ESCUELA NAVAL DE CADETES
"ALMIRANTE PADILLA"
Honor y Tradición

Article

Theoretical Perspective of the Hybrid EMD–SSA–VMD–EWT Approach and Machine Learning in Price Prediction

Perspectiva Teórica del Enfoque Híbrido EMD–SSA–VMD–EWT y Machine Learning en la Predicción de Precios

Adelaida Ojeda-Beltran [1]*

[1] School of Economics, Universidad del Atlántico, Barranquilla, 081007, Colombia; adelaidaojeda@mail.uniatlantico.edu.co
* Correspondence: adelaidaojeda@mail.uniatlantico.edu.co

Abstract: The prediction of maize prices in Colombia has become a challenge due to the high volatility that characterizes agricultural markets and the complex interaction among various endogenous and exogenous factors. This article aims to provide a comprehensive theoretical foundation that identifies the conceptual pillars for developing a hybrid prediction model based on advanced time series decomposition techniques, machine learning algorithms, and optimization metaheuristics. First, agricultural processes and the key stages of the maize value chain are examined, highlighting the influence of post-harvest activities, logistics, and marketing systems on price formation. Subsequently, contemporary decomposition methodsEMD, SSA, VMD, and EWTare reviewed as tools capable of extracting structure, reducing noise, and capturing hidden patterns in nonlinear and non-stationary signals. Third, the contributions of supervised machine learning are synthesized, with emphasis on models such as XGBoost, LightGBM, and neural networks (FCN and RNN), widely used in complex predictive scenarios. Finally, optimization metaheuristics, particularly Particle Swarm Optimization (PSO) and Cuckoo Search (CS), are examined, highlighting their ability to fine-tune parameters and enhance the predictive performance of hybrid models. The articulation of these conceptual pillars provides a robust framework that supports the design of more accurate predictive architectures adapted to the dynamics of the Colombian agricultural market.

Keywords: Theoretical perspectives; Machine Learning; Decomposition methods; Metaheuristic optimization techniques

Resumen: La predicción del precio del maíz en Colombia se ha convertido en un desafío debido a la alta volatilidad que caracteriza los mercados agrícolas y a la compleja interacción entre distintos factores endógenos y exógenos. Este artículo tiene como objetivo realizar una fundamentación teórica integral que identificando los ejes conceptuales para el desarrollo de un modelo híbrido de predicción basado en técnicas avanzadas de descomposición de series temporales y algoritmos de aprendizaje automático y metaheurísticas de optimización. En primer lugar, se analizan los procesos agrícolas y los eslabones de la cadena productiva del maíz, resaltando la influencia de las actividades de postcosecha, la logística y los sistemas de comercialización sobre la formación del precio. Posteriormente, se revisan los métodos contemporáneos de descomposición EMD, SSA, VMD y EWT entendidos como herramientas capaces de extraer estructura, reducir ruido y capturar patrones ocultos en señales no lineales y no estacionarias. En tercer lugar, se sintetizan los aportes del aprendizaje automático supervisado, con énfasis en modelos como XGBoost, LightGBM y redes neuronales (FCN y RNN), ampliamente utilizados en escenarios de predicción compleja. Finalmente,

Citation: Ojeda, A. Theoretical Perspective of the Hybrid EMD–SSA–VMD–EWT Approach and Machine Learning in Price Prediction. *OnBoard Knowledge Journal* **2025**, *1*, 0. https://doi.org/

Received: 05/05/2025
Accepted: 10/06/2025
Published: 27/06/2025

se examinan las metaheurísticas de optimización, particularmente Particle Swarm Optimization (PSO) y Cuckoo Search (CS), destacando su capacidad para ajustar parámetros y mejorar el rendimiento predictivo de modelos híbridos. La articulación de estos ejes conceptuales configura un marco robusto que respalda el diseño de arquitecturas predictivas más precisas y adaptadas a la dinámica del mercado agrícola colombiano.

Palabras clave: Perspectivas teóricas; Machine Learning; Métodos de descomposición; Metaheurísticas de optimización

1. Introdution

Agricultural production processes consist of a series of activities aimed at modifying a natural ecosystem for the production of food and inputs. These processes are developed in three phases: land preparation, harvesting, and post-harvest [24]. From the perspective of the agricultural or agri-food system (or value chain), harvesting can be considered a link or transitional element, or even a peak that separates two slopes: the pre-harvest phase, corresponding to the production activity itself, and the post-harvest phase, which extends from harvesting operations to consumption.

The post-harvest system comprises a number of sequential activities and functions that can be classified into two categories: technical activities and economic activities [23]. Technical activities include harvesting, field drying, threshing, cleaning, drying, storage, and primary processing; while economic activities include transportation, quality control, packaging, secondary processing, and marketing, as illustrated in Figure 1.

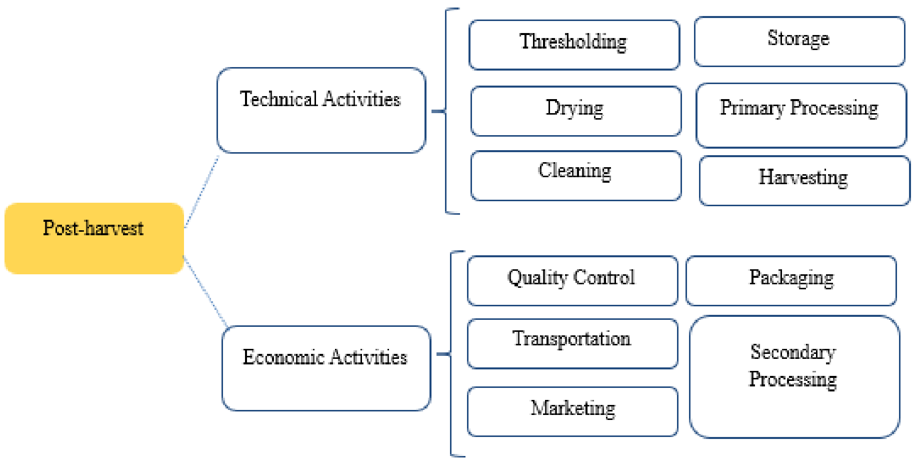


Figure 1. Post-harvest activities.
Note: Source [23]

In this way, agricultural production processes constitute the foundation upon which efficiency, sustainability, and quality of goods obtained in the rural sector are ensured. However, value generation in agriculture is not limited solely to the stages of planting, cultivation, and harvesting, but also extends to the way in which these products reach the final consumer [57]. At this point, marketing activities become fundamental, as they determine distribution dynamics, market competitiveness, and producers' access to fair and sustainable trading conditions.

1.1. Agricultural Marketing System

Marketing constitutes the final and decisive phase of the post-harvest system and is closely linked to transportation, since production is of little use if goods do not reach consumers at the right place and time. Its essential purpose is to move products from farms or harvesting sites to points of demand, ensuring that they meet attributes such as variety, degree of maturity, size, packaging, origin, and food safety, in accordance with current sanitary regulations [54].

For a marketing system to be transparent and profitable, it must reduce information asymmetries among stakeholders through clear regulations on weights, measures, labeling, and sanitary conditions, as well as through timely access to data on supply, demand, imports, consumer preferences, and logistics. In this regard, Information and Communication Technologies (ICT) and artificial intelligence become key tools, as they enable the analysis and anticipation of the behavior of these variables, providing more accurate and timely forecasts of agricultural market dynamics particularly product prices. For this reason, marketing processes acquire great importance within a productive chain [17].

1.2. Agricultural Productive Chain

According to [52], a productive chain is a linkage of multiple stages through the production of differentiated goods and services among firms; these stages include everything from inputs and elements of the production process to the final consumer or another form of the productive process. Additionally, as stated in [35], to understand the concept of productive chains it is necessary to consider the actors involved in the economic system who, in the long term, contribute to generating competitive advantages within the business environment.

According to [12], the productive chain emerges as a concept linked to the school of strategic planning, which allows competitiveness to be analyzed based on the internal characteristics of organizations and external factors related to their environment. In this sense, interactions with external agents such as suppliers, countries, customers, or distributors foster the creation of incentives and synergies that strengthen competitive advantage.

The term links, in turn, was initially introduced by [31] in their studies and works, constituting a key analytical category for understanding the dynamics of interdependence among the different stages of production.

1.3. Maize Productive Chain in Colombia

The maize productive chain in Colombia is configured as a strategic system of coordination among public and private actors, aimed at improving the competitiveness, productivity, and sustainability of this cereal in the country. According to [15], these chain organizations are established as advisory bodies to the National Government, responsible for negotiating policies, coordinating actions, and promoting sectoral development strategies. Colombia annually consumes 8.4 million tons of maize, of which 88% corresponds to yellow maize and 12% to white maize, while domestic production barely reaches 1.9 million tons cultivated on 462 thousand hectares. This situation forces the country to import approximately 83% of yellow maize and 36% of white maize. In addition, the country is a price taker in international markets, using the Chicago Board of Trade as a reference, which exposes the domestic market to global volatility.

At the regional level, maize cultivation is concentrated in departments such as Valle del Cauca, Meta, Tolima, and Córdoba, which constitute highly technified productive hubs. In these regions, yields exceed the national average, with Valle del Cauca standing out at 6.8 tons/ha and Meta at 7.2 tons/ha for white maize. In other regions, such as the department of Atlántico, there are dispersed plantings and hubs with complementary potential; however, insufficient supporting goods and services hinder the consolidation of industrial and commercial linkages, limiting the full utilization of their agricultural capacity [14]. Regarding the links of the chain, it is structured into five levels referenced in Figure 3: primary production; provision of inputs and services; industrial processing; wholesale and retail marketing; and finally, final consumption. These links are articulated through the National Council of Colombian Maize, the highest decision-making body, where producers, processors, marketers, the animal feed and human food industries, as well as public entities and input suppliers are represented [5]. [46], as shown in Figure 2.

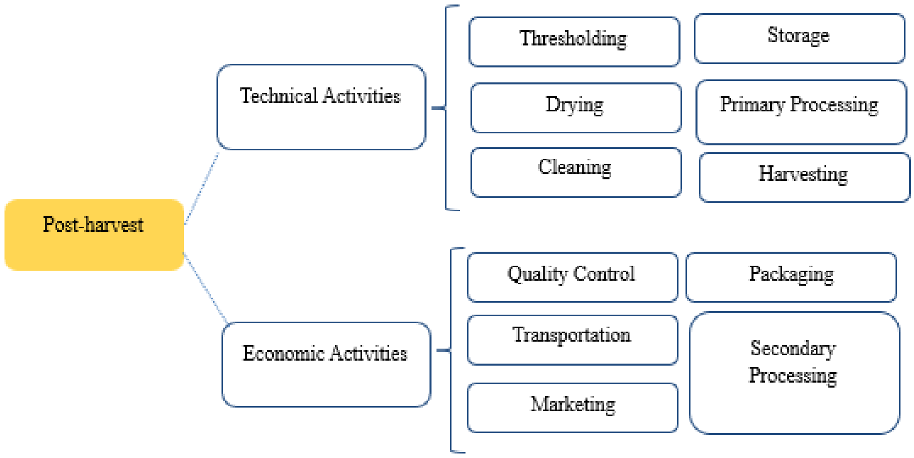


Figure 2. Agricultural Maize Value Chain in Colombia.
Note: Source [23]

1.4. Price Determination and Dynamics in Agricultural Markets

Supply and demand are the two fundamental components of any market. The quantities of a good that consumers wish and are able to purchase are referred to as the demand for that good. To demand means to be willing to buy, whereas to buy means to actually carry out the purchase. Demand reflects an intention, while purchase constitutes an action [11].

1.5. Price

Price is the number of monetary units (dollars, pesos, euros, etc.) that must be given in exchange for a good or service. Economists refer to this as the monetary price [51].

2. Artificial Intelligence

Artificial intelligence (AI) is a field of study that formally emerged in 1956 during the Dartmouth Conference, where the possibility was raised that machines could imitate reasoning processes similar to those of humans [8]. Since then, its purpose has been the handling, processing, and analysis of data in order to develop systems capable of performing tasks that traditionally require human intelligence. Among its most notable applications are robotics, expert systems for decision-making, image and text recognition and processing, as well as the development of autonomous agents in various productive sectors [47]. AI can be generally understood as the ability of a machine to simulate human cognitive functions such as learning, perception, reasoning, and problem solving [49].

Within this broad field lies machine learning (ML), which constitutes one of its most dynamic and rapidly developing branches in recent decades. Unlike AI in a general sense, ML focuses on the construction of algorithms capable of learning patterns directly from data.

2.1. Machine Learning

Among the existing Artificial Intelligence techniques, an important category known as Machine Learning (ML) corresponds to a set of algorithms capable of performing complex tasks by identifying non-trivial relationships within data for descriptive or predictive purposes [7].

Machine Learning techniques can be defined as a set of methods capable of automatically detecting patterns in data [45]. This concept of Machine Learning includes the use of detected patterns to make predictions or to support other types of decision-making under certain levels of uncertainty, which, according to [34], reduces the human effort required to apply learning.

Machine Learning techniques are applied within the context of data mining, which is the process of the "non-trivial extraction of implicit, previously unknown, and potentially useful knowledge from data" [25]. According to [41] and [16], data mining refers to the process of "automatic discovery of interesting and non-obvious patterns or models hidden in a database, which have great potential to contribute to key business aspects." Data mining is "a data exploitation mechanism consisting of the search for valuable information in large volumes of data," according to [21].

There are two types of machine learning algorithms: supervised and unsupervised. Supervised learning is used when there is knowledge of the desired outputs, and a training process is carried out to obtain those outputs. On the other hand, when information about expected outputs is not available, clustering techniques that do not require supervision are typically applied [44]. The former are primarily used to propose prediction-based solutions, while the latter usually focus on description.

2.1.1. Predictive Machine Learning

Predictive Machine Learning refers to the application of algorithms capable of analyzing historical and current data to generate predictions about future events. Its main objective is to identify patterns in data and use them to anticipate behaviors or outcomes [53]. This approach is widely applied in business, healthcare, agriculture, finance, and education, as it enables improved evidence-based decision-making [58].

2.1.2. Predictive Machine Learning – Classification Type

Predictive classification-based machine learning is used when the target variable is categorical in nature, that is, when possible outcomes are grouped into discrete classes or categories. Its main objective is to assign a label or class to each observation based on the input features or attributes that describe it. To achieve this, classification algorithms learn patterns and relationships from a labeled dataset, allowing them to generalize their knowledge and correctly predict new observations. Among the most commonly used models are logistic regression, decision trees, random forests, support vector machines (SVM), and neural networks [27].

2.1.3. Predictive Machine Learning – Regression Type

Predictive regression-based machine learning is applied when the target variable is numerical and continuous. Its purpose is to estimate future values based on past data [8]. This approach makes it possible to quantify relationships among variables and generate projections that help anticipate behavior in real-world contexts. It is widely used to predict product demand, financial asset prices, environmental pollution levels, or agricultural crop yields, among other scenarios in which values vary over time or as a function of external factors. Commonly used algorithms include Linear Regression, Ridge Regression, Lasso, Regression Trees, Gradient Boosting, and Deep Neural Networks.

2.1.4. Fully Connected Neural Networks (FCN)

Fully Connected Networks (FCNs), also referred to as dense networks, constitute one of the most representative architectures in deep learning. Their structure is based on each neuron in one layer being connected to all neurons in the subsequent layer, which ensures a complete flow of information [61]. This characteristic enables them to model highly complex relationships among input variables, which is why they are frequently used in classification, regression, and pattern recognition tasks [39]. From a computational perspective, each layer performs a linear transformation of the input values through a weight matrix and a bias vector, followed by the application of a nonlinear activation function [56]. This process endows the model with the ability to learn nonlinear representations and approximate complex functions.

2.1.5. Recurrent Neural Networks (RNN)

Recurrent Neural Networks (RNNs) are a type of neural network specifically designed for processing sequential data, such as time series, text, speech, or biological sequences [56]. Unlike Fully Connected Networks (FCNs), RNNs incorporate recurrent connections that allow them to retain and use information from previous states while processing new inputs.

This characteristic endows them with the ability to model temporal and contextual dependencies, which is essential in tasks where the order and relationships among elements in a sequence determine the meaning or dynamics of the phenomenon under analysis [19]. For example, in the case of text, the interpretation of a word depends on the preceding words; similarly, in a time series, future values largely depend on past patterns.

2.1.6. XGBoost (Extreme Gradient Boosting)

XGBoost is a supervised learning algorithm derived from the gradient boosting technique and has become one of the most widely used methods in the field of machine learning for regression and classification problems. Its foundation lies in the construction of an ensemble of decision trees generated sequentially, where each new tree aims to reduce the residual errors produced by the previous trees [33]. In this way, by combining multiple weak learners, the model becomes a highly powerful and accurate classifier or regressor.

One of the most relevant characteristics of XGBoost is its computational efficiency. The algorithm was designed to make intensive use of hardware resources, incorporating optimization techniques such as cache-aware memory usage, parallel processing, and branch pruning without significant loss of accuracy [13]. In addition, it includes L1 and L2 regularization mechanisms, which help control model complexity and mitigate the risk of overfitting, distinguishing it from more traditional boosting methods. Mathematically, XGBoost optimizes an objective function that combines prediction error with a penalty term for model complexity.

2.1.7. LightGBM (Light Gradient Boosting Machine)

LightGBM, developed by Microsoft, represents an evolution of gradient boosting methods, specifically designed to address the challenges of scalability and speed posed by large-scale datasets [37]. Like XGBoost, it is based on the construction of multiple decision trees that are aggregated sequentially; however, it introduces methodological innovations that make it a highly efficient tool in big data contexts [37].

One of its main distinguishing features is the use of the histogram-based learning technique, through which continuous values of predictor variables are grouped into intervals or “bins,” significantly reducing the computational complexity of calculating splits at tree nodes. This not only accelerates the training process but also reduces memory consumption, which is crucial when working with massive data volumes [55].

3. Optimization Metaheuristics

Optimization is the discipline concerned with finding the inputs of a function that minimize or maximize its value, which may be subject to constraints [50]. Optimization problems can be classified according to different factors such as their complexity, the presence or absence of constraints, their static or dynamic nature, linear or nonlinear formulation, and whether they are single-objective or multi-objective, among others. Regarding search techniques, these can be classified based on whether they guarantee obtaining the optimal result (exact techniques) or, alternatively, whether they allow the attainment of solutions close to the optimum (approximate techniques) [2].

Considering that combinatorial optimization consists of finding the best (optimal) solution among a finite set of alternative solutions, although exact techniques guarantee obtaining the optimal solution for any type of problem, they require a high computational cost. That is, to obtain the best solution, the required time grows exponentially with the size of the problem, and in some cases it becomes impossible to find it due to the time demanded. This has led to the emergence of techniques known as metaheuristics [4], which are high-level search procedures that apply one or more rules based on some source of knowledge in order to efficiently explore the search space [29].

There are different ways to classify and describe metaheuristics [59]. Depending on the selected characteristics, different taxonomies can be obtained: nature-inspired and non-nature-inspired, with or without memory, with one or multiple neighborhood structures, trajectory-based and population-based. Trajectory-based metaheuristics are those that use a single solution during the search process, and the result is also a single optimized solution. Trajectory-based metaheuristics include hill climbing (HC), simulated annealing (SA), tabu search (TS), greedy randomized adaptive search procedures (GRASP), variable neighborhood search (VNS), and iterated local search (ILS) [4].

The main population-based metaheuristics include genetic algorithms (GA) and evolutionary algorithms (EA), scatter search (SS), path relinking (PR), ant colony optimization (ACO), particle swarm optimization (PSO), estimation of distribution algorithms (EDA), and differential evolution (DE) [1;42].

Multi-objective metaheuristics can also be classified into trajectory-based methods and population-based methods. Trajectory-based methods include Pareto Archived Evolution Strategy (PAES) and Multi-Objective Simulated Annealing (MOSA), among others [3]. Population-based metaheuristics include Multi-Objective Tabu Search (MOTS), the Non-dominated Sorting Genetic Algorithm II (NSGA-II), Pareto Simulated Annealing (PSA), Single-Front Genetic Algorithm (SFGA), Strength Pareto Evolutionary Algorithm (SPEA/SPEA2), and Pareto Envelope-based Selection Algorithm (PESA/PESA-II). Some authors have also proposed hybrid approaches that combine aspects of two or more methods, such as Genetic Tabu Search (GTS), Multi-Objective Genetic Local Search (MOGLS), Multi-Objective Pareto Archived Evolution Strategy (M-PAES), multi-objective simulated annealing, and Multi-Objective Simulated Annealing and Tabu Search (MOSATS) [4;43].

3.1. Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO), introduced by [38], is inspired by the collective behavior observed in biological groups such as flocks of birds or schools of fish. This method is based on swarm intelligence and operates through the interaction of multiple agents (particles) that move within a solution space, influenced both by their own experience and by that of the group. Each particle updates its position by modifying its velocity according to two key influences: its personal best historical position (pbest) and the best position found by the group (gbest). This iterative process enables efficient exploration of the search space, achieving stable convergence in problems characterized by multiple parameters and high nonlinearity.

3.2. Cuckoo Search Algorithm (CS)

The Cuckoo Search (CS) algorithm, developed by [60], is inspired by the peculiar reproductive strategy of certain cuckoo species that lay their eggs in the nests of other birds, thereby increasing the survival chances of their offspring. This method employs what are known as Lévy flights, which allow long-range random movements and thus enhance global exploration capability. From a theoretical standpoint, CS is based on two fundamental mechanisms: evolutionary imitation, whereby nests containing less effective solutions are replaced, and global search supported by Lévy walks, which favor non-local jumps in the search space. This delicate balance between local exploitation and global exploration enables the algorithm to avoid being trapped in local optima and to find optimal solutions in complex optimization problems.

4. Decomposition Methods

Time series decomposition is a key component in the analysis of economic variables that tend to be highly volatile, such as agricultural product prices. Its objective is to decompose the original signal into components that are easier to interpret, which helps reduce noise, highlight hidden patterns, and improve the predictive capacity of models. In this research, four decomposition methods commonly used to handle non-stationary and non-linear time series have been integrated: EMD, SSA, VMD, and EWT. These methods allow the maize price series to be decomposed into different modes or components, thereby achieving a clearer and more structured representation that facilitates the predictive modeling process.

4.1. Empirical Mode Decomposition (EMD)

Empirical Mode Decomposition (EMD), introduced by [33], is used to decompose nonlinear and non-stationary signals into a set of functions known as Intrinsic Mode Functions (IMFs). Each IMF represents an oscillatory component with a different frequency, which facilitates the separation of noise, trends, and cyclical fluctuations present in the original series. This procedure is based on an iterative process known as sifting, which systematically extracts the modes of the signal by following the internal dynamics of the data without the need to make parametric assumptions. In the context of agricultural markets, EMD has proven to be highly effective in capturing abrupt fluctuations, stochastic noise, and irregular patterns related to supply, demand, and external shocks, thereby generating smoother signals that can be used in Machine Learning-based predictive models.

4.2. Singular Spectrum Analysis (SSA)

Singular Spectrum Analysis (SSA) is a robust technique for the decomposition and analysis of nonlinear time series, developed by [30] as a mathematical tool capable of extracting underlying structural components such as trends, periodic oscillations, and random noise. Unlike other traditional approaches, SSA combines principles of singular value decomposition and spectral analysis, enabling the reconstruction of complex signals through their disaggregation into representative principal components.

This approach is particularly useful for series that exhibit smooth long-term patterns and seasonal cycles, which are common in agricultural markets influenced by climatic, seasonal, and global supply dynamics. Its application allows the deterministic structure of the signal to be preserved, favoring the generation of reconstructed components that can be used in predictive models.

4.3. Variational Mode Decomposition (VMD)

Variational Mode Decomposition (VMD), developed by [18], is responsible for decomposing a time series into a predefined set of modes with specific frequency bands through a variational optimization process. Unlike EMD, VMD prevents mode mixing, resulting in more stable and mathematically well-controlled representations. Owing to its regulated structure, it becomes a highly effective tool for capturing significant oscillations in agricultural price behavior. This helps minimize noise and highlight cyclical patterns that are crucial for price forecasting in highly volatile markets.

4.4. Empirical Wavelet Transform (EWT)

The Empirical Wavelet Transform (EWT), presented by [28], is based on the construction of adaptive wavelet bases derived from the empirical spectrum of the signal. Through an automatic partitioning of the spectrum, this approach enables the extraction of specific frequency components, which helps capture abrupt transitions and local structures over time. EWT is particularly valuable for time series that exhibit structural changes and highly dynamic behavior, such as agricultural prices influenced by factors including climatic conditions, international demand, logistical aspects, and market events. Its application enhances precise signal decomposition, generating components with high predictive value.

5. Methodology

This article is framed within a theoretical–conceptual methodology, aimed at the analysis, integration, and systematization of existing knowledge on time series decomposition methods, machine learning algorithms, and optimization metaheuristics applied to agricultural price forecasting. A documentary research design is adopted, based on the review, selection, and integration of relevant scientific sources, with the purpose of constructing a conceptual model grounded in the main approaches, techniques, and analytical categories identified in the specialized literature.

The research follows an analytical–synthetic method, consisting of:

- Analysis, to decompose the literature into fundamental concepts (agricultural processes, decomposition methods, machine learning, and optimization).
- Synthesis, to integrate these concepts into a coherent theoretical framework that supports the methodological design of the proposed hybrid model.

6. Results

Based on the methodological design adopted in this research, three key theoretical axes that structure the analysis were identified: agricultural and commercial processes, advanced methods for time series decomposition, and machine learning techniques. Table 1 presents a detailed summary of the results obtained.

Table 1. Theoretical Foundation Matrix

Theoretical Axes	Definition	Authors	Units of Analysis / Subconcepts
Agricultural Processes	Set of productive and economic activities that transform agricultural inputs until their final commercialization within the agri-food chain	[14;22–24;48]	1. Agricultural processes 2. Post-harvest activities 3. Economic activities 4. Technical activities 5. Transportation 6. Marketing 7. Management 8. Storage 9. Quality control 10. Productive chain 11. Links 12. Wholesalers 13. Retailers
Decomposition Methods	Techniques that separate a time series into elementary components to identify patterns and reduce noise	[18;18;20;26; 28;28;28;32;33; 36]	14. Time series 15. Trend 16. Seasonality 17. Cycle 18. Noise 19. Signal 20. IMFs 21. EMD method 22. SSA method 23. VMD method 24. EWT method
Predictive Machine Learning	Supervised learning models capable of identifying patterns and generating future predictions on continuous variables such as price	[8], [9], [6]	25. Supervised ML 26. Regression 27. Classification 28. Tree-based models (XGBoost, LightGBM) 29. Neural networks (RNN, FCN)
Optimization Metaheuristics	Methods that help find the best (optimal) solution among a finite set of alternative solutions	[50], [29], [59], [1], [3], [43], [38], [60]	30. Particle Swarm Optimization (PSO) 31. Cuckoo Search (CS)

In accordance with the methodological structure of this study and with the objective of ensuring coherence between the theoretical framework, key concepts, and the established objectives, Table 2 presents the matrix that links the units of analysis, the related concepts, and the specific objective.

Table 2. Matrix of Main Units of Analysis

Main Units of Analysis	Associated Concept	Authors / Sources
Price	Monetary value of maize in wholesale markets	[11], [51]
Time series	Temporal observations of price values	[10]
EMD	Empirical decomposition into intrinsic mode functions	[32]
SSA	Singular spectrum decomposition	[30]
VMD	Variational mode decomposition	[18]
EWT	Empirical wavelet transform	[28]
XGBoost / LightGBM models	Boosting-based models for time series prediction	[33], [37]
Neural networks (RNN / FCN)	Models inspired by biological neurons	[40], [39], [56]
Cuckoo Search	Metaheuristic optimization technique	[60]
Particle Swarm Optimization (PSO)	Metaheuristic optimization technique	[38]

7. Conclusions

The theoretical foundation presented has been essential for establishing the conceptual and scientific axes capable of supporting the design of predictive models for price forecasting. By defining three conceptual axes: agricultural processes that shape productive and market dynamics, advanced methods for time series decomposition, and machine learning techniques focused on forecasting, a solid structure has been created that facilitates understanding of the phenomenon under analysis and guides the adopted methodological approach. Furthermore, by identifying and systematizing the derived units of analysis, such as price, time series components, and the algorithms employed, coherence between theory, experimental design, and the research objectives has been ensured.

Author Contributions: **Ojeda, A.:** Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization.
The author has read and agreed to the published version of the manuscript. Please refer to the CRediT taxonomy for the definitions of the terms. Authorship is limited to those who have made substantial contributions to the reported work.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable, since the present study does not involvehuman personnel or animals.

Informed Consent Statement: This study is limited to the use of technological resources, so nohuman personnel or animals are involved.

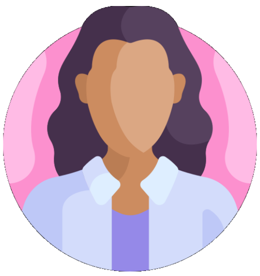
References

1. Aguilar, J. (2017a). Clasificación de problemas de optimización y métodos de búsqueda. *Revista Científica*.
2. Aguilar, J. (2017b). Resolución computacional de un problema de optimización combinatorio híbrido. *Ciencia e Ingeniería*, 38(2):99–106.
3. Alancay, D., Naretto, M. E., Castillo, J. I., and Álvarez, A. G. (2016a). Metaheurísticas: fundamentos y aplicaciones. *Revista de Investigación Operacional*.
4. Alancay, N., Villagra, S. M., and Villagra, N. A. (2016b). Metaheurísticas de trayectoria y poblacional aplicadas a problemas de optimización combinatoria. *Informes Científicos Técnicos – UNPA*, 8(1):202–220.

5. Arbeláez, M. A. and Ramírez, S. (n.d.). Política comercial de la cadena productiva del maíz amarillo en colombia.
6. Banerjee, T., Sinha, S., and Choudhury, P. (2022). Long term and short term forecasting of horticultural produce based on the lstm network model. *Applied Intelligence*, 52(8):9117–9147.
7. Bengio, Y. (2009). Learning deep architectures for ai. *Foundations and Trends in Machine Learning*, 2(1):1–127.
8. Benítez, R., Escudero, G., Kanaan, S., and Rodó, D. M. (2014). *Inteligencia artificial avanzada*. Editorial UOC, Barcelona, España.
9. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, New York.
10. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. Wiley, Hoboken, NJ, 5 edition.
11. Case, K. E., Fair, R. C., and Oster, S. M. (2017). *Principios de economía*. Pearson Educación, México D.F., 12 edition.
12. Castro, E. (2008). *Las cadenas productivas y la planificación estratégica*. Universidad Nacional de Colombia, Bogotá, Colombia.
13. Chen, Y., He, R., Zhu, M., and Wang, J. (2019). Application of xgboost for credit scoring: Comparison with logistic regression. *Expert Systems with Applications*, 136:262–273.
14. CIMMYT, CIAT, CIUF, UPRA, and AGROSAVIA (2020). Maíz para colombia: Visión 2030. Technical report, Ministerio de Agricultura y Desarrollo Rural, Bogotá, Colombia.
15. Congreso de la República de Colombia (2003). Ley 811 de 2003.
16. Corporación Universitaria del Norte (2017). Minería de datos en gestión del conocimiento de pymes de colombia. *Revista Virtual Universidad Católica del Norte*.
17. Departamento Nacional de Planeación (2014). Propuesta para desarrollar un modelo eficiente de comercialización y distribución de productos. Technical report, Misión para la Transformación del Campo.
18. Dragomiretskiy, K. and Zosso, D. (2014). Variational mode decomposition. *IEEE Transactions on Signal Processing*, 62(3):531–544.
19. Elmezughi, M. K., Salih, O., Afullo, T. J., and Duffy, K. J. (2022). Comparative analysis of major machine-learning-based path loss models for enclosed indoor channels. *Sensors*, 22(13):4967.
20. Fan, K., Xu, B., Zhang, M., Nan, M., and Huang, J. (n.d.). A new method for time series signal decomposition. *Unknown*.
21. Fayyad, U., Shapiro, G. P., and Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*.
22. Food and Agriculture Organization of the United Nations (2020). *El estado mundial de la agricultura y la alimentación 2020: Superar los desafíos relacionados con el agua en la agricultura*. FAO.
23. Food and Agriculture Organization of the United Nations (n.d.a). Actividades postcosecha y funciones económicas. Accessed: 2025-01-01.
24. Food and Agriculture Organization of the United Nations (n.d.b). Producción de cultivos | mecanización agrícola sostenible. Accessed: 2025-01-01.
25. Frawley, W. J., Shapiro, G. P., and Matheus, C. J. (1992). Knowledge discovery in databases: An overview. *AI Magazine*.
26. Garai, S. et al. (2023). Wavelets in combination with stochastic and machine learning models to predict agricultural prices. *Mathematics*, 11(13):2896.
27. Ghai, D., Tripathi, S. L., Saxena, S., Chanda, M., and Alazab, M. (2022). *Machine Learning Algorithms for Signal and Image Processing*. Wiley.
28. Gilles, J. (2013). Empirical wavelet transform. *IEEE Transactions on Signal Processing*, 61(16):3999–4010.
29. Glover, F. (1986). Future paths for integer programming and links to artificial intelligence. *Computers & Operations Research*, 13(5):533–549.
30. Golyandina, N., Nekrutkin, V., and Zhigljavsky, A. (2001). *Analysis of Time Series Structure: SSA and Related Techniques*. Monographs on Statistics and Applied Probability. Chapman and Hall/CRC, Boca Raton, FL.
31. Hirschman, A. O. (1958). *The Strategy of Economic Development*. Yale University Press, New Haven, Connecticut.
32. Huang, J., Zhang, M., Mujumdar, A. S., and Ma, Y. (2023). Technological innovations enhance postharvest fresh food resilience from a supply chain perspective. *Critical Reviews in Food Science and Nutrition*, pages 1–23.
33. Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C.-C., and Liu, H. H. (1998). The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A*, 454(1971):903–995.
34. Hutter, F., Kotthoff, L., and Vanschoren, J., editors (2019). *Automated Machine Learning: Methods, Systems, Challenges*. The Springer Series on Challenges in Machine Learning. Springer, Cham.
35. Isaza, J. (n.d.). Cadenas productivas: Enfoques y precisiones conceptuales.
36. Karaaslan, O. F. and Bilgin, G. (2020). Comparison of variational mode decomposition and empirical mode decomposition features for cell segmentation in histopathological images. In *2020 Medical Technologies Congress (TIPTEKNO)*, pages 1–4, Antalya, Turkey. IEEE.
37. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, volume 30.
38. Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of the International Conference on Neural Networks (ICNN'95)*, pages 1942–1948, Perth, Australia.
39. Langer, S. (2021). Analysis of the rate of convergence of fully connected deep neural network regression estimates with smooth activation function. *Journal of Multivariate Analysis*, 182:104695.

40. Linka, K. and Kuhl, E. (2023). A new family of constitutive artificial neural networks towards automated model discovery. *Computer Methods in Applied Mechanics and Engineering*, 403:115731.
41. Moine, J. and Haedo, A. (2011). Estudio comparativo de metodologías para minería de datos. Technical report, Universidad Nacional de La Plata.
42. Moreno Vega, J., Melián Batista, M., and Moreno Pérez, J. (2003). Metaheurísticas: Una visión global. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, 7(19):7–28.
43. Moreno-Vega, M., Padrón, J. M., and Verdegay, J. L. (2003). Metaheuristics: An overview of the current state-of-the-art. *European Journal of Operational Research*.
44. Mucherino, A., Papajorgji, P., and Pardalos, P. M. (2009). *Data Mining in Agriculture*. Springer, New York, NY.
45. Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, MA.
46. Navarro, R. E. Z. (2017). Plan de ordenamiento productivo para la cadena de maíz en colombia. Technical report, Unidad de Planificación Rural Agropecuaria (UPRA); Ministerio de Agricultura y Desarrollo Rural, Bogotá, Colombia.
47. Ocaña-Fernández, Y., Valenzuela-Fernández, L. A., and Garro-Aburto, L. L. (2019). Inteligencia artificial y sus implicaciones en la educación superior. *Propósitos y Representaciones*, 7(2):536–568.
48. OECD and FAO (2022). *OCDE-FAO Perspectivas Agrícolas 2013–2022*. OECD Publishing and FAO.
49. Osorio, N. E. A. (2020). El derecho de autor en la inteligencia artificial de machine learning. *Revista Jurídica*.
50. Pardalos, P. M. (2002). *Handbook of Applied Optimization*. Oxford University Press.
51. Parkin, M. (2015). *Microeconomía*. Pearson Educación, México D.F., 11 edition.
52. Porter, M. E. (1999). *Ser competitivo: Nuevas aportaciones y conclusiones*. Deusto, Bilbao, España.
53. Ramana, T. V., Ghantasala, G. S., Sathiyaraj, R., and Khan, M. (2024). *Artificial Intelligence and Machine Learning for Smart Community: Concepts and Applications*. CRC Press.
54. Saravia, C. D. (2009). *Comercialización y mercados agropecuarios*. Universidad Nacional de La Pampa, Santa Rosa, Argentina.
55. Seireg, H. R., Omar, Y. M. K., El-Samie, F. E. A., El-Fishawy, A. S., and Elmahalawy, A. (2022). Ensemble machine learning techniques using computer simulation data for wild blueberry yield prediction. *IEEE Access*, 10:64671–64687.
56. Sherstinsky, A. (2020). Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. *Physica D: Nonlinear Phenomena*, 404:132306.
57. Unidad de Implementación del Acuerdo de Paz (UIPAZ) (2022). Plan nacional para la promoción de la comercialización de la producción de la economía campesina, familiar y comunitaria. Technical report, Presidencia de la República, Bogotá, Colombia.
58. Varios autores (2023). *Sistemas de aprendizaje automático*. Ediciones de la U.
59. Vélez, J., Romero, P., and Cardona, L. (2007). Clasificación y comparación de metaheurísticas para optimización. *Revista Colombiana de Computación*.
60. Yang, X.-S. and Deb, S. (2009). Cuckoo search via lévy flights. In *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, pages 210–214, Coimbatore, India.
61. Yang, Y. and Fan, C. (2022). Efficient and robust time series prediction model based on remd-mmlp with temporal-window. *Expert Systems with Applications*, 207:117979.

Authors' Biography



Adelaida Ojeda-Beltran Business Administrator, graduated from the University of Atlántico, with a Master's degree in Organizational Management. Experienced in the use and management of virtual learning platforms such as Moodle and Blackboard. Certified by theServicio Nacional de aprendizaje (SENA) under the standard "Deliver distance training in accordance with technical and regulatory procedures." Experienced in the development of digital competencies, including the use and appropriation of Information and Communication Technologies (ICTs) in business contexts.