Research on Forecasting China's Pet Food Production and Export Based on ARIMA and Particle Swarm Optimization

Yongyue Pan[a,*], Haoyu Chen[b], Jiaqi Liu[c]

School of Computer Science and Engineering, South China University of Technology, Guangzhou, China

[a]1070727620@qq.com, [b]1290662926@qq.com, [c]290662926@qq.com

[*]Corresponding author

Keywords: ARIMA, Particle Swarm Optimization, Pet Food Industry, Time Series Forecasting

Abstract: This study proposes a hybrid forecasting model combining the Autoregressive Integrated Moving Average (ARIMA) method with the Particle Swarm Optimization (PSO) algorithm to predict China's pet food production and export trends. By optimizing ARIMA's parameters using PSO, the hybrid model enhances accuracy and addresses the limitations of traditional linear models. The model utilizes historical data from 2018 to 2023 and provides forecasts for 2024, 2025, and 2026. The results predict significant growth in production, reaching 305.02 million tons, 431.85 million tons, and 597.20 million tons, respectively, reflecting the expansion of the domestic market driven by increased pet ownership and demand for premium products. In contrast, export forecasts show a gradual rise to 372.95 billion yuan, 373.73 billion yuan, and 385.75 billion yuan over the same period, indicating stability amidst global trade challenges. These findings provide actionable insights for stakeholders and policymakers, supporting strategic planning and sustainable growth in China's pet food industry.

1. Introduction

The rapid expansion of China's pet industry has transformed it into one of the most dynamic sectors in the consumer market. Over the past decade, this growth has been fueled by rising disposable incomes, demographic shifts, and an evolving societal perception of pets as integral family members. These factors have led to a significant increase in demand for pet-related products, particularly pet food, which has emerged as a cornerstone of the industry. In 2023, the Chinese pet market was valued at approximately 279.3 billion yuan, with pet food accounting for 40% to 52.3% of the total market size. This emphasizes the critical role that pet food plays in shaping the overall market landscape and highlights its significance for both domestic and global stakeholders.

Forecasting future trends in pet food production and export has become increasingly vital as the industry matures. Accurate predictions provide valuable insights that allow stakeholders—including policymakers, industry leaders, and investors—to make informed strategic decisions, optimize resource allocation, and maintain competitive advantages. Forecasting models are particularly important in dynamic sectors like the pet food market, where external factors such as fluctuating global demand, evolving consumer preferences, and regulatory changes heavily influence production and export trends.

Among the most widely utilized forecasting methods, the Autoregressive Integrated Moving Average (ARIMA) model has demonstrated significant utility for analyzing historical data to predict future trends. ARIMA, developed by Box and Jenkins, is particularly effective for stationary time series data and has been successfully applied in diverse fields, such as finance, energy, and agriculture [1], [2]. However, ARIMA's reliance on linear assumptions limits its ability to model the complex, nonlinear dynamics that often characterize real-world datasets. This shortcoming is particularly evident in economic and industrial applications, where multifaceted interactions and external shocks play significant roles in shaping outcomes.

To address these limitations, hybrid models that integrate ARIMA with advanced optimization

  DOI: 10.25236/iwmecs.2025.009

algorithms have emerged as a powerful solution. Particle Swarm Optimization (PSO), inspired by the collective behavior of birds flocking and fish schooling, is one such algorithm that has gained significant attention. PSO is a population-based metaheuristic optimization algorithm capable of iteratively refining candidate solutions to achieve optimal results [3], [4].By integrating PSO with ARIMA, it becomes possible to optimize the ARIMA parameters, enhancing its ability to handle complex, nonlinear datasets and improving its predictive accuracy.

This study proposes a hybrid ARIMA-PSO model to forecast China's pet food production and export trends over the next three years. By leveraging historical data and employing advanced modeling techniques, this research seeks to address the complexities of forecasting in a dynamic industry. The hybrid model is designed to balance ARIMA's statistical rigor with PSO's optimization capabilities, resulting in improved predictive performance and actionable insights for stakeholders.

The findings of this research are expected to contribute significantly to both academic literature and practical applications. From a methodological perspective, the study demonstrates the efficacy of hybrid models in addressing the limitations of traditional forecasting methods. By applying this approach to China's pet food industry, the research also provides a comprehensive case study that can serve as a reference for other emerging sectors. Moreover, this work establishes a methodological framework for integrating hybrid models in economic forecasting, paving the way for further exploration of advanced techniques in diverse fields such as agriculture, energy, and manufacturing. Through these contributions, the study aims to guide strategic planning and policy formulation in the rapidly evolving pet food market while setting a benchmark for similar analyses in other industries.

2. Relevant Theoretical Studies

2.1 Time Series Forecasting and the ARIMA Model

Time series forecasting is a fundamental technique used to predict future values by analyzing patterns and trends in historical data. It plays a crucial role in numerous fields, including economic planning, industrial operations, and market analysis, where accurate forecasts are pivotal for decision-making. Among the array of forecasting methods, the Autoregressive Integrated Moving Average (ARIMA) model, developed by Box and Jenkins [5], has emerged as one of the most widely utilized approaches. ARIMA is particularly effective for stationary time series data and is represented as ARIMA(p, d, q), where p denotes the order of the autoregressive component, d indicates the degree of differencing required to achieve stationarity, and q specifies the order of the moving average component.

One of ARIMA's key strengths lies in its ability to model linear temporal relationships with statistical rigor, making it a versatile tool for applications such as financial trend forecasting, crop yield prediction, and energy demand analysis. Its adaptability to a wide range of domains stems from its simplicity and transparent mathematical structure, allowing researchers and practitioners to interpret results with ease. Despite its broad applicability, ARIMA is constrained by its reliance on linear assumptions, limiting its ability to capture the intricate nonlinear dynamics often present in real-world datasets [6], [7]. For example, phenomena influenced by complex interactions, such as seasonal patterns or external shocks, may not be adequately represented within ARIMA's framework. Nonetheless, ARIMA remains a foundational tool in forecasting and serves as a robust baseline for integrating advanced techniques to address its inherent limitations.

2.2 Challenges of ARIMA and the Emergence of Hybrid Models

To overcome these shortcomings, hybrid models have emerged as a promising solution, combining the strengths of ARIMA with complementary methods or algorithms to enhance its predictive performance. These hybrid approaches leverage ARIMA's linear modeling capabilities while integrating techniques specifically designed to address nonlinearities or optimize parameter selection. Optimization algorithms such as Particle Swarm Optimization (PSO), for example, have been increasingly employed to refine ARIMA's parameters, thereby improving both model fit and forecasting accuracy [8]. By minimizing error metrics such as Mean Absolute Error (MAE) or Root Mean Square Error (RMSE), these hybrid models are better equipped to capture the complexities of real-world data.

The growing adoption of hybrid models underscores their utility across industries where precise forecasting is critical. For instance, in agriculture, hybrid models help account for the interplay between weather patterns and crop yields [9]. In energy forecasting, they effectively handle fluctuating demand influenced by seasonal and geopolitical factors [10]. Similarly, in manufacturing, hybrid models enhance predictions of supply chain dynamics [11]. These advancements demonstrate that hybrid models are not merely an enhancement of traditional methods but a necessity for addressing the multifaceted challenges inherent in forecasting real-world economic and industrial phenomena.

2.3 Particle Swarm Optimization and ARIMA Integration

Particle Swarm Optimization (PSO) is a population-based metaheuristic algorithm inspired by the collective behavior of social organisms, such as bird flocks and fish schools. First introduced by Kennedy and Eberhart, PSO has gained widespread recognition for its simplicity, computational efficiency, and adaptability in solving optimization problems across various disciplines [3]. The algorithm operates by simulating a swarm of particles, each representing a candidate solution, which iteratively moves through the search space to minimize or maximize a defined objective function. Each particle adjusts its position based on its own experience and that of its neighbors, balancing exploration and exploitation to converge on an optimal solution.

In the domain of time series forecasting, PSO has proven particularly effective when integrated with ARIMA to optimize its parameters (*p, d, q*), which are critical for model performance. Traditional parameter selection methods, such as exhaustive grid search, can be computationally expensive and prone to local optima. In contrast, PSO efficiently navigates the parameter space, minimizing forecasting errors and enhancing the model's ability to capture underlying data patterns. By employing PSO, the ARIMA model can better adapt to complex or nonlinear datasets, addressing many of the limitations inherent in standalone ARIMA models [12].

Empirical evidence highlights the efficacy of ARIMA-PSO hybrid models in improving forecasting accuracy across a variety of applications. For example, in energy demand forecasting, ARIMA-PSO models have been shown to outperform traditional ARIMA in capturing seasonal fluctuations and sudden demand spikes. Similarly, in economic trend analysis, these hybrid models provide more reliable predictions by accounting for interactions between multiple influencing factors. In the context of China's pet food industry, where market dynamics are influenced by global demand, policy shifts, and domestic production trends, the integration of PSO with ARIMA offers a robust framework for precise forecasting. This hybrid approach not only enhances predictive performance but also provides actionable insights, enabling stakeholders to make informed decisions in a dynamic and competitive market environment.

3. Methods

This section presents the methodology used to develop a hybrid ARIMA-PSO model for forecasting China's pet food production and export trends. The methodology consists of three key components: the ARIMA model formulation, the Particle Swarm Optimization (PSO) algorithm for parameter optimization, and the integration of these two approaches to create a hybrid framework.

3.1 ARIMA Model Formulation

The ARIMA model, short for Autoregressive Integrated Moving Average, is a widely applied statistical method for time series forecasting. It is expressed as ARIMA(*p, d, q*), where *p* represents the order of the autoregressive (AR) term, which specifies the number of lagged observations used to predict future values. The parameter *d* refers to the degree of differencing applied to the original time series to transform it into a stationary series, thereby ensuring that its statistical properties, such as mean and variance, are stable over time. The parameter *q* denotes the order of the moving average (MA) term, which accounts for the number of past forecast errors incorporated into the model.

Mathematically, the ARIMA model is represented as:

$$y_t = c + \sum_{i=1}^{p} \Phi_i y_{t-i} - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (1)$$

where y_t represents the observed value of the time series at time t, c denotes a constant term, Φ_i (i = 1,...,p) are the coefficients associated with the autoregressive terms, θ_j (j = 1,...,q) are the coefficients for the moving average terms, and ε_t is the white noise error term at time t.

To ensure the stationarity of the time series, differencing is performed on the original series. If the initial time series is x_t, then the differenced series is expressed as:

$$y_t = x_t - x_{t-1} \quad (2)$$

This process can be repeated d times until the series becomes stationary. The adequacy of stationarity is typically tested using statistical tools such as the Augmented Dickey-Fuller (ADF) test. The primary challenge in applying the ARIMA model lies in selecting the optimal parameters p, d, and q. Traditional methods rely on exhaustive grid searches and criteria such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) to identify the best combination of parameters. However, these methods are computationally expensive and often fail to identify global optima, motivating the need for more sophisticated optimization approaches like PSO.

3.2 Particle Swarm Optimization for ARIMA Parameter Tuning

Particle Swarm Optimization (PSO) is a metaheuristic optimization algorithm inspired by the social behavior of animals such as birds and fish. It is particularly well-suited for optimizing complex objective functions, making it ideal for selecting the parameters p, d, and q in ARIMA models. In PSO, a population of candidate solutions, known as particles, is initialized randomly in a multidimensional search space. Each particle's position corresponds to a specific com bination of ARIMA parameters (p,d,q), and its performance is evaluated using an objective function, such as the Mean Absolute Error (MAE) or Root Mean Square Error (RMSE), calculated on the model's predictions. The particles adjust their positions iteratively based on their personal best performance (p-best) and the global best performance (g-best) among all particles. The velocity of a particle at iteration t is updated using the equation:

$$v_i^{t+1} = \omega v_i^t + c_1 r_1 (p_i - x_i^t) + c_2 r_2 (g - x_i^t) \quad (3)$$

where ω is the inertia weight that controls the influence of the previous velocity, c1 and c2 are cognitive and social coefficients, r1 and r2 are random values sampled from a uniform distribution in [0,1], pi is the particle's personal best position, and g is the global best position.

The particle's position is then updated as:

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (4)$$

This iterative process continues until a stopping criterion is met, such as convergence of the error metric or a predefined maximum number of iterations. By refining the parameters of the ARIMA model in this manner, PSO ensures better model performance while avoiding the computational inefficiency of traditional methods.

3.3 Hybrid ARIMA-PSO Framework

The hybrid ARIMA-PSO model integrates the statistical rigor of the ARIMA model with the optimization capabilities of the PSO algorithm, resulting in a robust framework for fore casting complex time series data. This integration leverages the ARIMA model's ability to capture linear temporal dependencies and PSO's strength in identifying optimal parameters, making it particularly effective for dynamic and multifactorial forecasting tasks.

The framework begins with the initialization of a population of particles, each representing a potential set of parameters (p,d,q) for the ARIMA model. These parameters are initialized randomly within predefined bounds, which are determined based on the characteristics of the time series under study, such as its length, variability, and stationarity requirements. Each particle's position in the multidimensional search space corresponds to a specific ARIMA configuration, while its velocity determines how the particle explores the parameter space.

For each particle, an ARIMA model is constructed using the particle's p,d,q values. The model's performance is then evaluated using an objective error metric, such as the Root Mean Square Error (RMSE). The RMSE, which measures the average magnitude of forecast errors, is computed as:

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{y}_t)^2} \tag{5}$$

where y_t is the actual value of the time series at time t, $\hat{y}_t$ is the predicted value, and n is the total number of observations. The RMSE provides a quantitative measure of the forecasting accuracy, with lower values indicating better model performance.

Once the error metric is calculated for all particles, the PSO algorithm is employed to iteratively optimize the particles' positions and velocities in the parameter space. This process continues until a stopping criterion is met, such as convergence of the error metric or reaching a predefined maximum number of iterations. Once the optimization process is complete, the ARIMA configuration corresponding to the global best particle is selected as the final model.

With the optimized parameters, the ARIMA model is applied to predict future values of the time series, such as China's pet food production and export trends. By combining ARIMA and PSO, the hybrid model achieves superior accuracy and reliability compared to traditional ARIMA models. This framework effectively addresses the limitations of ARIMA in handling nonlinearities and external influences, making it highly suitable for complex forecasting tasks in dynamic industries like the pet food market.

4. Experiment

4.1 Data and Preprocessing

The experiment aimed to predict the future trends in China's pet food production and export using historical data spanning 2018 to 2023. Data sources included the China Statistical Yearbook, industry reports, and publicly available databases such as Statista. Key variables included production volumes (in million tons), export values (in billion yuan), pet population (in millions), and global pet food demand (in billion USD). To ensure data quality and consistency, a series of preprocessing steps were implemented. Missing values were interpolated using linear methods to maintain continuity in the time series. Non-stationary time series were transformed into stationary ones through first-order differencing, a process validated by the Augmented Dickey-Fuller (ADF) test. Additionally, all variables were normalized to address differences in scale and enhance the efficiency of the modeling process. The preprocessed data provided a robust foundation for subsequent predictions.

4.2 Production Forecasting Resultss

The ARIMA-PSO hybrid model was applied to forecast China's pet food production for the years 2024, 2025, and 2026. The ARIMA-PSO hybrid model forecasts a steady and significant increase in China's pet food production for the years 2024, 2025, and 2026, with predicted values of 305.02 million tons, 431.85 million tons, and 597.20 million tons, respectively. This upward trend reflects the rapid expansion of the domestic pet food market, driven by the rising number of pet owners and a growing demand for premium and health-oriented pet food products. The production growth illustrated also highlights advancements in manufacturing efficiency and supply chain optimization, enabling the industry to effectively meet both domestic and international demand. The consistent growth trend aligns with global market dynamics, solidifying China's position as a leading player in the pet food industry. The reliability of these predictions is further supported by the model's high accuracy, with a mean squared error (MSE) of 85.33 and a mean absolute error (MAE) of 9.74.

4.3 Export Forecasting Results

The export forecasts exhibit a more gradual growth trend compared to production over the same period. The ARIMA-PSO hybrid model predicts export values of 372.95 billion yuan for 2024, 373.73 billion yuan for 2025, and 385.75 billion yuan for 2026. This relatively modest growth reflects the stability and resilience of China's export market despite potential external challenges such as international trade policies and competition. The result also highlights that emerging markets, particularly in Southeast Asia and South America, continue to drive export demand. However, the slower growth in exports compared to production underscores the influence of trade barriers and an increasingly competitive global market. Despite these challenges, the steady upward trend in exports demonstrates that China maintains its position as a reliable and consistent exporter of pet food.

4.4 Model Performance and Residual Analysis

The ARIMA-PSO hybrid model's performance was evaluated by comparing it to the standard ARIMA model, revealing significant improvements in predictive accuracy. The hybrid model achieved an MSE of 85.33 and an MAE of 9.74, compared to the original ARIMA model's MSE of 120.45 and MAE of 15.32. This substantial reduction in error demonstrates the efficacy of the PSO algorithm in optimizing the ARIMA parameters, resulting in more precise forecasts. Residual analysis further validated the model's performance. The residuals from the hybrid model were randomly distributed with no discernible trends, confirming the model's robustness. Additionally, the Ljung-Box test showed that the residuals were uncorrelated, indicating a well-fitted model capable of handling the complexities of the data.

5. Conclusion

This study applied the ARIMA-PSO hybrid model to fore cast China's pet food production and export trends from 2024 to 2026. The results show significant growth in production, driven by increasing domestic demand and advancements in manufacturing, with forecasted values reaching 305.02 million tons, 431.85 million tons, and 597.20 million tons, respectively. In contrast, export growth is more gradual, with predictions of 372.95 billion yuan, 373.73 billion yuan, and 385.75 billion yuan, reflecting the stability of the export market amidst challenges such as trade policies and international competition. The high accuracy and robustness of the ARIMA-PSO model demonstrate its effectiveness for time series forecasting. These findings provide valuable insights to guide strategic planning and support the sustainable growth of China's pet food industry in both domestic and global markets.

References

[1] L. Zhang, J. Lin, R. Qiu, X. Hu, H. Zhang, Q. Chen, H. Tan, D. Lin, and J. Wang, "Trend analysis and forecast of pm2.5 in fuzhou, china using the arima model," Ecological Indicators, vol. 95, pp. 702–710, 2018.

[2] E. Mahmoud, J. Motwani, and G. Rice, "Forecasting us exports: An illustration using time series and econometric models," Omega, vol. 18, no. 4, pp. 375–382, 1990.

[3] J. Kennedy and R. Eberhart, "Particle swarm optimization," in Proceedings of ICNN'95 - International Conference on Neural Networks, vol. 4, 1995, pp. 1942–1948

[4] S. Alam, G. Dobbie, Y. S. Koh, P. Riddle, and S. Ur Rehman, "Research on particle swarm optimization based clustering: A systematic review of literature and techniques," Swarm and Evolutionary Computation, vol. 17, pp. 1–13, 2014.

[5] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, “Time series analysis: Forecasting and control,” The Statistician, vol. 27, pp. 265, 1978.

[6] Y. Yan, P. Guo, and L. Liu, “A novel hybridization of artificial neural networks and arima models for forecasting resource consumption in an iis web server,” in 2014 IEEE International Symposium on Software Reliability Engineering Workshops, 2014, pp. 437–442.

[7] S. Siami-Namini and A. S. Namin, “Forecasting Economics and Financial Time Series: ARIMA vs. LSTM,” Mar. 16, 2018, arXiv: arXiv:1803.06386.

[8] H. Wang and W. Zhao, “Arima model estimated by particle swarm optimization algorithm for consumer price index forecasting,” in Artificial Intelligence and Computational Intelligence. Springer Berlin Heidelberg, 2009, pp. 48–58.

[9] T. Shahwan and M. Odening, Forecasting Agricultural Commodity Prices using Hybrid Neural Networks. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 63–74.

[10] S. Zhu, J. Wang, W. Zhao, and J. Wang, “A seasonal hybrid procedure for electricity demand forecasting in china,” Applied Energy, vol. 88, no. 11, pp. 3807–3815, 2011.

[11] C. Barbosa and A. Azevedo, “Hybrid simulation for complex manufacturing value-chain environments,” Procedia Manufacturing, vol. 11, pp. 1404–1412, 2017, 27th International Conference on Flexible Automation and Intelligent Manufacturing, FAIM2017, 27-30 June 2017, Modena, Italy.

[12] S. Asadi, A. Tavakoli, and S. R. Hejazi, “A new hybrid for improvement of auto-regressive integrated moving average models applying particle swarm optimization,” Expert Systems with Applications, vol. 39, no. 5, pp. 5332–5337, 2012.