Research on Multi-objective Planning Optimization Algorithm Based on Cluster Analysis

Yiwei Zhang [*,†], Zhou Zhang [†]

School of Computing and Data Science, Xiamen University Malaysia, Selangor, Malaysia, 43900

* Corresponding Author Email: zhangyiwei1002@gmail.com

† These authors contributed equally.

Abstract. AI has introduced new solutions to multi-objective decision-making in agricultural resource optimization. Still, the current methods struggle with handling complicated constraints and heterogeneous resource types effectively. This study proposes an integrated approach combining linear programming (LP) with K-means++ clustering to enhance the adaptability in diverse agricultural contexts. Specifically, the linear programming (LP) model firstly optimizes resource allocation according to real-world constraints, while K-means++ clustering groups crops with similar characteristics, selecting representative crops from each cluster, and allows simplification of crop structure and enhancing resilience. Compared to traditional K-means, K-means++ improves the stability of the initialization, ensuring more reliable crop clustering results, and largely avoids local minima. This approach has been applied to a case study in a village of Northern China, demonstrating its effectiveness. Using the LP model alone increases annual revenue from 5.66 to 6.58 million yuan annually, while combining with K-means++, revenue increases further to 7.89 million yuan, exceeding 2.23 million yuan compared to the original. This research also integrates risk modeling, simulating disruptions from the environment and market, demonstrating strong robustness in maintaining revenue stability in different scenarios. This hybrid model not only improves profitability but also provides a scalable and robust strategy for sustainable agricultural planning.

Keywords: Linear programming, K-means++, Multi-objective optimization, Robustness.

1. Introduction

Nowadays, rational resource configuration and utilization are crucial aspects in modern production. Traditionally, crop decisions were experience-based, lacking systematic optimization. However, precision agriculture based on AI and data-driven technologies is a transformative power that could increase and enhance farm productivity, sustainability, and climate change [1].

There are several studies applied optimization techniques to agricultural planning. For instance, Zhou et al. [2] applied a linear programming approach to the allocation of agricultural machines for multi-task cross operations on farms. Their model includes constraints such as machine availability, task duration, and cost minimization, which proves the effectiveness of reducing machinery costs by optimizing crop and equipment pairing. Majeke and Majeke [3] applied LP in farm-level planning, with an increase in production capacity and profit. For clustering in agriculture, K-means is often used in agriculture, such as grouping crop or soil characteristics for better planning, but it still has poor initialization and local optima problems. Vardakas and Likas [4] implement the global K-means++ algorithm, which offers revised initialization and convergence properties. In addition, Polasi et al. [5] used goal programming in R to prioritize crop plans based on sustainability goals, such as maximizing yield and minimizing resource use. These studies underscore the importance of mathematical modeling and data-driven techniques in agricultural decision-making. While these methods are effective in evaluating and ranking crop strategies, they are inadequate when dealing with complex challenges such as high-dimensional clustering and dynamic optimization of resource structures. In particular, existing models largely rely on fixed crop plans and do not fully use the capabilities of advanced clustering algorithms. Despite recent advances, there is still a gap in research on combining augmented clustering methods (e.g., K-means++) with linear programming to obtain approaches that can be scaled to different situations and robustly solve agricultural optimization problems with multiple constraints and uncertainties.

This study proposes an integral optimization framework that combines LP with K-means++ clustering to cope with complex agricultural problems involving multiple decision variables and constraints. This method effectively handles high-dimensional data while maintaining good interpretability, making it suitable for a variety of agricultural contexts. To verify its effectiveness, the framework is applied to a crop cultivation case study in a rural village in Northern China. The results demonstrate that the proposed method with good practicality, which can achieve efficient optimization of resources in real production environments.

The framework also supports Climate-Smart Agriculture (CSA) principles, aiming for sustainable productivity and resilience [6]. As noted by Kakraliya et al. [7], achieving CSA goals often requires combining locally adapted strategies. In our case study, LP was first used to maximize village income while minimizing ecological impact, considering plot type, maturity, and crop characteristics. Then, K-means++ was applied to group crops and select representatives with traits such as drought and pest resistance. Compared to the LP-only model, the integrated approach achieved higher revenue and stronger robustness under several uncertainty situations.

2. Construction of Linear Programming Model under Realistic Constraints

2.1. Datasets and Information Organization

The data utilized in this paper were collected from (https://www.mcm.edu.cn/); the China Statistical Yearbook (https://www.stats.gov.cn/sj/ndsj/2024/indexeh.htm), and the China Meteorological Administration (https://www.cma.gov.cn/). To ensure accuracy, the datasets were selected based on the officiality of the database, completeness, trustworthiness, and consistency across time and regions. For the crop price section, this paper allowed the price ranges to be normally distributed and appropriately modified a small amount of data where prices were not justified based on market conditions.

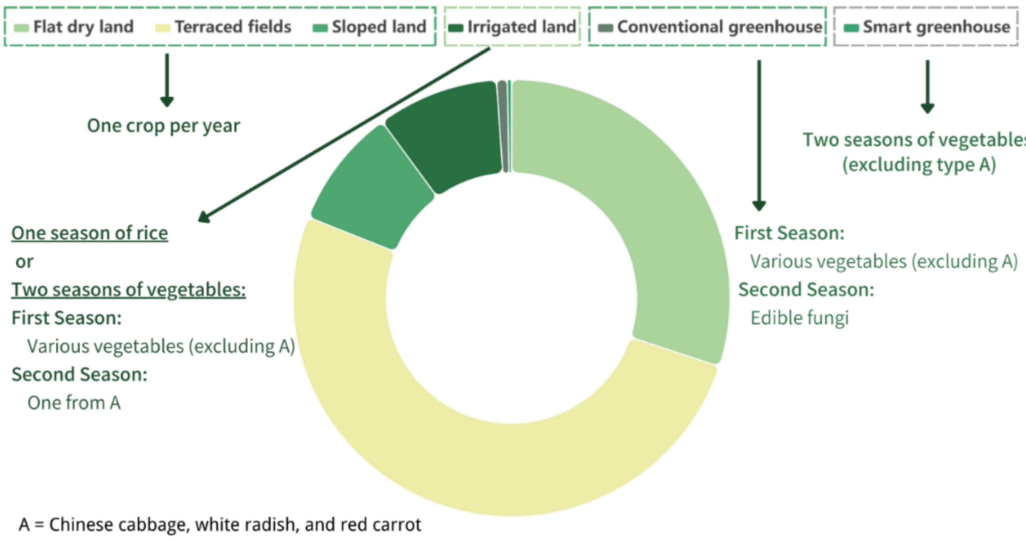

Figure 1. Cropping Regimes by Plot Type in a North China Village

In order to more clearly express the percentage of each plot type, its characteristics and crop requirements, the agricultural characteristics of Northern China and related data sets are combined, as shown in Figure 1: This figure shows the village's percentage of each plot type, crop requirements and maturity seasons [8].

As can be seen from Figure 1, flat dry land, terraced fields, and sloped land are more predominant and all have one season of crop a year, while the other three types of plots have one or two seasons a year and have different requirements for the crops grown in different seasons.

2.2. Linear Programming Model Under Multiple Constraints

Before the modern optimization technology appeared, agricultural planning primarily relied on the experience of farmers, using trial-and-error practices and historical patterns to manage land and crop cultivation. As the development of mathematical modeling and computing tools, the agricultural problem in real world, such as determining optimal crop-planting strategies over multiple fields and seasons, can now be effectively addressed by linear programming, it is observed that through the interactive conflicts in between multiple objectives a solution vector can be analyzed [2]. In this paper, a linear programming model is proposed that incorporates a range of practical constraints, including land type suitability, seasonal crop rotations, and management regulations, to maximize agricultural profitability. As demonstrated in the thesis [3], using the LP model could better utilize resources and promote the overall productivity compared with traditional methods.

To develop an optimal planting plan for the village, this study is profit-oriented, and with more constraints, a model with reliability and more stable output results should be used. Linear programming is favored for its high computational efficiency, ease of handling a large number of constraints, and high reliability, while single-objective optimization is highly respected for its clear objective, wide range of applications and ease of implementation. Therefore, in this paper, linear programming and single-objective optimization are used to maximize the gains while taking into account the various planting factors and setting the constraints appropriately.

To accurately consider the type of arable land suitable for each crop, the type and size of each plot, and to plan the planting programme more accurately, this paper sets the following area variables:

$$x_{i,j,t} \tag{1}$$

Where i is the crop number; j is the plot number; t is the time parameter, because this study considers the case of two seasons a year, for the sake of convenience of calculation, the value of t ranges from $4047 \ll t \ll 4060$, 2 seasons from 2024 to 2030, respectively. Subscripts in subsequent formulas maintain this notation setting.

To reflect practical agronomic requirements, nine constraints were formulated and grouped into the following five categories: Land suitability, Seasonality and plot rules, Crop rotation and soil maintenance, Greenhouse management, and Minimum area constraints. The overall objective is to maximize total profit across all plots and periods, while satisfying agronomic and management constraints:

$$\max I = \sum_i \sum_j \sum_t (p_i \cdot \min(Q_{i,t}, W_i) + 0.5 \cdot p_i \cdot \max(Q_{i,t} - W_i, 0) - c_i \cdot x_{i,j,t}) \tag{2}$$

Subject to:

(1) Constraints for dryland, terraces, and sloped lands $(j \in DTS)$: One crop can be planted per season:

$$\sum_{i \in C_{DTS}} x_{i,j,t} \le A_j, \forall j \in DTS\,, if\ t \equiv 1(mod2) \tag{3}$$

$$\sum_{i \in C_{DTS}} x_{i,j,t} = 0, \forall j \in DTS\,, if\ t \equiv 0(mod2) \tag{4}$$

(2) Rice-only constraint on irrigated land $(j = I)$:

$$x_{rice,j,t} \le A_j, \forall j \in I, if\ t \equiv 1(mod2) \tag{5}$$

$$x_{rice,j,t} = 0, \forall j \in I, if\ t \equiv 0(mod2) \tag{6}$$

(3) Two-season vegetable structure ($i \in C_V$), excluding type A ($i \in C_A$) in the first season:

$$\sum_{i \in C_I} x_{i,j,t} \le A_j, \forall j \in I\,, if\ t \equiv 1(mod2) \tag{7}$$

$$\sum_{i \in C_I} y_{i,j,t} \le 1,\ \forall j \in I, if\ t \equiv 1(mod2) \tag{8}$$

$$\sum_{i \in C_A} x_{i,j,t} \le A_j, \forall j \in I\,, if\ t \equiv 0(mod2) \tag{9}$$

$$\sum_{i\in C_A} y_{i,j,t} \leq 1, \forall j \in I\,, if\ t \equiv 0(mod2) \quad (10)$$

(4) Crop continuity for irrigated land: To avoid overlapping crops in seasonal sequences:

$$y_{rice,j,t} + \frac{\sum_{i\in C_I} y_{i,j,t} + \sum_{i\in C_A} y_{i,j,t+1}}{2} \leq 1, \forall j \in I\ if\ t \equiv 1(mod2) \quad (11)$$

$$y_{rice,j,t-1} + \frac{\sum_{i\in C_I} y_{i,j,t-1} + \sum_{i\in C_A} y_{i,j,t}}{2} \leq 1, \forall j \in I, if\ t \equiv 0(mod2) \quad (12)$$

(5) Conventional greenhouse ($j \in CG$):Vegetables ($i \in V$) in first, edible fungi ($i \in F$) in second:

$$\sum_{i\in C_V} x_{i,j,t} \leq A_j, \forall j \in CG\,, if\ t \equiv 1(mod2) \quad (13)$$

$$\sum_{i\in C_V} y_{i,j,t} \leq 1, \forall j \in CG\,, if\ t \equiv 1(mod2) \quad (14)$$

$$\sum_{i\in C_F} x_{i,j,t} \leq A_j, \forall j \in CG\,, if\ t \equiv 0(mod2) \quad (15)$$

$$\sum_{i\in C_F} y_{i,j,t} \leq 1, \forall j \in CG\,, if\ t \equiv 0(mod2) \quad (16)$$

(6) Smart greenhouse ($j \in SG$): Two vegetable seasons allowed, but excluding type-A crops:

$$\sum_{i\in C_V} x_{i,j,t} \leq A_j, \forall j \in SG \quad (17)$$

(7) Avoid repeated planting of crops:

$$y_{i,j,t} + y_{i,j,t+1} \leq 1, \forall i, t, \forall j \in CG, SG \quad (18)$$

$$y_{rice,j,t} + y_{rice,j,t+2} \leq 1, \forall i, t, \forall j \in I, if\ y_{rice,j,t} = 1 \quad (19)$$

$$y_{i,j,t} + y_{i,j,t+2} \leq 1, \forall i, t, \forall j \in DTS \quad (20)$$

(8) Legume frequency constraint: At least one legume must be planted every 3 years per plot:

$$\sum_t^{t+4} x_{i,j,t} \geq A_j, \forall i, j, t \quad (21)$$

$$\sum_t^{t+4} y_{i,j,t} \geq 1, \forall i, j, t \quad (22)$$

(9) Minimum planting unit constraint:

$$x_{i,j,t} \geq 0.3, if\ y_{i,j,t} = 1 \quad (23)$$

The results of the planting programme solved using linear programming for the period 2024-2030 are as follows: (Due to the huge amount of data, only the planting arrangements for the five plots before 2024 are shown in Table 1)

Table 1. Linear programming of optimal crop cultivation programmes (partial)

Season	Plot Name	Crop 1	Planting Area of Crop 1 (mu)	Crop 2	Planting Area of Crop 2 (mu)
First season	A1	Mung Bean	2.346	Wheat	77.65
First season	A2	Wheat	42	Sweet potato	12
First season	A3	Corn	35		
First season	A4	Mung Bean	23.36	Corn	48
First season	A5	Corn	68		

3. The Combination of Clustering Algorithm and LP

3.1. Brief Introduction

To address the complexity of the data, clustering algorithms are employed to reduce the solution space and improve optimization efficiency. Meanwhile, in this case, considering the substitutability and complementarity between crops also makes the clustering usage more reasonable.

In this example, K-means++ was selected as the clustering algorithm due to its enhanced performance. Compared with the classic K-means method, K-means++ is an optimal clustering method, introducing a more elaborate initial cluster center selection mechanism, and largely reduces the computational load [4]. This mechanism aims to make sure that the selected initial cluster centers are maximally dispersed across the dataset by intelligently considering the distances between data points. This strategy also largely reduces the risk of the algorithm falling into the local optimal trap. Therefore, it enhances the stability and accuracy of clustering results.

3.2. Process of K-means++

Firstly, the dataset DATA is established based on attributes such as crop type, total crop value, crop selling price, plot type and crop maturity. A random sample point is picked up initially as the first cluster center. Subsequently, the shortest distance from each sample point to the nearest cluster center is calculated and denoted as dis_i. Based on these distances, each sample point is assigned a probability value P(b) of being selected as the next cluster center. Finally, a sample point is selected according to the probability distribution to be the next cluster center.

$$P(b) = \frac{dis_i(b)^2}{\sum_{b \in DATA} dis_i(b)^2} \tag{24}$$

On this basis, repeat the previous steps until K cluster centers have been identified, and calculate the Euclidean distance between each sample and all cluster centers.

$$D_{il} = \sqrt{\sum_{k=1}^{n} (b_{ik} - K_{lk})^2} \qquad l = 1,2,\dots,K; i = 1,2,\dots,m \tag{25}$$

In total, there are m samples and n indicators. After distance comparison, each sample point is categorized into the category belonging to the cluster center with which it has the minimum distance. After an iterative process, this research recalculated and updated the position of the cluster center based on the currently classified sample points, and the new cluster center is as follows:

$$K_l' = \frac{1}{m'} \sum_{b_i \in K_l} b_i \tag{26}$$

This process of recalculating distances, reassigning samples, and updating cluster centers continues iteratively until the cluster centers converge and no longer change significantly, which means that the clustering process is considered complete.

3.3. Result of K-means++

As shown in Table 2, crops were clustered into four categories: grains (excluding rice), rice, vegetables (excluding type A), and a group including Chinese cabbage, white radish, carrot, and edible mushrooms. The representative crops closest to each cluster center are oat, rice, cucumber, and white mushroom, respectively. To simplify real-world planting, only these four representatives and all legume crops (both grains and vegetables) were selected for modeling, reducing the complexity of the linear programming formulation.

Table 2. Crop clustering results using the K-means++ algorithm

Crop ID	Crop Name	Cluster	Crop ID	Crop Name	Cluster	Crop ID	Crop Name	Cluster
1	Soy bean	2	15	Turnip	2	29	Cucumber	1
2	Black bean	2	16	Rice	0	30	Lettuce	1
3	Red Bean	2	17	Broad Bean	1	31	Chili Pepper	1
4	Green Bean	2	18	Sword Bean	1	32	Water Spinach	1
5	Creeping Bean	2	19	Hyacinth Bean	1	33	Yellow Heart Cabbage	1
6	Wheat	2	20	Potato	1	34	Guangcai	1
7	Corn	2	21	Tomato	1	35	Cabbage	3
8	Millet	2	22	Eggplant	1	36	White Radish	3
9	Sorghum	2	23	Spinach	1	37	Red Carrot	3
10	Sweet Corn	2	24	Green Pepper	1	38	Crown Daisy	3
11	Barley	2	25	Mustard Greens	1	39	Amarant	3
12	Pumpkin	2	26	Cabbage	1	40	White Mushroom	3
13	Sweet Potato	2	27	Celtuce	1	41	Pleurotus Eryngii	3
14	Oat	2	28	Rape	1			

3.4. Integration of Clustering Results into LP Model

After applying k-means++ clustering, the previous LP model is revised based on the clustering results: This study selected representative crops from each cluster (referring to Section 3.3), so as to simplify the structure of crops. This study retained the same LP model and constraints as introduced in Section 2.2. The only difference is the input dataset, which only uses the reduced crop categories from clustering results. This adjustment allows the model to retain its original constraint structure while improving computational efficiency and robustness in crop selection under uncertainty.

4. Economic Variable Modeling and Profit Optimization

To evaluate the performance of strategies under uncertainty, this study separately modelled 2 common risks: (1) natural disruptions such as pests and drought, and (2) market price fluctuations affecting crop revenue. [9].

4.1. The establishment of the simulation model

The frequency and intensity of droughts and pest disasters extremes in northern China have increased as global average temperatures have risen, and are expected to increase in the future, posing significant risks to several sectors, including agriculture, compound dry and hot events or extremes (CDHEs) can cause reduced crop yield, posing a threat to food security [9], Therefore, this study places particular emphasis on simulating their impacts. Assuming that there is a 40% probability of large-scale pests happening in this area in a random year [10], leading to a 20% reduction in grain legume production compared to the previous year. Similarly, drought risk was considered in the model with a 70% occurrence in any year between 2024 and 2030 [11]. As vegetables are highly sensitive to drought, this disease will cause a 30% decrease in production, while the damage is particularly severe if planted on watered land, where vegetable production can be reduced by 80%.

As shown in Figure 2 and 3, without clustering, the model has average revenue of 6.58 million yuan annually, but dropped below 5 million yuan in some years, and this research speculate that pests and drought diseases happened in these two years, showing that this model is vulnerable while facing production perturbation; After the introducing K-means++ clustering, optimized crop structures led to a more stable revenue stream, with average revenue rising to 7.89 million yuan, and for all the years in research, it maintained above 5 million yuan, with a strongly enhancement of stability.

4.2. Price Volatility Risk Simulation

To simulate market risk, crop prices were allowed to vary by ±10% annually. As shown in Figure 2, without clustering, the model is more sensitive to price fluctuations, and even though the average annual return is higher than the traditional method, the difference in return between years is larger, indicating that the stability of the return is relatively low. After optimizing with clustering shown in

Figure 3, the model shows higher robustness in the face of price uncertainty, with less fluctuation of the return change from year to year. The average annual return remains at a high level, and the fluctuation of return changes from year to year is small.

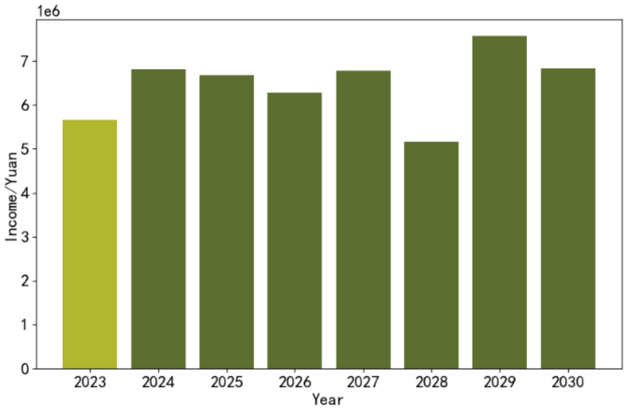

Figure 2. Annual Revenue Using LP

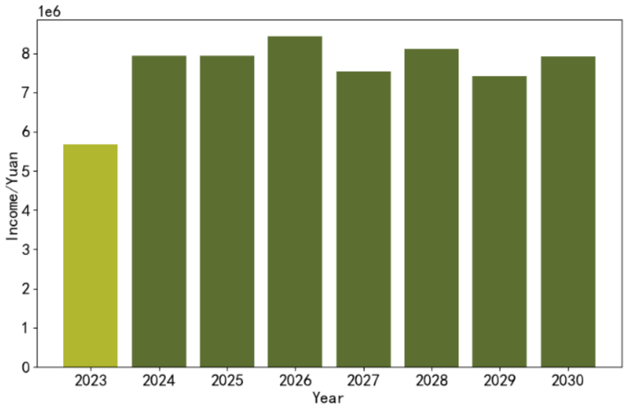

Figure 3. Annual Revenue Using LP & K-means++

Overall, natural diseases and price fluctuations act as two distinct and independent sources of risk that can affect agricultural production severely. But by introducing K-means++ clustering combined with linear programming, it becomes possible to optimize crop planting structures so as to reduce the adverse effects of both risks, enhancing the stability of agricultural revenue.

5. Conclusion

This paper proposed an optimization framework that combines linear programming and K-means++ clustering. Considering a case of a village in northern China, this framework aimed to improve the economic efficiency and stability of agricultural systems under multiple constraints and uncertainties. The models fully considered the type of land, seasonal planting structure and risks of natural disaster (such as drought and pest diseases). Compared with a traditional method with annual revenue 5.66 million yuan, it could get the average income 6.58 million yuan per year while using LP increasing; based on this LP method, this study introduces K-means++ cluster for optimization of crop structure, achieving the average annual revenue to 7.89 million yuan, with 2.23 million yuan raising. However, the current model still has some restrictions, like the crop production settings being more certain and the clustering criteria being relatively simplified. The model performance can be further improved by introducing climate change simulations or adaptive learning algorithms in the future.

References

[1] Zhao J, Liu D, Huang R. A Review of Climate-Smart Agriculture: Recent Advancements, Challenges, and Future Directions [J/OL]. Sustainability, 2023, 15 (4): 3404.

[2] Zhou X, Sharma A, Mohindru V. Research on linear programming algorithm for mathematical model of agricultural machinery allocation [J]. International Journal of Agricultural and Environmental Information Systems (IJAEIS), 2021, 12 (3): 1-12.

[3] Majeke F, Majeke J. A farm resource allocation problem: case study of small scale commercial farmers in Zimbabwe [J]. Journal of Sustainable Development in Africa, 2010, 12 (2): 315-320.

[4] Vardakas G, Likas A. Global k-means++: an effective relaxation of the global k-means clustering algorithm [J]. Applied Intelligence, 2024, 54 (19): 8876-8888.

[5] Alotaibi A, Nadeem F. A review of applications of linear programming to optimize agricultural solutions [J]. International Journal of Information Engineering and Electronic Business, 2021, 15 (2): 11.

[6] MIZIK T. Climate-Smart Agriculture on Small-Scale Farms: A Systematic Literature Review [J]. Agronomy, 2021, 11 (6): 1096.

[7] Kakraliya S K, Jat H S, Singh I, et al. Performance of portfolios of climate smart agriculture practices in a rice-wheat system of western Indo-Gangetic plains [J]. Agricultural Water Management, 2018, 202: 122-133.

[8] Liao X, Wang M, Tang Y. Research on the Temporal and Spatial Distribution of Crop Planting Structures in Mountainous Areas of North China [J]. Advances in Applied Mathematics, 2025, 14: 418.

[9] Zhang Y, Hao Z, Zhang Y. Agricultural risk assessment of compound dry and hot events in China [J]. Agricultural Water Management, 2023, 277: 108128.

[10] Wang S. Research on Problems and Countermeasures in the Process of Soybean Pest and Disease Control--Taking Jancheng County as an Example [J]. Digital Agriculture and Smart Farming Machinery, 2024, (06):102-105.

[11] Song C. Predicting wheat yield reduction under drought stress based on unmanned airborne hyperspectral remote sensing technology [D]. East China University of Technology, 2022.