# Replenishment Strategy Model Based on Cost-plus Pricing

Xiaoyun Huang *, Ziwei Zhang

School of Statistics and Information, Shanghai University of International Business and Economics,Shanghai, China, 201620

* Corresponding Author Email: hxiaoyun2021@163.com

**Abstract.** To formulate the commodity pricing replenishment strategy to meet the short-term profit demand of merchants, we established a pricing replenishment prediction model based on the sales-cost-plus pricing correlation. It integrates four-dimensional profit factors, taking the maximization of total profit as the core guidance and considering the objective impact of date. With ARIMA and random forest algorithm, the commodity cost and sales volume are predicted from the perspective of sales cyclical fluctuation. Finally, we constructed the optimal pricing linear function based on the goal programming problem. The optimal pricing replenishment strategy can be determined by adjusting the commodity pricing. The calculation results show that the prediction accuracy of the model is higher than 0.71, which can provide merchants with pricing and replenishment strategies with high reliability.

**Keywords:** Cost-plus Pricing, ARIMA, Random Forest, Replenishment Strategy.

## 1. Introduction

Commodity pricing and replenishment strategy has always been a crucial link in enterprise operation, which directly affects profitability and customer satisfaction. Reasonable pricing can attract more consumers and increase sales. And the efficient replenishment strategy can ensure sufficient inventory and reduce the risk of shortage. In order to maximize profits, enterprises need to formulate commodity pricing and replenishment strategies that can not only improve market share, but also ensure reasonable inventory on the basis of mastering market rules, competitor dynamics and consumer demand. In China, with the rapid development of the e-commerce industry, commodity pricing and replenishment strategies are increasingly valued by enterprises. Various intelligent, data-driven pricing and replenishment solutions have also emerged.

In recent years, with the rapid development of big data and artificial intelligence technology, most domestic researchers try to optimize the replenishment decision and pricing of goods by building mathematical models. Chen Jun and Kang sha[1](2023) established a joint decision-making model of pricing and inventory replenishment considering commodity price and inventory level based on the current situation of dual channel sales of agricultural products. Zhang Xingong and Mo Ning[2](2020) used the three-parameter Weibull distribution to express the deterioration rate. Then built an inventory ordering and pricing model of perishable goods considering deterioration cost and price discount. Cui Ligang et al.[3](2023) constructed a joint replenishment and pricing model of fresh products based on fresh-keeping technology investment by introducing fresh-keeping technology investment parameters, and proposed an adaptive differential evolution algorithm to solve the model. To sum up, in recent years, most of the researches on replenishment decision and pricing of commodities by Chinese scholars have focused on commodities with short shelf life. And by using mathematical models, designing algorithms and other methods to solve the optimal pricing and replenishment decision. These studies can improve the accuracy of replenishment decision and pricing for Chinese enterprises in the actual operation. At the same time, it provides theoretical support for improving the profitability of enterprises.

For the formulation of commodity replenishment strategy based on commodity sales and sales price, it is usually based on the pricing level and sales situation of various categories of commodities in the past period. Then, combined various commodity costs (including purchase price, transportation loss, etc.) at the corresponding time with the help of relevant prediction and planning models, giving the commodity replenishment and pricing strategy for a certain period in the future. However, in the process of actual strategy formulation. While paying attention to the total sales volume, businesses will also pay attention to other price constraints such as the profit of goods. During the process of model formulating, it is usually included in the calculation in the form of profit margin. The commodity price and replenishment volume determined only based on the total sales volume may have low economic value in the actual application process.

The author proposes a multi-dimensional pricing replenishment model with profit margin. When actually making strategies, replenishment strategies are often targeted at the commodity category dimension. Therefore, the object of model analysis and prediction is commodity category. Based on the four dimensional variables of wholesale price, sales unit price, total profit and cost profit rate, this paper analyzes the relationship between the total sales volume of each commodity and pricing (taking cost-plus pricing as an example). On this basis, ARIMA and random forest algorithm are used to predict the sales volume of goods in the next week guiding supermarkets to make pricing and replenishment decisions. Finally, based on the vegetable commodity sales data provided by http://www.mcm.edu.cn, we analyzes the sales data and draws relevant conclusions.

## 2. Correlation between Total Sales and Cost-Plus Pricing

### 2.1. Unit Cost Profit Margin

Cost profit margin, which means: (sales revenue - total cost)/total cost, is calculated as follows:

$$r_i^m = (p_i^m - c_i^m)/c_i^m \tag{1}$$

Includes:

- $r_i^m$ indicates the cost profit margin of item *i* with category *m*.
- $p_i^m$ indicates the sales unit price of item *i* with category *m*.
- $c_i^m$ indicates the cost profit margin of item *i* with category *m*.

### 2.2. Total Profit Per Item

Since the same item may involve multiple sales records at different times. In order to obtain the total sales volume in daily units, we introduced the variable total daily sales volume $s_i^m$ to summarize the sales records of a single item on a single day. The daily total profit $profit_i^m$ of each item can be calculated based on the total sales volume. The calculation formula is as follows.

$$profit_i^m = s_i^m * (p_i^m - c_i^m) \tag{2}$$

### 2.3. Calculating Spearman Rank Correlation Coefficient

In order to preliminarily determine the correlation between sales volume and pricing, the Spearman correlation coefficient[4] can be introduced to analyze the change trend and correlation degree between the cost profit margin of individual products and categories and each sales price and its constraints (total sales volume, total profit and wholesale price) in the case of unknown data whether it meets the normal distribution and whether there are abnormal values.

### 2.4. Weighting Process

Based on the concepts of unit cost profit margin, total sales volume and total profit, for the determination of the relationship between the sales volume of each category and the cost-plus pricing, it is also necessary to classify and count the items under the same category. For the daily total sales volume and total profit of a category, the sum of the corresponding unit value is calculated as follows:

$$s^m = \sum_{i \in m} s_i^m \tag{3}$$

$$profit^m = \sum_{i \in m} profit_i^m \tag{4}$$

In order to make the summary effect more accurate, explain the characteristics of the data itself more effectively and improve the accuracy and reliability of the prediction model, the item with low sales volume and less sales times are weighted. Taking the sales volume as the weight reference index, the weight of each item in each category is obtained by calculating the daily sales volume proportion of it in its category, recording as normalization coefficient $a_i^m$. The calculation formula is as follows.

$$a_i^m = s_i^m / \sum_{i \in m} s_i^m \tag{5}$$

According to the normalization coefficient, the weighted wholesale price, sales unit price and cost profit margin of each category can be obtained. The calculation formula is as follows.

$$c^m = \sum_{i \in m} a_i^m * c_i^m \tag{6}$$

$$p^m = \sum_{i \in m} a_i^m * p_i^m \tag{7}$$

$$r^m = \sum_{i \in m} a_i^m * r_i^m \tag{8}$$

## 3. Daily Replenishment-Pricing Strategy model

### 3.1. Model Overview

The daily replenishment-pricing decision problem can generally be described as: for the replenishment quantity of a commodity, there are many factors affecting it. Mainly includes the wholesale price of the commodity, the sales price of the commodity in the previous days, the recent loss rate of the commodity and whether there is any special time effect (such as seasonality).

For consumers, commodity prices will directly affect commodity sales. So, the solution to this problem should be based on the relationship between total sales and pricing, considering the possible cyclical fluctuations in commodity sales (such as high sales in holidays) at the same time. And then predict commodity wholesale prices in combination with historical data.

Then, based on the wholesale price, sales price, special time impact and commodity loss rate, we can formulate the daily replenishment volume and pricing strategy. When considering the optimal strategy, the profit and profit margin should be considered together with the sales volume to maximize the interests of businesses.

### 3.2. Algorithm principle

#### 3.2.1. Auto Regressive Integrated Moving Average model (ARIMA)

ARIMA (p, d, q) [5] is a time series prediction method. This model is an autoregressive moving average model ARMA (p, q) after d-order difference stationarity operation when the studied data series do not meet the stationarity conditions. Where AR is autoregressive and p is the number of autoregressive terms. I is the difference, and d is the number (order) of differences made to make it a stationary sequence. Ma is the moving average, q is the number of moving average items [6]

The determination of the order of autoregressive terms p and moving average terms q in ARIMA(p,d,q) is based on the parameter selection of Bayesian information criterion (BIC) from the perspective of fitting, and the model with the best fitting effect is obtained under the existing data.

$$S_{BIC} = -2lnL + klnn \tag{9}$$

At the same time, it can also be judged according to the truncating/trailing values of autocorrelation function (ACF) and partial autocorrelation function (PACF). Judgment basis:

- If PACF is truncated after phase p, the order of the truncated is the parameter p determined by the model.
- If ACF is truncated after phase q, the order of truncated is the parameter q determined by the model.

### 3.2.2. Random Forest

Random forest is a bagging method in ensemble learning proposed by breiman [7] and adele [8]. Its basic principle is to use bootstrap resampling method to obtain different sample sets from the original data set, randomly select node attributes for each sample set, and select the best node attributes to split to form a basic decision tree. The final result is obtained by comprehensive analysis such as majority result method and voting method in multiple unrelated base decision trees [9].

The decision tree divides the value of the current space at each division by using the method of space hyperplane division. The specific methods are as follows:

a. Define the segmentation area(determine segmentation variable *j* and the segmentation point *s*):

$$R_1(j,s) = \{x|x_{(j)} \geq s\} \tag{10}$$

$$R_2(j,s) = \{x|x_{(j)} > s\} \tag{11}$$

b. Find the optimal segmentation variable j and the segmentation point s:

$$min\ [min \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + min \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2] \tag{12}$$

c. Output the corresponding value:

$$\widehat{c_1} = ave \qquad [y_i | x_i \in R_1(j,s)] \tag{13}$$

$$\widehat{c_2} = ave \qquad [y_i | x_i \in R_2(j,s)] \tag{14}$$

Where *x* is the input value; *y* is the target value; $\hat{c}$ is the corresponding output value.

## 4. Example Analysis

To verify the effectiveness and scientificity of the model, six kinds of vegetable commodities information of a supermarket dealer from July 1, 2020 to June 30, 2023 were selected. It includes the category of each vegetable item, the daily sales flow details and wholesale prices of each item. And the recent loss rate data of each item. There are more than 80000 pieces of valid information.

### 4.1. Correlation between Total Sales and Pricing for Different Vegetables

According to the calculated weighted price constraints and total sales volume of each vegetable category, with the help of Spearman correlation coefficient, the correlation about the change trend direction and intensity between the total sales volume and the cost-plus pricing are analyzed with multiple perspectives. The thermodynamic diagram is shown in Figure 1:

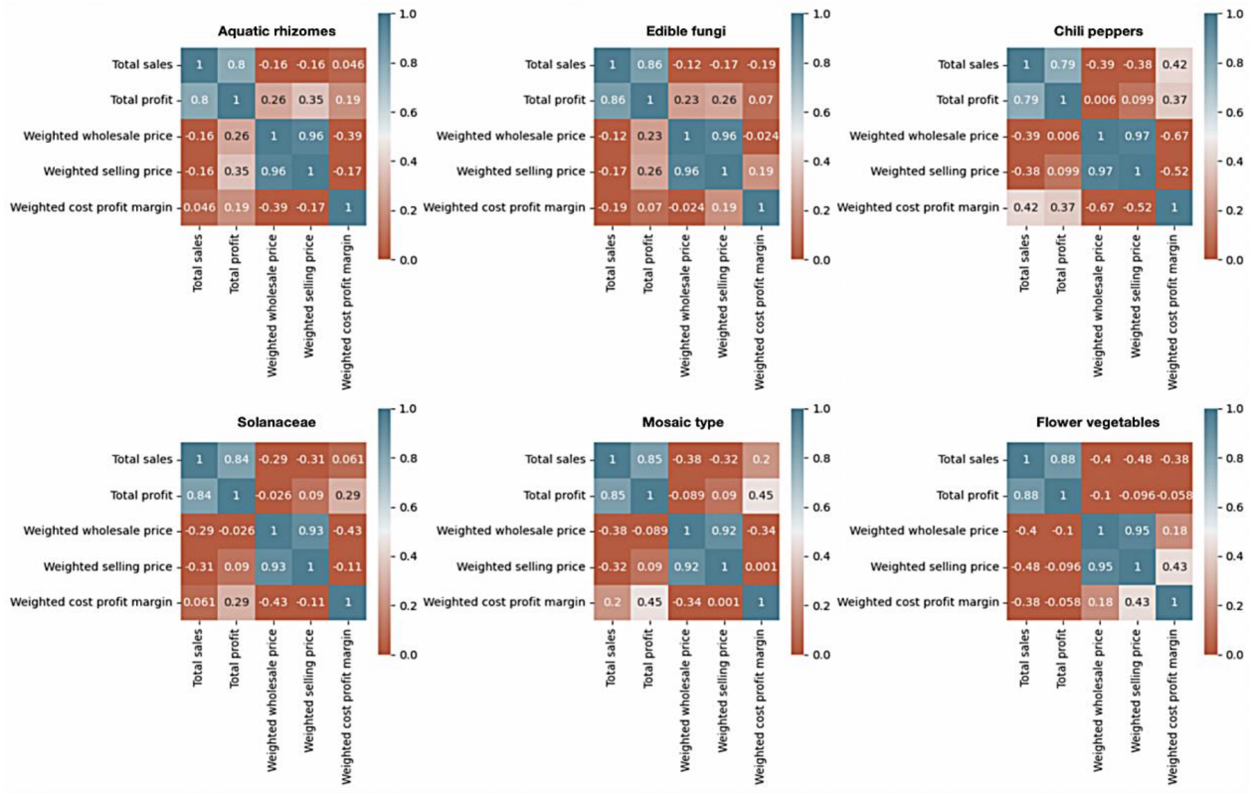


**Figure 1.** Thermodynamic Diagram for Different Vegetable Categories

According to the above figure, we have found that the correlation between cost profit margin and total sales volume was weak, and the coefficients were basically less than 0.3. At the same time, there is a certain relationship between the cost profit margin and the total profit and sales price, especially for cauliflower and pepper. The Spearman correlation coefficient between the cost profit margin and the total profit of these two kinds of vegetables is about 0.4, which is a medium degree of positive correlation. To some extent, it can be considered that higher cost profit margin can bring higher total profit. However, combined with reality, this positive correlation should have the maximum value. When the cost profit rate reaches a peak, the total profit will show a downward trend.

To sum up, for some vegetable categories, there is a positive correlation between the total sales volume and the cost plus pricing to some extent (because the improvement of cost profit margin is often accompanied by higher sales pricing). For more vegetables, there is no obvious correlation between total sales and cost plus pricing. The reason for vegetables with moderate positive correlation may be that they are commodities with less elasticity of demand. Within a certain sales pricing range, appropriately increasing the cost profit margin can improve the total profit.

### 4.2. Daily Replenishment-Pricing Strategy for Different Vegetables

Since the cost-plus pricing depends on the wholesale price of vegetables on that day, it is necessary to predict the wholesale price of various vegetables in the near future. In the above correlation analysis of different categories of vegetables, it is found that the wholesale price of each category of vegetables has some large fluctuations with the passage of time, and may have repeated fluctuations in a special time cycle, which may be a non-stationary series.

In order to predict the wholesale prices of various categories, ADF test is performed on this time series. Because there are many kinds of vegetables, we take cauliflower category as an example to predict by ARIMA model. As shown in the Table 1 below, the p value of cauliflower is less than 0.05. Therefore, the original hypothesis is rejected, and the sequence can be considered as a stationary sequence, and there is no need to carry out difference processing on the sequence.

**Table 1.** ADF Inspection

| t-statistic | P-value |
|---|---|
| -4.4876 | 0.000207 |

Because it is difficult to judge the truncated/trailing points of ACF function and PACF function of cauliflower (as shown in the Figure 2 and Figure 3 below), the p value and q value of ARIMA model are judged by AIC principle. The order to minimize the AIC value is the order of AR and ma. Finally, p is 3 and q is 2.

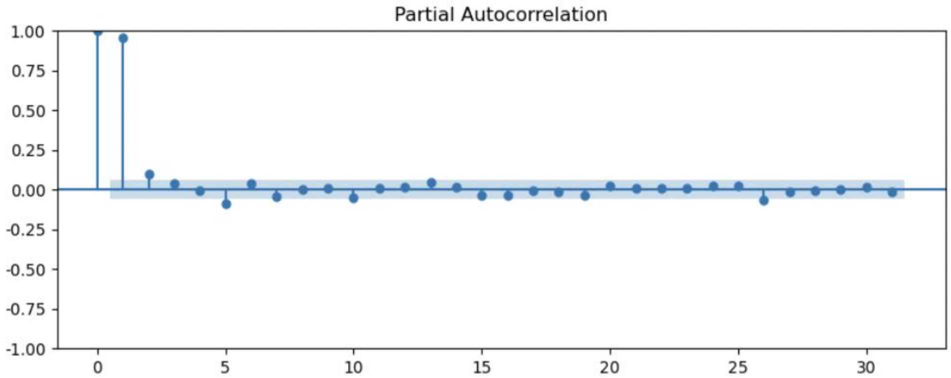


**Figure 2.** Partial Autocorrelation Diagram for Flower Vegetables

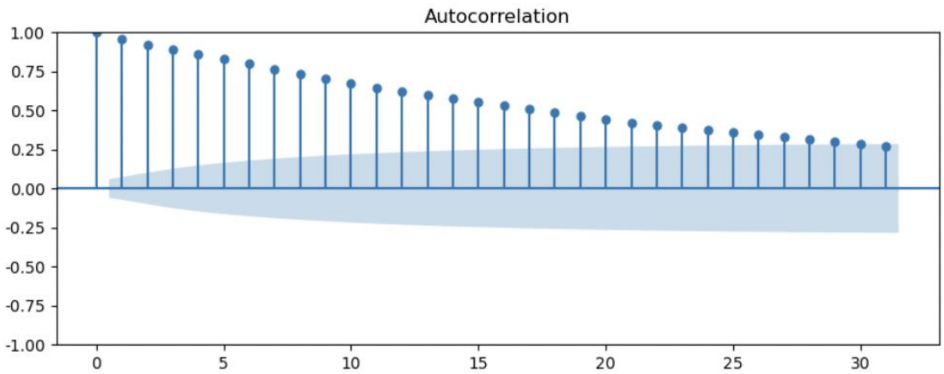


**Figure 3.** Autocorrelation Diagram for Flower Vegetables

In order to further determine the daily replenishment volume and pricing strategy, it is also necessary to further analyze the relationship between sales unit price and sales volume on the basis of correlation analysis. Because the correlation result is not significant and unique, it is not suitable for linear regression analysis. We used the method of random forest to model the multi decision tree system, while combining the wholesale price, sales price and time attribute characteristics (judgment of working days, holidays and seasons) to comprehensively predict the total daily sales of vegetables.

For the processed data, the partition ratio of test set and training set is 0.2. The optimal parameters of the model are determined by grid search method. The specific superparameter settings are shown in the Table 2.

**Table 2.** Random Forest Parameter Setting Table

| N_estimators | Criterion | Max_depth | Min_samples_leaf | Min_samples_split |
|---|---|---|---|---|
| [10,500] | Gini | [5,50] | [2,10] | [2,20] |

By comparing the predicted results with the real values, the accuracy is higher than 0.71. Therefore, through this method, the daily sales volume of vegetables under the comprehensive judgment of wholesale price, sales price and time characteristics can be more convincingly predicted. and then the replenishment volume data can be obtained by combining the loss rate of each single product.

Combined with the prediction of daily wholesale price and sales volume of each category in the coming week by the above model, the total daily sales volume of each category of vegetables under different sales unit prices can be obtained by substituting different sales unit prices. In order to formulate the maximum revenue strategy, it is necessary to define the linear optimization function [11] in combination with the goal planning problem [10] (because the wholesale price, total sales volume, sales unit price and profit are linear relationships). The specific standard calculation formula is as follows.

$$\max s_0^m * (p_0^m - c_0^m) \tag{15}$$

$$s.t.\ \ c_0^m \le p_0^m \le 3c_0^m \tag{16}$$

$$d^m s_0^m \le e^m \tag{17}$$

Where $s_0^m$ is the daily sales volume of a category predicted under comprehensive constraints. $p_0^m$ is the sales unit price. $c_0^m$ is the daily wholesale price of a category predicted by ARIMA model. $d^m$ is the loss rate of category $m$. $e^m$is the replenishment quantity of item $m$.

According to the above planning model, the replenishment pricing strategy of cauliflower category is optimized. Finally, the replenishment pricing strategy of each category of vegetables in the coming week (July 1-7, 2023) is obtained as shown in the Table 3 below.

**Table 3.** Flower Vegetables Replenishment-Pricing Strategy

| Vegetable Category | Sale Date | Weighted wholesale price | Weekday or not | Holiday or not | Season | Replenishment | Pricing | Projected Revenue |
|---|---|---|---|---|---|---|---|---|
| Cauliflower | 2023-07-01 | 7.790 | 0 | 1 | 1 | 51.200 | 12.885 | 260.865 |
| Cauliflower | 2023-07-02 | 7.819 | 0 | 1 | 1 | 51.307 | 12.829 | 257.035 |
| Cauliflower | 2023-07-03 | 7.813 | 1 | 0 | 1 | 36.198 | 12.919 | 184.843 |
| Cauliflower | 2023-07-04 | 7.799 | 1 | 0 | 1 | 34.563 | 12.899 | 176.258 |
| Cauliflower | 2023-07-05 | 7.817 | 1 | 0 | 1 | 35.795 | 12.826 | 179.284 |
| Cauliflower | 2023-07-06 | 7.806 | 1 | 0 | 1 | 36.198 | 12.909 | 184.720 |
| Cauliflower | 2023-07-07 | 7.808 | 1 | 0 | 1 | 36.198 | 12.912 | 184.759 |

## 5. Conclusion

Taking vegetables as an example, starting from the relationship between total sales and cost plus pricing, and based on the conclusion of correlation analysis, with the help of ARIMA and random forest algorithm, a daily replenishment pricing strategy model for commodity sales forecasting planning is proposed. The example analysis shows that:

1) The change trend and correlation intensity of total sales and cost plus pricing are significantly low, and some categories have a certain positive correlation.

2) Considering the daily flow data from the perspective of cyclical fluctuation of commodity sales, the time series model Arima is used to predict the wholesale price of vegetables. At the same time, the time attribute characteristics of total sales are divided from three aspects of holidays, working days and seasons, and the final prediction effect is better.

3) Considering the vegetable sales volume from the perspective of wholesale price, sales price and time attribute characteristics, combined with the optimization idea, the replenishment pricing strategy for the coming week was formulated, with an accuracy rate of more than 0.71, providing an effective

method for commodity pricing and replenishment.

## References

[1] Chen Jun, Kang Sha. Joint Decision on Pricing and Inventory Replenishment of Agri-food with Dual-channel Sales [J]. Industrial engineering, 2023, 26 (3): 39.

[2] Zhang Xingong, Mo Ning. Inventory Ordering and Pricing Strategy of Perishable Goods based on Weibull Function and Price Discount [J]. Journal of Chongqing Normal University (NATURAL SCIENCE EDITION), 2020, 37 (4): 1 – 5.

[3] Cui Ligang, Li Yali, Liu Jinxing, et al. Joint decision-making of multi product replenishment and pricing considering preservation technology investment [J]. Industrial engineering and management, 2023, 28 (03): 17 – 26.

[4] Nikitina A M, Chernukha M I. Nonparametric statistics. Part 3. Correlation coefficients [J]. Theory and practice of meat processing, 2023, 8 (3): 237 - 251.

[5] Li J. Automatic Pricing and Replenishment Decision Analysis of Vegetable Products Based on ARIMA Optimization Model [J]. Agricultural Forestry Economics and Management, 2023, 6 (3).

[6] ZHAO Z, WANG X S. Ultra-short-term multi-step wind power prediction based on ceemd and improved time series model [J]. Acta energiae solaris sinica, 2020, 41 (7): 352 - 358.

[7] BREIMAN L. Random Forests [J]. Machine Learning, 2001, 31: 106472.

[8] ADELE C, DAVID R C, JOHN R Ensemble Machine Learning [M]. Boston: Springer, 2012.

[9] Yangruijun, Zhao Nan, fan Yaofeng, et al. Urban Air Quality Assessment based on Random Forest Model [J]. computer engineering and design, 2017, 38 (11): 3151 - 3156.

[10] Jayabharathiraj J. Goal programming model for predicting the parameters involved in growth of cancer cells [J]. Int. J. of Operational Research, 2019, 34 (3): 450 - 465.

[11] Prabodhika A, Wickramarachchi A, Niwunhella D, et al.Selecting the best logistics service providers by evaluating the sustainability performance using AHP and linear programming problem [J]. International Journal of Business Performance and Supply Chain Modelling, 2022, 13 (4): 359 - 379.