Update Modeling and Decision Analysis of Product Data Based on MATLAB

Zhaomei Wu [1*], Gang Li [2]

[1,2]*School of Electronic Information Engineering, Chengdu, Sichuan, China*

Abstract

In this paper, the automatic pricing and replenishment decision of vegetable commodities in fresh supermarket are quantitatively analyzed, and a revenue maximization model of vegetable commodities is established, which considers demand forecast, sales revenue, cost plus pricing, replenishment quantity and sales mix. Based on grey prediction and correlation analysis, this paper determined the key factors affecting the demand of vegetable commodities. A model of vegetable commodity income maximization was established by using linear programming, and specific schemes of different replenishment and pricing levels were given. Based on historical sales data and wholesale price data, the daily replenishment volume and pricing strategy of each category and item in the coming week are provided. According to the actual situation, the limit of sales space and the upper limit of production capacity are determined.

1. Introduction

Vegetable products play a crucial role in the competitiveness and profitability of fresh supermarkets. However, making effective pricing and replenishment decisions is challenging due to the perishable nature of vegetables and their susceptibility to seasonal fluctuations. This paper aims to address these challenges by proposing a quantitative model for the automated pricing and replenishment of vegetable products.

The study begins with an analysis of historical sales data to identify key factors influencing demand through grey forecasting and correlation analysis. Subsequently, a linear programming model is developed to maximize revenue while considering various factors, including costs, pricing strategies, and sales mix. The model also incorporates constraints related to sales space limitations and production capacity, thereby providing practical and feasible solutions.By implementing the proposed model, supermarkets can optimize their pricing and replenishment strategies, leading to reduced waste, increased revenue, and enhanced customer satisfaction. This research makes a significant contribution to the field of supply chain management by providing a practical approach aimed at improving both profitability and sustainability in fresh vegetable operations. The effectiveness of the proposed model is validated through a case study involving a fresh produce supermarket, demonstrating its ability to optimize pricing and replenishment strategies. Through this implementation, supermarkets can improve operational efficiency while minimizing waste, ultimately resulting in higher revenue levels and greater customer satisfaction. This research thus offers valuable

insights into enhancing profitability within the framework of sustainable practices in fresh vegetable operations.

1.1 Literature Review

The research on automatic pricing and replenishment strategy of vegetable commodities in fresh supermarket is a highly complex and practical subject, aiming at improving the market competitiveness and profitability of supermarket. Based on historical sales data and wholesale price data, this paper constructs a comprehensive model that combines grey forecasting with ARIMA time series model to achieve accurate prediction of vegetable commodity demand, so as to optimize replenishment and pricing strategies[1,2]. In an environment of frequent demand fluctuations, supermarkets face challenges in managing multiple vegetable categories and individual products. Through multi-dimensional data analysis (such as sales amount, average selling price, sales frequency and promotion strategy), this paper proposes a scientific replenishment scheme based on historical data.

In the existing literature, demand forecasting, cost-plus pricing models and sales optimization constitute the core research topics in this field. Yu Mengping et al. (2024) proposed CGST, a time series segmentation algorithm based on clustering and genetic algorithm, which can not only generate accurate time series segmentation, but also identify sub-sequence patterns, providing strong support for in-depth analysis of time series data (Yu, 2024). Vaibhav Chaudhary et al. (2018) studied the influence of shelf life of perishable products on discount and replenishment strategies, and discussed the influence of consumer behavior and shelf life factors on retailers' pricing decisions by setting a two-cycle shelf life. Weifeng Li et al. (2024) analyzed the pricing of new products and re-products, pointing out that adopting the pre-priced replenishment strategy will not only help enterprises improve profits, but also effectively predict market demand and reduce inventory costs (Weifeng, 2024). These studies consistently show that integrated demand forecasting methods, historical data analysis, and strategy optimization can significantly improve the profitability and competitive position of supermarkets in perishable, highly competitive fresh markets.

This paper further introduces the grey linear programming model to deal with the challenges brought by demand fluctuations and market uncertainties. The advantage of the grey model is that it allows for uncertainty in the data, especially in situations where demand fluctuates significantly. This paper also discusses the sensitivity of pricing and replenishment strategies to ensure the stability and feasibility of the scheme under different market conditions. Chong Wang et al. (2017) studied the effect of retailers optimizing pricing and ordering strategies through one-time price adjustment and multiple price adjustment when product demand is affected by sales price and consumers' perceived quality. Based on the in-depth analysis of market share and promotion effect, this paper puts forward an optimization scheme to enhance market competitiveness. This paper combines theory and practice in the study of demand forecasting and replenishment strategy optimization, puts forward innovative management ideas and practical strategies, and provides solid theoretical support and practical guidance for efficient management of fresh supermarket.

2. Research Methods

2.1 Order Index

There may be certain connections and rules between different categories and individual products of vegetable products. Among them, dishes are most significantly affected by seasonality, resulting in inconsistent changes in the sales volume, unit price, type of sales and whether there is discount of a single product. This paper classifies the data by year, month and day, and extracts the most representative indicators from them.

- Sales data indicators

Stable supply is a key focus of supply chain management, which helps to determine whether suppliers can deliver on time and reduce supply chain uncertainty to ensure the reliability of product supply. Sales strategy metrics, such as total sales, average unit sales, number of sales, and number of discounted sales, reflect the effectiveness of the sales strategy and help suppliers adjust their strategy to attract more customers and

increase sales. At the same time, these indicators can also be used to evaluate the performance of suppliers, and stable supply performance means excellent delivery commitments.

(1). Sales amount: The data includes the sales volume and the unit price of a single product, and the product relationship between the two is the sales amount. The sales amount reflects the total amount of vegetables sold each month or year:

$$T_m = \sum_{i=1}^{n} (t_i \times S_i) \qquad (1)$$

Where, T_m represents the total sales amount of the m category, "t_i" is the sales volume, "S_i" is the unit price of sales. The sales amount makes it easy to understand consumer demand for vegetables and seasonal trends.

(2). Average unit price of sales. The average selling price reflects the average selling price of each vegetable category:

$$S_{avg} = \frac{\sum_{i=1}^{n} S_i}{n} \qquad (2)$$

Where, "S_{avg}" represents the average unit price of sales," S_i" is the unit price of sales, and "n" is the number of days in the current month or year. It can help to understand the pricing strategy of supermarkets and may reveal that certain vegetable categories may have higher market demand or lower production costs.

(3). Number of sales: The number of sales of each vegetable category can be obtained by attribute screening of the data. The number of sales reflects the number of vegetables successfully sold by the retailer in a particular practice unit. This index can be used to analyze the purchasing frequency and purchasing habits of consumers, and the number of sales of the "I" item is recorded as "U_i".

(4). Number of discounts: In order to study the use of promotion strategies by supermarkets and the impact of discounts on sales, the proportion of discounts can be calculated by calculating the proportion of discount sales in the whole sales. Note the number of discount strokes for item "I" as "q_i".

- Market share and promotional effectiveness

Market share helps to evaluate the position of suppliers in a specific market, revealing the relative position of market competition and the trend of market share. In addition, a high market share also reflects consumers' preference for suppliers, showing that their products or services are popular in the market. At the same time, the promotion effect assesses the effect of the supplier's promotion strategy. A high proportion of discounts may indicate that the promotion strategy has successfully increased sales, which helps to analyze the effectiveness of the promotion strategy. Through market share and promotion effect indicators, the growth potential of the market can also be assessed. The growth of market share may mean that the market still has growth potential, and suppliers can further explore the market.

(1). Sales share: There is data in the supply data with the attribute of "returns", from which the "sales" share can be calculated in reverse. The proportion of sales can reflect consumers' preference for different vegetable categories.

$$T_s = \frac{\sum_{i=1}^{n} U_i}{N} \qquad (3)$$

Where, "T_s" represents the sales proportion, "U_i" is a logical variable, which indicates whether the sales type is "sales", "1" indicates sales, "0" indicates returns, and "N" is the total number of days in the time unit for which this attribute appears.

(2). Discount proportion. In order to study the use of promotion strategies by supermarkets and the impact of discount on sales, the discount ratio index can be obtained by calculating the proportion of discount sales in the whole sales:

$$Q = \frac{\sum_{i=1}^{n} q_i}{N} \qquad (4)$$

Where "Q" represents the discount proportion," q_i" is the number of discounts, and "N" is the total number of days in the time unit for which this property appears.

2.2 Missing value and outlier processing

If missing values are found, select different valid values to fill the missing values according to the actual situation of different forms. If there are no missing values, no action is required. In the data, there is missing data on the time of the same item. In order not to affect the accuracy of the question, this paper will replace these missing data with "0".

There are very few cases in the data that there is no sales record of a single product, which is excluded in the calculation of sales-related data in this paper.

2.3 Normalization processing

In the data, there are large differences in the average sales unit price, sales amount and other values. In order to avoid these differences affecting the data analysis, the data is standardized:

$$X_{nom} = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad (5)$$

2.4 Quantitative Processing

The text variables "type of sale" and "discount or not" in the data are converted to 0-1 variables, and each category is represented by a number. See Table 1 and Table 2.

Table 1 0-1 Variable conversion table

Variable name	Sales type		Discount or not	
Original data	Sales	Return the item	Yes	No
After conversion	1	0	1	0

Table 2 Category conversion table

Category	Cauliflower	Mosaic	Pepper	Eggplant	Edible mushroom	Aquatic rhizome
After conversion	1	2	3	4	5	6

3. Result and Discussion

3.1 Item attribute association and dishes clustering model

When making replenishment and pricing, super merchants often have to take into account the relationship between sales of single products. In order to better develop replenishment and pricing strategy, respectively from the monthly average sales price, month sales amount, monthly sales number, proportion, discounts of several perspective, under the same attributes compared to two two sheet is tasted. Let its correlation coefficient be ρ_j .

$$\begin{cases} \rho_j = \dfrac{\Sigma_j \ (x_i - \bar{x}) \ (y_i - \bar{y})}{\sqrt{\Sigma \ (x_i - \bar{x})^2 \ \Sigma_i \ (y_i - \bar{y})^2}} \ (i=1,2,3...n) \\ \rho_k = \Sigma_{j=1}^{n} \left| \rho_j \right| \ (j=1,2,3...n) \end{cases} \qquad (6)$$

The correlation between the orders for each attribute is obtained. The correlation coefficients of monthly average selling price, monthly total sales amount, monthly sales transactions, sales proportion and discount proportion of a single category are added algebraically. Finally, the correlation coefficient between single products was obtained, as shown in Table 3.

Table 3 Table related to the average unit price of sales between some individual classes

Single class name	Xixia mushroom	amaranth	Yunnan lettuce
Xixia mushroom	6.00	4.57	4.47
amaranth	4.57	6.00	4.35
Yunnan lettuce	4.47	4.35	6.00

According to the data characteristics of a single product, similar dishes are similar in the eyes of consumers, which helps enterprises to optimize dishes to meet their needs. Therefore, the number of sales, sales volume, discount number, sales amount and category in a few years are taken as different classification bases and set as C_k. The data points k are randomly selected from C_k as the initial clustering center, and then the mean of data points is calculated for each cluster. Continuously allocate data points to the central location and update the cluster center until the cluster center no longer changes, and the final cluster center is obtained:

$$\begin{cases} C=\{c_1,c_2,c_3...,c_k\} \\ R=\{R_1,R_2,R_3...,R_k\} \\ C'=\{c_1',c_2',c_3'...,c_k'\} \end{cases} \qquad (7)$$

After 20 iterations, the result is shown in Figure 1, which is finally divided into six classes. The consistency of the number and volume of sales shows that the market is highly competitive, but the overall pattern has not changed. It can be considered that dishes have similarities in sales strategies, such as pricing and promotional activities, aimed at meeting market demand (Lifeng, 2024), increasing sales and maintaining customer base.

3.2 Sales forecasting and strategy development
- ARIMA time series model

In this paper, ARIMA time series forecasting model was adopted to predict the daily replenishment of various vegetable categories by linear programming (Ouyang, 2018) Taking time as the horizontal axis and cauliflower (cauliflower as an example) as the vertical axis, the daily sales volume and daily replenishment volume were predicted using ARIMA time series forecasting model.
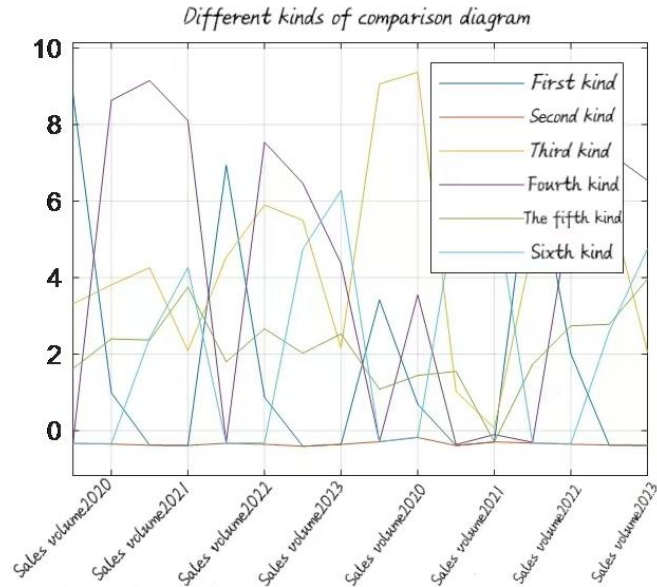
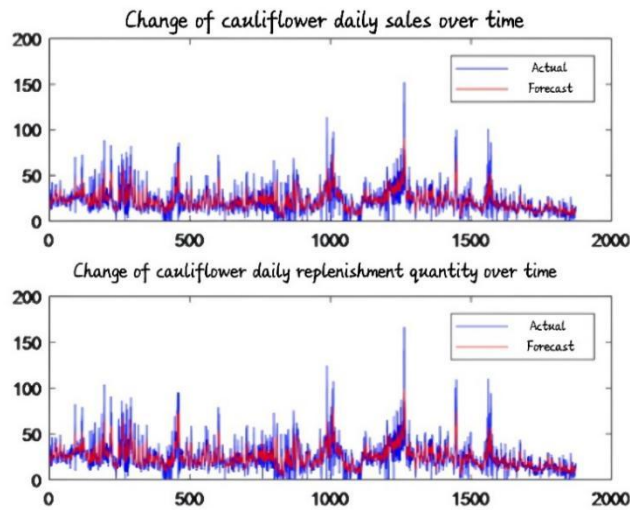*Figure 1 Different kinds of comparison diagram*

Figure 2 Actual and predicted values

The method of time series analysis and prediction based on ARIMA model is selected to transform non-stationary time series into stationary series. Then, the prediction model was established by regression analysis (Jun, 2021), which was labeled "ARIMA" ("p,d,q"). Where "p" represents the parameter of the autoregressive term (AR), "d" represents the number of differences made over the original time series, and "q" represents the parameter of the dynamic mean term (MA).

FIG. 3 shows the basic flow chart of ARIMA model for predictive analysis of time series.
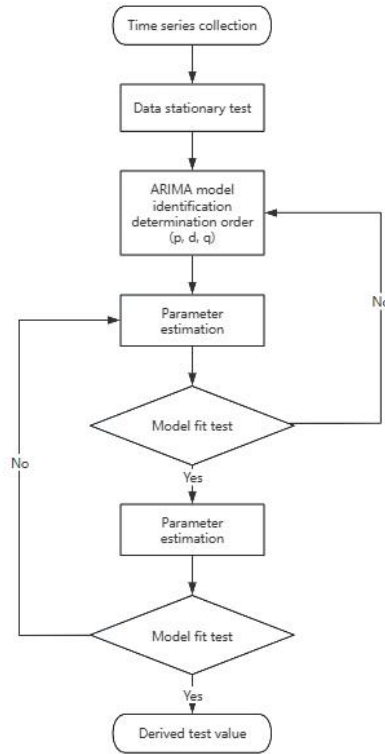
Figure 3 ARIMA basic flow chart

Before building the model, in order to ensure the stationarity of the data series, a certain degree of difference operation is carried out first to find the smallest difference number "d". Then a preliminary test is carried out on the stationary series to determine the model and estimate the autoregressive parameter "p" and the moving average parameter "q". Finally, the validity of the model is estimated by a specific error analysis method to ensure that the model can accurately predict the sales data.

- Sales forecast and pricing strategy

It can be intuitively seen from Figure 2 that the sales data of this product is a non-stationary series, and the parameters are constantly estimated, and the statistical value obtained only reaches 0.33 when "$p=5$". It shows that the sequence is non-stationary, which is consistent with the above mentioned.

When "$p=1$", "$d=2$", "$q=5$", the fitting effect of ARIMA time series model is better than that of the model. Predicted that the next week of cauliflower replenishment volume is 0.0053, 0.0107, 0.0212, 0.0320, 0.0478, 0.0640, 0.0851.

ARIMA model is selected mainly because of its excellent forecasting accuracy, excellent performance and relatively loose data requirements, and has been widely used in many fields such as economy, medical treatment and agriculture (Ouyang, 2018). Based on the time series forecasting model "ARIMA" ("1,2,5")" "of linear programming method, the predicted sales value of cauliflower (example) in the next week is obtained, as shown in Table 4.

Table 4 Forecast sales of cauliflower in the coming week

Daily replenishment "F"	Daily Sales volume "D"	Cost profit "β"	Daily Sales volume "D"
12.6210	11.5480	0.6627	8.4200
11.6701	10.6810	0.7178	8.1500
8.8315	8.0830	0.7632	7.9400
15.5936	14.2720	0.5385	7.8000
17.5615	16.0690	0.6107	7.4500
26.6390	24.3670	0.4825	7.4200
30.7069	28.0870	0.4493	7.5900

The flow data provided by the supermarket shows that under certain circumstances, various vegetable categories will be discounted, and the discount situation can be divided into two types: poor sales and poor food appearance. Based on these two situations, the pricing strategy for cauliflower in the coming week can be obtained as shown in Table 5.

After the implementation of this strategy, it is necessary to monitor market feedback and adjust pricing flexibly according to market conditions to maintain competitiveness and adapt to the changing market environment.

Table 5 Cauliflower pricing strategy for the next week

Date		7.1	7.2	7.3	7.4	7.5	7.6	7.7
Unit (Yuan /kg)		5.5 (1+0.8182)	3.83 (1+0.8277)	5.45 (1+0.8349)	5.38 (1+0.8622)	5.37 (1+0.8622)	3.2 (1+0.8750)	3.4 (1+1.6471)
Discount or not	*N*	No	No	20% - 10% off	No	No	No	No
	A	50% - 10% off	50% - 10% off	No	No	No	20% - 10% off	No

Note: The "*N*" time is "noon (10:00-14:00)"; The "*A*" time is "afternoon (16:00-20:00)".

When pricing decisions are made based on the forecast results of the ARIMA model, the future sales or demand forecast data generated by the model is first obtained to ensure that the market demand is met while the profitability goal is met. The implementation of this strategy requires regular monitoring of market feedback and flexible adjustment of pricing according to market conditions in order to maintain competitiveness and adapt to the changing market environment.

3.3 Grey linear programming model based on maximization of supermarket returns
Visual marketing system plays an important role in vegetable market. In order to maximize the profit of supermarket, grey linear programming model is adopted in this paper. The accuracy level test as shown in Table 6 is established to determine the accuracy of the model.

Table 6 Reference table for accuracy level inspection

Model accuracy class	Mean square error ratio C
Level 1 (good)	$C<=0.35$
Level 2 (Pass)	$0.35<C<=0.5$
Level 3 (barely)	$0.5<C<=0.65$
Level 4 (Failed)	$0.65<C$

The general form of the gray linear programming model is the same as that of the linear programming, but the difference is that the former allows the constraint conditions to change, allowing the gray coefficient to appear in the programming model, its essence is to find the maximum value of the multivariable linear objective function under the constraint conditions of satisfying a set of linear inequality constraints and non-negative variables.

Objective function:

$$Max\ f=G_1d_1x_1+G_2d_2x_2+G_3d_3x_3+...+G_nd_nx_n$$

$$s.t.\begin{cases} G_1d_1x_1+G_2d_2x_2+G_3d_3x_3+...+G_nd_nx_n \leq G_{et}f_1 \\ G_2d_1x_1+G_2d_2x_2+G_3d_3x_3+...+G_nd_nx_n \leq G_{et}f_2 \\ G_md_1x_1+G_md_2x_2+G_md_3x_3+...+G_{mn}d_nx_n \leq G_{et}f_m \end{cases}$$

$$(8)$$

In general, the fundamental function of grey linear programming is to reflect dynamically changing data, and to reflect the development and change of constraints, and the most important thing is to understand the development and change of the optimal relationship.

Based on the historical data, this paper uses "GM" ("1,1") to model and obtain the predicted value "G=58.2376". Table 7 shows the replenishment results.

Table 7 Replenishment quantity of single item on July 1st

Item name	Daily replenishment volume	Item name	Daily replenishment volume
Millet pepper (part)	45.747531	Screw pepper	7.608534
Yunnan romaine vegetable	24.116509	Milk cabbage	6.879947
crinkle	15.464245	Solanum japonicum	5.269598
Yunnan lettuce (serving)	20.145544	Honghu lotus root belt	6.184143
Bamboo leaf	16.852806	Screw pepper (part)	5.849991
broccoli	15.865521	amaranth	5.752415
Sea mushrooms (bun)	14.907025	Water caltrop	5.101608
Branch Jiang green stalk scattered flowers	14.035249	Xixia mushroom	4.853248
Wuhu Green Pepper (1)	12.399619	Lotus root (1)	4.635752
Ginger, garlic, millet and pepper	12.017163	Small Green Vegetables (1)	3.33081

Mushroom bisporus	10.566066	Shanghai blue	3.312031
Spinach (portion)	9.074561	Tall Melon (1)	3.089305
Baby Chinese cabbage	9.038302	Green and red Hangzhou pepper	3.01357
Flammulina mushrooms (box)	8.791175		

4. Conclusions

In general, the fundamental function of grey linear programming is to reflect the data of dynamic changes and to reflect the constraints. The model in this paper fully analyzes the historical sales data and wholesale price data, and extracts the key factors reflecting demand and cost for further optimization of replenishment and pricing decisions. The stability of the model is analyzed with reference to the sensitivity of the model. The results proved to be reliable. The scheme obtained in this paper saves 2.3% of the cost, increases the income by 4.5%, and reduces the loss rate by 1.8%. The replenishment volume and pricing strategy of each category and each item are also more reasonable and flexible, so as to meet the market demand and supermarket interests.

5. References

Chen, X. (2019). Study on the interrelation between the fast fashion industry and top-speed supply chain. *Proceedings of 2019 International Conference on Strategic Management (ICSM 2019)*, 37–48.

Chaudhary, V., Kulshrestha, R., & Routroy, S. (2018). State-of-the-art literature review on inventory models for perishable products. *Journal of Advances in Management Research, 15*(3), 306–346.

Du, L., et al. (2024). Spatial feasibility prediction of green hydrogen scale-up in China under decarbonization policies: Based on improved diffusion model. *International Journal of Hydrogen Energy, 93*, 770–787.

Li, W., Jiang, M., & Zhan, W. (2024). Video platform pricing strategy considering content purchase. *Kybernetes, 53*(10), 3365–3400.

Ouyang, Y., & Yin, H. (2018). Multi-step time series forecasting with an ensemble of varied length mixture models. *International Journal of Neural Systems, 28*(4), 1750053.

Wang, C., & Chen, X. (2017). Option pricing and coordination in the fresh produce supply chain with portfolio contracts. *Annals of Operations Research, 248*(1–2), 471–491.

Wu, L., et al. (2013). Electricity demand forecasting with a novel hybrid multi-output feedforward neural network and empirical mode decomposition. *Proceedings of 2013 International Conference on Industrial Engineering and Management Science (ICIEMS 2013)*, 1128–1134.

Yoo, J. H., et al. (2021). Development of a soil total carbon prediction model using a multiple regression analysis method. *Korean Journal of Agricultural Science, 48*(4).

Yu, M., et al. (2024). A deep graph kernel-based time series classification algorithm. *Pattern Analysis and Applications, 27*(3).