The Relationship Between Vegetable Sales Volume and Cost-Plus Pricing Strategies: A Multi-Method Analysis and Forecast

Haozhe Wang [1, *, #], Yi Liu [1, #], Tinghui Wang [2]

[1] School of Computer Science, Inner Mongolia University, Hohhot, China, 010021

[2] School of Mathematical Sciences, Inner Mongolia University, Hohhot, China, 010021

* Corresponding author: 0211122678@mail.imu.edu.cn

[#]These authors contributed equally.

Abstract. In today's highly competitive supermarket industry, optimizing sales strategies for perishable goods such as vegetables is crucial for maintaining profitability and market share. This study provides an analytical framework for optimizing vegetable sales strategies in supermarkets, focusing on perishability and consumer behavior. Through regression analysis, it explores the sales volume and pricing relationship, identifying the most effective model with a Mean Squared Error (MSE) criterion. Time series models forecast future needs, reflecting on historical sales' cyclical and seasonal trends. A Particle Swarm Optimization (PSO) algorithm, tailored with constraints, offers predictions on sales volumes and pricing, aiming to enhance profitability. Iterative refinements reveal that the model achieves stability and accuracy in predicting profits. The findings emphasize the critical role of data-driven strategies in the competitive supermarket sector, advocating for adaptive approaches in response to evolving market conditions.

Keywords: Particle swarm optimization, Correlation analysis, Time series forecasting, Vegetable strategy optimization.

1. Introduction

The supermarket industry operates in a dynamic and competitive environment where the effective management of perishable goods, such as vegetables, plays a critical role in ensuring customer satisfaction, maintaining profitability, and achieving sustainability goals [1, 2]. Vegetables, with their short shelf life and varying consumer demand patterns, represent a particularly challenging category of goods to manage [3]. Factors such as seasonal changes, weather conditions, and holidays significantly affect their consumption patterns, thereby complicating inventory management and pricing strategies [4].

In this work, the methodology combines regression analysis and time series forecasting to understand sales and pricing dynamics for various vegetable categories. We then apply a Particle Swarm Optimization (PSO) algorithm for predicting sales volumes and optimizing pricing strategies. This streamlined approach enables us to develop data-driven insights and strategic recommendations, focusing on enhancing operational efficiency and market responsiveness without detailing specific results.

2. Future Forecast of Vegetable Sales

The data for this study was sourced from http://www.mcm.edu.cn/. In this research, the total monthly sales volume and cost-plus pricing for each vegetable category were calculated. Subsequently, scatter plots of cost-plus pricing versus sales volume over time were created to preliminarily determine their relationship. For different vegetable models, various fitting functions were employed for fitting, specifically, we experimented with Linear functions, Logistic regression, and Exponential-logarithmic mixing. By exploring different fitting functions for various types of vegetables, we aimed to identify the most suitable fitting algorithm for each category.

The fitting process is based on minimizing the Euclidean distance between each discrete point and the fitting curve, continuously training the parameters using the gradient descent algorithm [5]. In essence, it involves connecting a series of points on a plane with a smooth curve. To evaluate the effectiveness of the fit, Mean Squared Error (MSE) is used as the criterion [6], which is defined as the average of the squared differences between the fitted function values and the original scatter plot values. The specific formula is as follows:

$$\mathrm{MSE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{Y_{pre}} - \mathrm{Y})^2 \tag{1}$$

In the formula, Y_{pre} represents the function values corresponding to the fitting curve, while Y denotes the function values of the original scatter points. By calculating the value of MSE, we can better assess the quality of the fitting function. A smaller MSE value indicates a higher precision of the fit.

Given that the scatter plot exhibits an increasing trend with a linear relationship, we initially use a linear function 2 to perform regression fitting on the scatter plot [7].

$$\mathrm{Y1} = \mathrm{beta\_1X} + \mathrm{beta\_2} \tag{2}$$

After fitting the six vegetable categories and calculating the MSE for each category, the overall fitting results are quite good. However, for the data on chili peppers, the scatter points are not concentrated and there are many outliers, making the linear regression fitting less effective. Upon further observation of the scatter plot, it is noted that the growth of aquatic root and stem vegetables starts slow, speeds up in the middle, and then slows down again. Therefore, we attempt to use a logistic regression curve for fitting, with the logistic regression function being calculated as formula 3 [8]:

$$\mathrm{Y2} = \frac{1}{1+\mathrm{e}^{-\mathrm{beta\_1(X-beta\_2)}}} \tag{3}$$

By employing the logistic regression function for fitting, the fit for aquatic root and stem vegetables has improved. However, for some categories, the fitting of scatter plots in the lower left and upper right areas is not very good, with the fitting curve deviating significantly. Moreover, observing the distribution of mature specimens, some category scatter plots exhibit both exponential and logarithmic trends. Therefore, we will use a hybrid function that considers both exponential and logarithmic elements for fitting, with the fitting function as given in formula 4 [9]:

$$\mathrm{Y3} = \mathrm{a} + \mathrm{b} * \log(\mathrm{c} * \mathrm{X}) + \mathrm{d} * \mathrm{e}^{\mathrm{e}'\mathrm{X}} \tag{4}$$

Where a, b, c, d, and e' are fitting parameters. Upon verification, the exponential-logarithmic hybrid function effectively resolves issues of uneven fitting and demonstrates excellent performance, particularly within the aquatic root and stem vegetables and chili pepper categories. A comparison of the three fitting methods is shown in Table 1:

Table 1. Comparison of the three fitting methods

Categories	Target	Linear Function	Logistic regression	Exponential logarithmic mixing
Mosaic and leafy	mse	0.039189075	0.039706131	0.043915202
	parameter	beta_1 = 0.586611 beta_2 = 0.221728	beta_1 = 2.576553 beta_2 = 0.472031	a = 40.774710, b = 0.646663, c = 7.106964, d = -40.768822, e = 0.013611
Cauliflower	mse	0.067665476	0.065712664	0.069262135
	parameter	beta_1 = 0.758751 beta_2 = 0.075210	beta_1 = 3.172966 beta_2 = 0.555521	a = 24.211767, b = 0.337003, c = 3.723640, d = -24.020078, e = -0.006211
Aquatic rhizomes	mse	0.115058661	0.117239514	0.111781194
	parameter	beta_1 = 1.091222 beta_2 = -0.007658	beta_1 = 5.823126 beta_2 = 0.441717	a = -22.038478, b = 0.210604, c = 0.000000, d = 42.799605, e = 0.013590
Eggplant	mse	0.092816853	0.091148768	0.097695567
	parameter	beta_1 = 0.729372 beta_2 = 0.075599	beta_1 = 3.150546 beta_2 = 0.576330	a = -3.744563, b = 0.183054, c = 0.000000, d = 8.274195, e = 0.036931
Pepper	mse	0.036137686	0.035063289	0.034815029
	parameter	beta_1 = 0.624239 beta_2 = 0.291578	beta_1 = 3.270930 beta_2 = 0.319389	a = 23.956223, b = 0.384006, c = 4.034872, d = -23.411741, e = 0.014758
Edible fungus	mse	0.067581699	0.065887186	0.068283106
	parameter	beta_1 = 0.873758 beta_2 = 0.040576	beta_1 = 3.944147 beta_2 = 0.519363	a = -0.671267, b = 0.137373, c = 1.373163, d = 0.949517, e = 0.478690

By comparing the three fitting methods, we can determine the best-fitting approach for each category using the Mean Squared Error (MSE). Hence, the final fitting results showing the relationship between cost-plus pricing and sales volume for each category are illustrated in Figure 1:

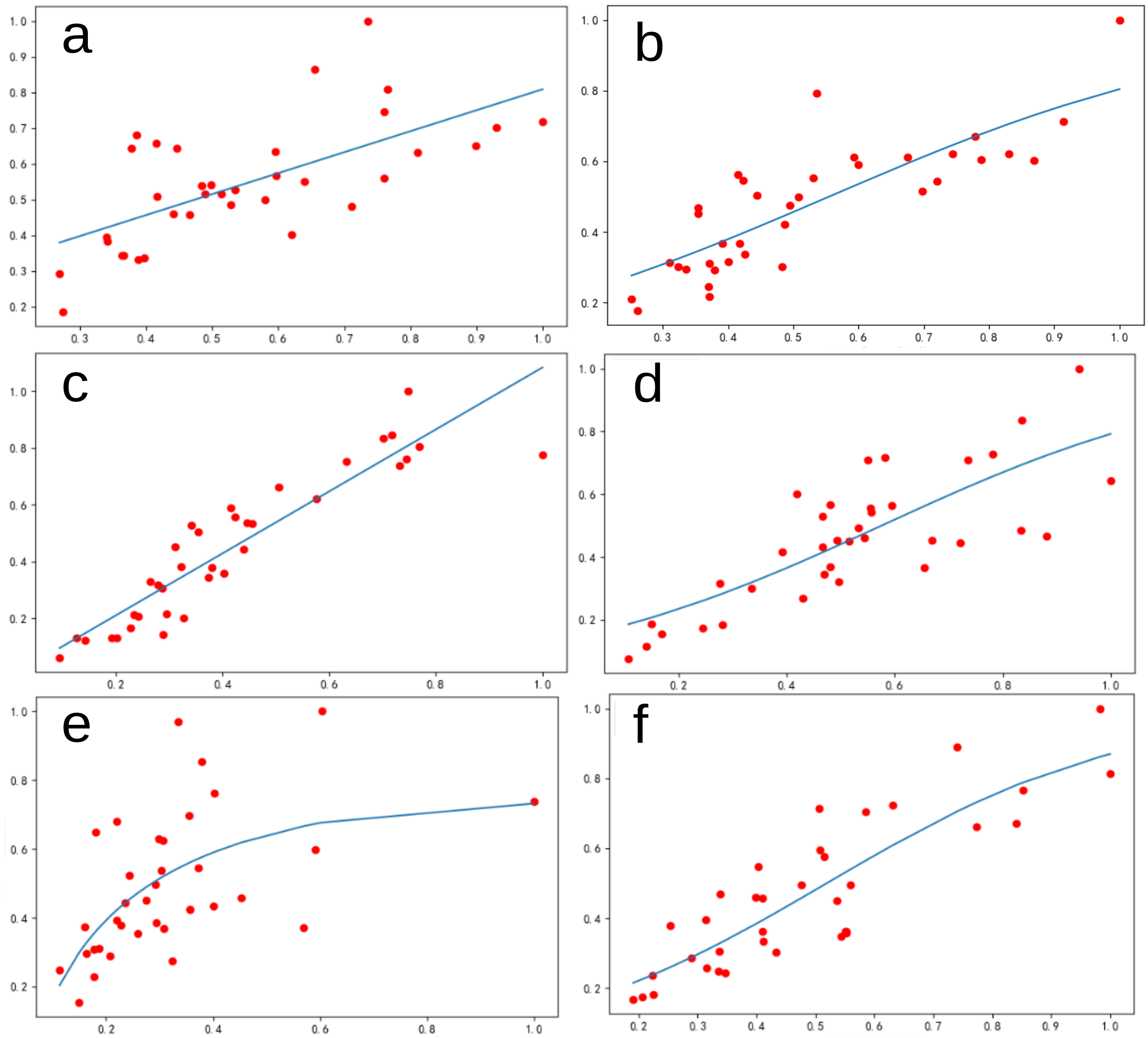

Figure 1. Illustration of the Fitting Effects for Different Vegetable Categories. (a) Mosaic leafy (b) Cauliflower (c) Aquatic rhizomes (d) Eggplant(e) Pepper (f) Edible Fungus

Further, we observed that the variety of vegetable supplies is more abundant from April to October, exhibiting distinct seasonal characteristics. In light of this, we opted to use the ARIMA time series model for future data forecasting. Initially, it is necessary to conduct an ADF test on the data to ascertain the stability of the time series, as illustrated in Figure 2.

The partial autocorrelation function X is obtained, and based on the truncation of the partial autocorrelation function, the initial model order is judged to be 5. The parameters are estimated using the least squares method, and the model residual variance and AIC values for orders within 10 are calculated. The AIC criterion is applied for model order selection, choosing the order that minimizes the AIC value. The corresponding orders for each category of vegetables are shown in Table 2.

Table 2. Types of vegetables and their corresponding orders

mosaic and leafy	cauliflower	aquatic rhizomes	Eggplant	Pepper	Edible fungus
8	7	10	6	4	10

At this point, the study can use the processed sample data series to perform single-step predictions, thereby obtaining the daily sales volume of each vegetable category over time. The forecast results are presented as shown in Table 3.

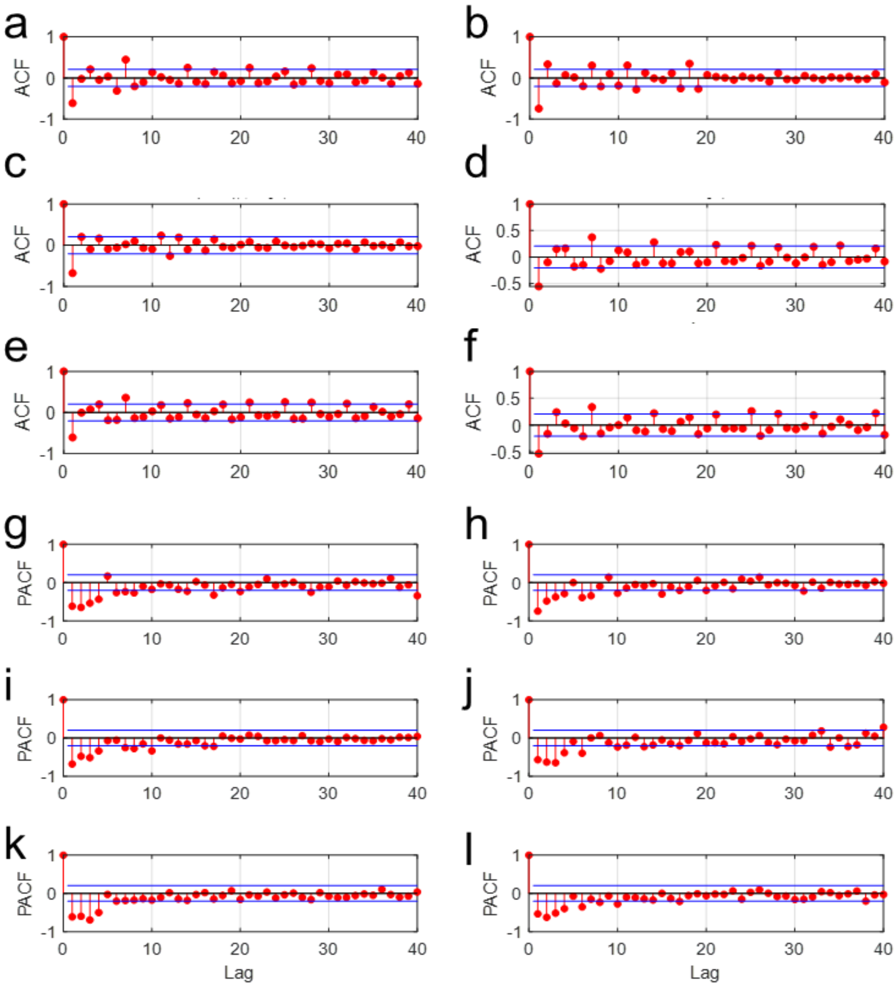

Figure 2. ACF and PCAF charts. a-f correspond to the ACF charts for mosaic and leafy, cauliflower, aquatic rhizomes, eggplant, pepper, and edible fungus, respectively. g-l correspond to the PCAF charts for the same types of vegetables.

Table 3. Forecast Results Table

	Mosaic and leafy	Cauliflower	Aquatic rhizomes	Eggplant	Pepper	Edible fungus
2023/7/1	-117.7423	-29.1844	-4.7735	-48.2403	-97.8769	-43.9528
2023/7/2	229.1092	29.7298	36.1361	37.5494	162.2737	64.7264
2023/7/3	406.9213	36.9630	7.2718	80.5531	282.3641	137.6280
2023/7/4	-57.4750	10.9022	13.9739	1.3866	19.2275	2.5973
2023/7/5	193.8986	16.9812	47.1296	24.3564	115.4684	115.6569
2023/7/6	179.8178	-7.9749	12.6345	44.3198	60.1900	36.7403
2023/7/7	33.8513	-2.2761	4.0856	17.5830	10.0926	-8.0213

3. Exploring the Supermarket Revenue Maximization Model

3.1. Model preparation

First, we need to identify the factors that influence profit. From the problem statement, we understand that profit is affected by sales volume, the selling price of the products, the purchase price of the products, and the rate of product loss. Total profit is defined as the difference between the total revenue earned and the total expenses incurred on a given day. This can be represented by Equation 5.

$$W = G - O \tag{5}$$

Here, *W* represents profit, *G* denotes the total revenue for the day, and *O* represents all other expenses incurred that day apart from the revenue. The revenue is calculated as the product of the selling price and the sales volume for the day.

$$G = Z * P \tag{6}$$

Here, *P* represents the selling price of the product. The expenses incurred on a given day, excluding revenue, can be represented as the product of the day's total replenishment volume and the purchase price of the products.

$$O = S * H \tag{7}$$

Here, *H* represents the purchase price of the products. Consequently, the total profit of the products can be expressed as follows:

$$W = Z * P - \frac{Z}{1-\alpha} * H \tag{8}$$

3.2. Particle swarm optimization

By processing and matching data, we identified the sales performance of individual items from June 24, 2023, to June 30, 2023. As the goal was to determine the total number of salable items, we considered the total profit of each item from the previous week as a unit. Utilizing Equation 8, we calculated the profit for all items during this period and sorted the results in descending order to identify the items with the highest profits. This approach enabled us to find 33 salable items.

To solve the maximization model, we plan to utilize the Particle Swarm Optimization (PSO) algorithm [10]. PSO is a collective intelligence method that guides particles—simplified as points in a search space—toward optimal solutions based on their and their peers' past experiences. It eschews evolutionary operators, instead adjusting each particle's trajectory by blending its personal best position with the swarm's best findings. This dynamic adjustment encourages both exploration and exploitation within the search space, akin to a flock of birds homing in on food through shared intelligence. Starting with a randomly generated set of particles as potential solutions, PSO iteratively refines these through a fitness function, aiming for the global optimum by balancing individual insights with collective wisdom.

When forecasting based on the previous week's sales data for individual items, it's crucial to consider factors affecting profit, as maximizing profit for all products is our goal when predicting replenishment and pricing for July 1. To achieve overall profit maximization, it's sufficient to focus on maximizing profit for each item individually. Our analysis indicates that an item's pricing and loss rate remained relatively stable over the previous week. Thus, the constraints are determined solely by an item's total sales volume and selling price. The requirement, as stated in the problem, is that the order quantity for each item must meet a minimum display quantity of 2.5 kilograms, and given the constraint to meet market demand as much as possible, this implies that the sales volume must be at least 2.5 kilograms. With these considerations, we've clarified the relationship between the objective function and decision variables, framing our task as a constrained multi-objective optimization forecasting problem. To solve this, we will develop and employ a regularization-based PSO forecasting model.

Here, we normalize the constraints and introduce a swarm of particles to construct a regularization-based PSO forecasting model. This model predicts the pricing and sales volume for July 1st, along with the anticipated maximum profit for individual items, as shown in Table 8. With the pricing strategy for individual items now determined, the predicted replenishment volume can be calculated by inserting the forecasted sales volume into the equation. Figure 3 illustrates the iteration process for a product forecast, where the profit stabilizes and tends towards optimization after more than 3600 iterations, indicating the model's strong predictive capability.

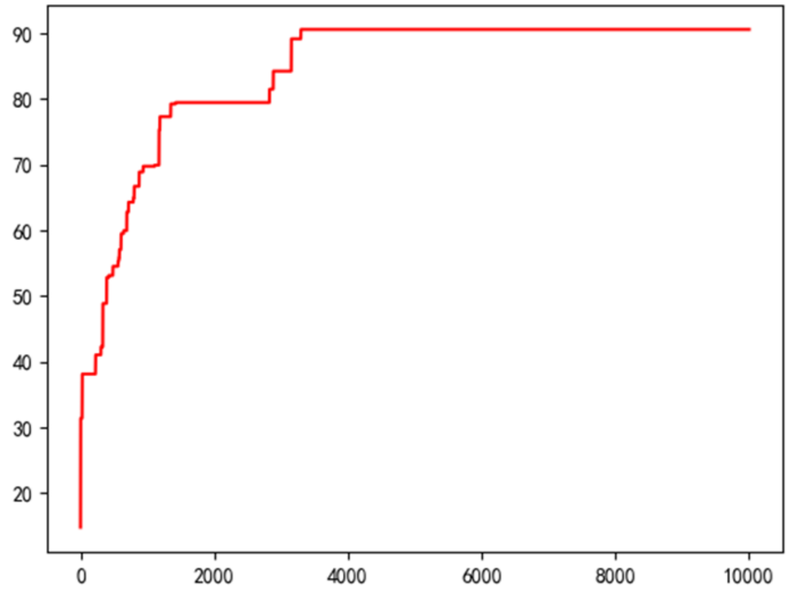

Figure 3. PSO iteration process

3.3. Market Demand Analysis and Strategy Formulation

By comprehensively collecting and analyzing data on customer demand, the impact of different weather conditions and holidays on vegetable sales, the stability and cost control of the supply chain, and competitors' pricing and promotion strategies, supermarkets can gain a complete insight into market demands. This multifaceted compilation of information not only helps supermarkets accurately predict sales trends but also enables them to formulate more targeted restocking and pricing strategies, thereby enhancing their competitiveness and operational efficiency in the market.

4. Conclusion

In conclusion, the comprehensive analysis and modeling efforts underscore the critical role of data-driven strategies in optimizing sales and pricing decisions for vegetable categories within the supermarket industry. By employing regression analysis, the study effectively mapped the relationship between sales volume and cost-plus pricing, facilitating strategic adjustments to pricing schemes. The use of time series models for forecasting replenishment volumes and pricing strategies, grounded in an understanding of cyclical and seasonal trends, further exemplifies the approach's precision. The novel application of a regularization-based PSO algorithm for profit maximization, adept at navigating the challenges posed by constraints, showcases the capability to predict sales volumes and pricing decisions accurately. This research highlights the necessity of agile and informed decision-making in response to market demands, supply chain dynamics, and other environmental factors, ultimately enabling more profitable and efficient operations in the competitive landscape of supermarket vegetable sales.

References

[1] PEREZ-MESA J C, PIEDRA-MUNOZ L, GALDEANO-GOMEZ E, et al. Management strategies and collaborative relationships for sustainability in the agrifood supply chain [J]. Sustainability, 2021, 13 (2): 749.

[2] TURAN C, OZTURKOGLU Y. A conceptual framework model for an effective cold food chain management in a sustainability environment [J]. Journal of Modelling in Management, 2022, 17 (4): 1262 - 1279.

[3] LIU X, Le BOURVELLEC C, YU J, et al. Trends and challenges on fruit and vegetable processing: Insights into sustainable, traceable, precise, healthy, intelligent, personalized and local innovative food products [J]. Trends in Food Science & Technology, 2022,125: 12 - 25.

[4] HU J, ZHANG X, CHEN H Y, et al. When it rains, it pours? The impact of weather on customer returns in the brick-and-mortar retail store [J]. Journal of Retailing and Consumer Services, 2024, 77: 103664.

[5] MALIK A, LISCHUK Y, HENDERSON T, et al. Generating Constant Screw Axis Trajectories with Quintic Time Scaling for End-Effector Using Artificial Neural Network and Machine Learning: 2021 IEEE Conference on Control Technology and Applications (CCTA) [C]: IEEE, 2021.

[6] HODSON T O. Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not [J]. Geoscientific Model Development Discussions, 2022, 2022: 1 - 10.

[7] BAZDARIC K, SVERKO D, SALARIC I, et al. The ABC of linear regression analysis: What every author and editor should know [J]. European science editing, 2021, 47.

[8] NHU V, SHIRZADI A, SHAHABI H, et al. Shallow landslide susceptibility mapping: A comparison between logistic model tree, logistic regression, naïve Bayes tree, artificial neural network, and support vector machine algorithms [J]. International journal of environmental research and public health, 2020, 17 (8): 2749.

[9] HOLMSTROM K, PETERSSON J. A review of the parameter estimation problem of fitting positive exponential sums to empirical data [J]. Applied Mathematics and Computation, 2002, 126 (1): 31 - 61.

[10] JAIN M, SAIHJPAL V, SINGH N, et al. An overview of variants and advancements of PSO algorithm [J]. Applied Sciences, 2022, 12 (17): 8392.