more comprehensive analytical tool that bridges empirical modeling with economic theory, thereby further advancing the practical utility of commodity price forecasts for strategic decision-making. Future research could prioritize enhancing the interpretability of ML models in this domain. While the GPR framework offers inherent transparency, applying post hoc interpretation techniques might quantify the contribution of each lagged price to the forecast. This could help provide market participants with additional actionable insights into which historical observations most strongly influence predictions.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

Xiaojie Xu is the Editorial Board Member for *FinTech and Sustainable Innovation* and was not involved in the editorial review or the decision to publish this article. The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Bingzi Jin: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization.
Xiaojie Xu: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization.

References

- [1] Beltran, A. O., Suaza-Medina, M. E., Zarazaga-Soria, F. J., De-La-Hoz-Franco, E., & Escorcia-Gutierrez, J. (2025). Multivariate integration of time series with ML for corn price forecasting in Colombia. *Expert Systems with Applications*, 299, 129822. <https://doi.org/10.1016/j.eswa.2025.129822>
- [2] Theofilou, A., Nastis, S. A., Michailidis, A., Bourmaris, T., & Mattas, K. (2025). Predicting prices of staple crops using machine learning: A systematic review of studies on wheat, corn, and rice. *Sustainability*, 17(12), 5456. <https://doi.org/10.3390/su17125456>
- [3] Cheng, W., & Li, F. (2025). Analysis of corn price forecast in China based on Lasso-XGBoost-SHAP. *Cogent Food & Agriculture*, 11(1), 2546076. <https://doi.org/10.1080/23311932.2025.2546076>
- [4] Khadka, R., Sanwa, R. L., Khadka, N., & Chi, Y. N. (2025). Forecasting the global price of corn using neural network and hybrid approach. *IEEE Access*, 13, 167424–167438. <https://doi.org/10.1109/ACCESS.2025.3611320>
- [5] Wang, L., An, W., & Li, F.-T. (2024). Text-based corn futures price forecasting using improved neural basis expansion network. *Journal of Forecasting*, 43(6), 2042–2063. <https://doi.org/10.1002/for.3119>

- [6] Guo, Y., Tang, D., Tang, W., Yang, S., Tang, Q., Feng, Y., & Zhang, F. (2022). Agricultural price prediction based on combined forecasting model under spatial-temporal influencing factors. *Sustainability*, 14(17), 10483. <https://doi.org/10.3390/su141710483>
- [7] Zeng, L., Ling, L., Zhang, D., & Jiang, W. (2023). Optimal forecast combination based on PSO-CS approach for daily agricultural future prices forecasting. *Applied Soft Computing*, 132, 109833. <https://doi.org/10.1016/j.asoc.2022.109833>
- [8] Wang, J., Wang, Z., Li, X., & Zhou, H. (2022). Artificial bee colony-based combination approach to forecasting agricultural commodity prices. *International Journal of Forecasting*, 38(1), 21–34. <https://doi.org/10.1016/j.ijforecast.2019.08.006>
- [9] Yang, C., Zhai, Y., & Liu, Z. (2025). Enhancing corn industry sustainability through deep learning hybrid models for price volatility forecasting. *PLOS One*, 20(6), e0323714. <https://doi.org/10.1371/journal.pone.0323714>
- [10] Kaabi, J., Harrath, Y., & Price, E. (2025). Utilizing ensemble learning techniques to enhance corn price prediction: A case study on South Dakota. *Agronomy*, 15(11), 2447. <https://doi.org/10.3390/agronomy15112447>
- [11] Zomchak, L., & Kukhotska, T. (2025). Building food security resilience in Ukraine: The autoregressive approach to food price forecasting. In P. Štarchoň, S. Fedushko, & K. Gubiniova (Eds.), *Developments in information and knowledge management systems for business applications: Volume 8* (pp. 403–424). Springer. https://doi.org/10.1007/978-3-031-80935-4_19
- [12] Peter, O. A., & Kwanashie, M. (2025). Impact of cereal production on food inflation in Nigeria: A time series analysis. *African Journal of Agricultural Science and Food Research*, 18(1), 167–201. <https://doi.org/10.62154/ajasfr.2025.018.010628>
- [13] Derouez, F., & Alqattan, Y. S. (2025). Agricultural insurance and food security in Saudi Arabia: Exploring short and long-run dynamics using ARDL approach and VECM technique. *Sustainability*, 17(10), 4696. <https://doi.org/10.3390/su17104696>
- [14] Kamalanathan, R., & Abdullah, A. A. S. (2025). Determining the impact of rainfall on major crop yields in Tamil Nadu using vector error correction model. *Current Agriculture Research Journal*, 13(1), 78–87. <http://dx.doi.org/10.12944/CARJ.13.1.08>
- [15] Alade, I. O., Oyedele, M. O., Rahman, M. A. A., & Saleh, T. A. (2022). Prediction of the lattice constants of pyrochlore compounds using machine learning. *Soft Computing*, 26(17), 8307–8315. <https://doi.org/10.1007/s00500-022-07218-1>
- [16] Zhao, T., Chen, G., Suraphee, S., Phoophiwfa, T., & Busabodhin, P. (2025). A hybrid TCN-XGBoost model for agricultural product market price forecasting. *PLOS One*, 20(5), e0322496. <https://doi.org/10.1371/journal.pone.0322496>
- [17] Dinçer, H., Yüksel, S., An, J., & Mikhaylov, A. (2024). Quantum and AI-based uncertainties for impact-relation map of multidimensional NFT investment decisions. *Finance Research Letters*, 66, 105723. <https://doi.org/10.1016/j.frl.2024.105723>
- [18] Hyndman, R. J., & Athanasopoulos, G. (2014). *Forecasting: Principles and practice* (2nd ed ed.). Australia: OTexts.
- [19] Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J. (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association*, 114(526), 804–819. <https://doi.org/10.1080/01621459.2018.1448825>
- [20] Mahto, A. K., Alam, M. A., Biswas, R., Ahmed, J., & Alam, S. I. (2021). Short-term forecasting of agriculture commodities in context of Indian market for sustainable agriculture by using the artificial neural network. *Journal of Food Quality*, 2021(1), 9939906. <https://doi.org/10.1155/2021/9939906>
- [21] Vitale, J., & Robinson, J. (2025). In-season price forecasting in cotton futures markets using ARIMA, neural network, and LSTM machine learning models. *Journal of Risk and Financial Management*, 18(2), 93. <https://doi.org/10.3390/jrfm18020093>
- [22] Sanwa, R., Khadka, Limbu., R., & Chi, Y. N. (2025). Forecasting global monthly cotton prices: The superiority of NNAR models over traditional models. *Frontiers in Artificial Intelligence*, 8, 1628744. <https://doi.org/10.3389/frai.2025.1628744>
- [23] Wen, G., Ma, B.-L., Vanasse, A., Caldwell, C. D., Earl, H. J., & Smith, D. L. (2021). Machine learning-based canola yield prediction for site-specific nitrogen recommendations. *Nutrient Cycling in Agroecosystems*, 121(2), 241–256. <https://doi.org/10.1007/s10705-021-10170-5>
- [24] Harris, J. (2017). *A machine learning approach to forecasting consumer food prices*. Master's Thesis, Dalhousie University.
- [25] Srichaiyan, P., Tippayawong, K. Y., & Boonprasope, A. (2025). Forecasting soybean futures prices with adaptive AI models. *IEEE Access*, 13, 48239–48256. <https://doi.org/10.1109/ACCESS.2025.3546786>
- [26] Deina, C., do Amaral Prates, M. H., Alves, C. H. R., Martins, M. S. R., Trojan, F., Stevan Jr, S. L., & Siqueira, H. V. (2022). A methodology for coffee price forecasting based on extreme learning machines. *Information Processing in Agriculture*, 9(4), 556–565. <https://doi.org/10.1016/j.inpa.2021.07.003>
- [27] Silva, N., Siqueira, I., Okida, S., Stevan Jr, S. L., & Siqueira, H. (2019). Neural networks for predicting prices of sugarcane derivatives. *Sugar Tech*, 21(3), 514–523. <https://doi.org/10.1007/s12355-018-0648-5>
- [28] Lu, Q., Sun, S., Duan, H., & Wang, S. (2021). Analysis and forecasting of crude oil price based on the variable selection-LSTM integrated model. *Energy Informatics*, 4(2), 47. <https://doi.org/10.1186/s42162-021-00166-4>
- [29] Abedinia, O., Amjadi, N., Shafie-khah, M., & Catalão, J. P. (2015). Electricity price forecast using combinatorial neural network trained by a new stochastic search method. *Energy Conversion and Management*, 105, 642–654. <https://doi.org/10.1016/j.enconman.2015.08.025>
- [30] Sohrabi, P., Jodeiri Shokri, B., & Dehghani, H. (2023). Predicting coal price using time series methods and combination of radial basis function (RBF) neural network with time series: Predicting coal price using time series methods and combination of radial basis function (RBF) neural network with time series. *Mineral Economics*, 36(2), 207–216. <https://doi.org/10.1007/s13563-021-00286-z>
- [31] Huang, Y., Dai, X., Wang, Q., & Zhou, D. (2021). A hybrid model for carbon price forecasting using GARCH and long short-term memory network. *Applied Energy*, 285, 116485. <https://doi.org/10.1016/j.apenergy.2021.116485>
- [32] Huynh, T. L. D. (2020). The effect of uncertainty on the precious metals market: New insights from transfer entropy and neural network VAR. *Resources Policy*, 66, 101623. <https://doi.org/10.1016/j.resourpol.2020.101623>
- [33] Zhou, J., & Xu, Z. (2023). A novel three-stage hybrid learning paradigm based on a multi-decomposition strategy, optimized relevance vector machine, and error correction for multi-step

- forecasting of precious metal prices. *Resources Policy*, 80, 103148. <https://doi.org/10.1016/j.resourpol.2022.103148>
- [34] Sadorsky, P. (2021). Predicting gold and silver price direction using tree-based classifiers. *Journal of Risk and Financial Management*, 14(5), 198. <https://doi.org/10.3390/jrfm14050198>
- [35] Gligorić, Z., Gligorić, M., Halilović, D., Beljić, Č., & Urošević, K. (2020). Hybrid stochastic-grey model to forecast the behavior of metal price in the mining industry. *Sustainability*, 12(16), 6533. <https://doi.org/10.3390/su12166533>
- [36] Li, Z., Yang, Y., Chen, Y., & Huang, J. (2023). A novel non-ferrous metals price forecast model based on LSTM and multivariate mode decomposition. *Axioms*, 12(7), 670. <https://doi.org/10.3390/axioms12070670>
- [37] Aruna, S., Umamaheswari, P., Sujipriya, J., & Umamaheswari, R. (2021). Prediction of potential gold prices using machine learning approach. *Annals of the Romanian Society for Cell Biology*, 25(4), 1385–1396.
- [38] Neal, R. M. (1996). *Bayesian learning for neural networks*. USA: Springer. <https://doi.org/10.1007/978-1-4612-0745-0>
- [39] Neal, R. M. (1997). Monte Carlo implementation of Gaussian process models for Bayesian regression and classification. *arXiv Preprint: physics/9701026*, .
- [40] Brahim-Belhouari, S., & Vesin, J. M. (2001). Bayesian learning using Gaussian process for time series prediction. In *Proceedings of the 11th IEEE Signal Processing Workshop on Statistical Signal Processing*, 433–436. <https://doi.org/10.1109/SSP.2001.955315>
- [41] Brahim-Belhouari, S., & Bermak, A. (2004). Gaussian process for nonstationary time series prediction. *Computational Statistics & Data Analysis*, 47(4), 705–712. <https://doi.org/10.1016/j.csda.2004.02.006>
- [42] Lahmiri, S., Bekiros, S., & Avdoulas, C. (2023). A comparative assessment of machine learning methods for predicting housing prices using Bayesian optimization. *Decision Analytics Journal*, 6, 100166. <https://doi.org/10.1016/j.dajour.2023.100166>
- [43] Brock, W. A., Scheinkman, J. A., Dechert, W. D., & LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, 15(3), 197–235. <https://doi.org/10.1080/07474939608800353>
- [44] Xiong, T., Li, C., & Bao, Y. (2018). Seasonal forecasting of agricultural commodity price using a hybrid STL and ELM method: Evidence from the vegetable market in China. *Neurocomputing*, 275, 2831–2844. <https://doi.org/10.1016/j.neucom.2017.11.053>
- [45] Lv, J., & Wu, X. (2022). Corn price prediction in China's futures market during COVID-19. *Academic Journal of Business Management*, 4(11), 7–12. <https://doi.org/10.25236/AJBM.2022.041102>
- [46] Satapathy, A., Mishra, A. K., Mishra, S. N., Mishra, R. K., & Sarangi, K. K. (2021). Price forecasting of maize in Nabarangur district of Odisha. *Environment and Ecology*, 39(4A), 1373–1381.
- [47] Pierre, G., & Kaminski, J. (2019). Cross country maize market linkages in Africa: Integration and price transmission across local and global markets. *Agricultural Economics*, 50(1), 79–90. <https://doi.org/10.1111/agec.12467>
- [48] de Albuquerque, V. P., de Medeiros, R. K., de Jesus, D. P., & de Oliveira, F. A. (2022). The role of transition regime models for corn prices forecasting. *Revista de Economia e Sociologia Rural*, 60(2), e236922. <https://doi.org/10.1590/1806-9479.2021.236922>
- [49] Guo, J., & Tanaka, T. (2020). Dynamic transmissions and volatility spillovers between global price and US producer price in agricultural markets. *Journal of Risk and Financial Management*, 13(4), 83. <https://doi.org/10.3390/jrfm13040083>
- [50] Jyothi Shiva Kumar, N. M., & Kotreshwar, G. (2017). Price discovery dynamics of Indian maize futures market. *Asian Journal of Research in Business Economics and Management*, 7(5), 72–85.
- [51] Sun, F., Meng, X., Zhang, Y., Wang, Y., Jiang, H., & Liu, P. (2023). Agricultural product price forecasting methods: A review. *Agriculture*, 13(9), 1671. <https://doi.org/10.3390/agriculture13091671>
- [52] Singh, N., & Sindhu, R. (2024). Crop price prediction using machine learning. *Journal of Electrical Systems*, 20, 2258–2269.
- [53] Mayabi, T. W. (2019). *An artificial neural network model for predicting retail maize prices in Kenya*. Master's Thesis, University of Nairobi.
- [54] Teste, F., Gangloff, H., Chen, M., Ciais, P., & Makowski, D. (2024). Leveraging satellite data with machine and deep learning techniques for corn yield and price forecasting. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 4413716. <https://doi.org/10.1109/TGRS.2024.3448205>
- [55] Jain, A., Marvaniya, S., Godbole, S., & Munigala, V. (2020). A framework for crop price forecasting in emerging economies by analyzing the quality of time-series data. *arXiv Preprint: 2009.04171*, .
- [56] Sanusi, O. I., Safi, S. K., Adeeko, O., & Tabash, M. I. (2022). Forecasting agricultural commodity price using different models: A case study of widely consumed grains in Nigeria. *Agricultural and Resource Economics: International Scientific E-Journal*, 8(2), 124–140. <https://doi.org/10.51599/10.51599/are.2022.08.02.07>
- [57] Choudhary, K., Jha, G. K., Jaiswal, R., Venkatesh, P., & Parsad, R. (2023). Agricultural price forecasting based on variational mode decomposition and time-delay neural network. *Statistics and Applications*, 21, 235–257.
- [58] Wu, B., Wang, Z., & Wang, L. (2024). Interpretable corn future price forecasting with multivariate time series. *Journal of Forecasting*, 43(5), 1575–1594. <https://doi.org/10.1002/for.3099>
- [59] Murugesan, G., & Radha, B. (2023). An extrapolative model for price prediction of crops using hybrid ensemble learning techniques. *International Journal of Advanced Technology and Engineering Exploration*, 10(98), 1–20. <https://doi.org/10.19101/IJATEE.2021.876382>
- [60] Bull, A. D. (2011). Convergence rates of efficient global optimization algorithms. *Journal of Machine Learning Research*, 12(10), 2879–2904.
- [61] Despotovic, M., Nedic, V., Despotovic, D., & Cvetanovic, S. (2016). Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation. *Renewable and Sustainable Energy Reviews*, 56, 246–260. <https://doi.org/10.1016/j.rser.2015.11.058>
- [62] Capistrán, C., & Timmermann, A. (2009). Forecast combination with entry and exit of experts. *Journal of Business & Economic Statistics*, 27(4), 428–440. <https://doi.org/10.1198/jbes.2009.07211>

- [63] Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger, & A. G. T. Timmermann (Eds.), *Handbook of economic forecasting* (pp. 135–196). North Holland.
- [64] Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of forecasting*, 13(2), 281–291. [https://doi.org/10.1016/S0169-2070\(96\)00719-4](https://doi.org/10.1016/S0169-2070(96)00719-4)
- [65] Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 20(1), 134–144. <https://doi.org/10.1198/073500102753410444>

How to Cite: Jin, B., & Xu, X. (2026). Machine Learning Corn Price Predictions. *FinTech and Sustainable Innovation*. <https://doi.org/10.47852/bonviewFSI620210385>

Appendix

Abbreviation or Symbol	Meaning
Abbreviation	
VAR	Vector Autoregressive
VECM	Vector Error Correction Model
ARIMA	Autoregressive Integrated Moving Average
SARIMA	Seasonal Autoregressive Integrated Moving Average
AR	Autoregressive
AI	Artificial Intelligence
ML	Machine Learning
GPR	Gaussian Process Regression
SVR	Support Vector Regression
RT	Regression Tree
BDS	Brock–Dechert–Scheinkman
BEKK-GARCH	Generalized Autoregressive Conditional Heteroskedasticity-Baba, Engle, Kraft, and Kroner
DCC-GARCH	Generalized Autoregressive Conditional Heteroskedasticity-Dynamic Conditional Correlation
LSTM	Long Short-Term Memory
EIPSP	Expected Improvement per Second Plus
CV	Cross-Validation
RMSE	Root Mean Square Error
RRMSE	Relative Root Mean Square Error
MAE	Mean Absolute Error
CC	Correlation Coefficient
MDM	Modified Diebold–Mariano
Symbol	
k	Kernel
b	Basis Function
α	Scale-Mixture Parameter
σ_l	Characteristic Length Scale
σ_f	Standard Deviation of the Signal
$a(\cdot)$	Acquisition Function