Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 13, No. 2, 2024

Research on Optimal Crop Planting Strategy Based on Linear Programming with Sample Average Approximation

Shi Chen[1, *], Kengtai Peng[2]

[1]School of Automation, Guangdong University of Technology, Guangzhou, China
[2]School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China
[*]Corresponding author: 2470683884@qq.com

Abstract: Using linear programming, sample average approximation and other methods, this paper mainly focuses on the optimization of crop planting strategies in China, and conducts in-depth research to formulate the optimal planting scheme for the period from 2024 to 2030. Firstly, we preprocesses data on rural plot types and crop characteristics, and proposes the optimal crop planting scheme based on linear programming by considering crop sales, planting costs, mu yield, etc. Subsequently, this paper introduces uncertainties such as fluctuations in market demand and climate change, and constructs a dynamic planning model of planting scheme through the Sample Average Approximation (SAA) method. Finally, combined with the correlation analysis between crops, this paper proposes a comprehensive optimization model of planting scheme based on Spearman correlation, which effectively improves the overall return and ensures the robustness of the scheme.

Keywords: Linear Programming, Cluster Analysis, Sample Average Approximation, Correlation Analysis.

1. Introduction

With the development of agricultural markets and the increase in crop diversity, rural crop planting decisions have become more complex. Especially under such special geographic and climatic conditions as the mountainous regions of North China, a reasonable planting strategy not only needs to consider geographic and climatic conditions, but also needs to face fluctuations in market demand, crop yields and cost changes. In this study, we first pre-processed data on rural land types and crop attributes, and designed an optimal crop planting plan using linear programming methods, taking into account factors such as crop sales, planting costs and yields [1-2]. Then, considering the volatility of market demand and the uncertainty of climate, this paper used the sample average approximation (SAA) method to construct a dynamic planning model to cope with these uncertainties [3-4]. Finally, by analyzing the correlation among different crops and combining with Spearman correlation analysis, this paper proposes a comprehensive and optimized planting scheme model, aiming to improve the overall return and ensure the stability of the scheme [5-6].

2. Optimization Model of Crop Planting Scheme Based on Linear Programming

We first collected and obtained the type, area, and type of crops that can be planted in 34 plots in a village in the mountainous region of North China, as well as 2023 planting information related to each type of crop, including planting plots, planting area, acreage, and planting cost. Table 1 shows some of the data collected, including information on the land number, plot type, and acre yield of the hill village.

Table 1. Information on land and crops to be planted in 2023 in a rural area in China

Crop name	Plot type	Planting season	Yield per mu/kg	Cost/RMB	Unit price/Yuan	Planting area	Crop type	Whether legume
soya bean	semi-arid	Single season	400	400	2.50-4.50	72	Grain	yes
black soya bean	semi-arid	Single season	500	500	6.50-8.50	0	Grain	yes
Red beans	semi-arid	Single season	400	400	7.50-9.00	0	Grain	yes
Mung beans	semi-arid	Single season	350	350	6.00-8.00	68	Grain	yes
Crawler beans	semi-arid	Single season	415	415	6.00-7.50	0	Grain	yes
Wheat	semi-arid	Single season	800	800	3.00-4.00	80	Grain	no
Corn	semi-arid	Single season	1000	1000	2.50-3.50	90	Grain	no
Grain	semi-arid	Single season	400	400	6.00-7.50	55	Grain	no
Sorghum	semi-arid	Single season	630	630	5.50-6.50	0	Grain	no
Millet	semi-arid	Single season	525	525	6.50-8.50	0	Grain	no
Buckwheat	semi-arid	Single season	110	110	30.0-50.0	0	Grain	no
Pumpkin	semi-arid	Single season	3000	3000	1.00-2.00	0	Grain	no
Sweet potato	semi-arid	Single season	2200	2200	2.50-4.00	0	Grain	no
Shinola	semi-arid	Single season	420	420	5.00-6.00	0	Grain	no
Barley	semi-arid	Single season	525	525	3.00-4.00	0	Grain	no

2.1. Data preprocessing based on cluster analysis

In order to handle the large amount of complex data collected, optimize the planting scheme and simplify the decision-making process, we used the K-means clustering algorithm to classify the plots and crops. We labeled different lands and then normalized the area and serial number information before using it for cluster analysis. The land types are divided into four clusters and visualized as shown in Figure 1.

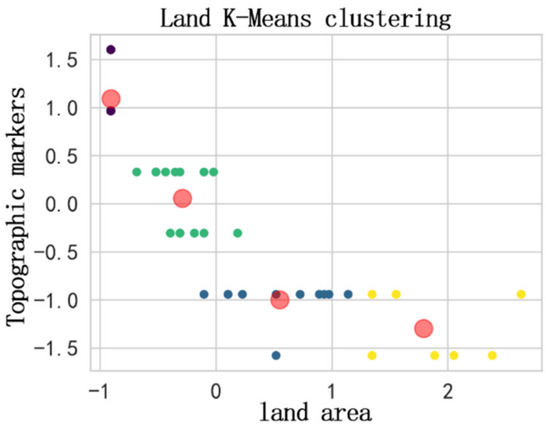

Figure 1. K-means clustering of land types

After land clustering, different plots are divided into 4 clusters, the leftmost center has the advantage of being able to grow in two seasons while having no land constraints and the disadvantage of having a smaller area, while the rightmost center has the advantage of having a large area but is affected by various aspects like land constraints, climate constraints etc. From left to right, each cluster center represents an increase in land area and constraints. Considering the above factors, we can achieve greater benefits by planning the land properly when planting.

Since there are crops characterized by three dimensions: acre yield, selling price, and planting cost, we label different crops and then perform PCA dimensionality reduction on the crop data before using it for cluster analysis. As shown in Figure 2, the crop information is divided into four clusters and visualized.

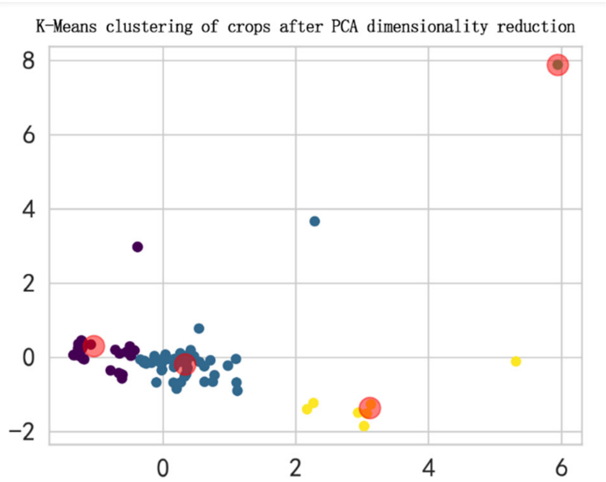

Figure 2. K-means clustering of crops based on PCA dimensionality reduction

In the figure of clustering after downscaling crops using PCA, the crops are divided into 4 categories and are more concentrated. The 4 clustering centers are defined in order from the left as high economic efficiency center, medium-high economic efficiency center, medium-low economic efficiency center, and low economic efficiency center. On this basis, planting can be considered to plant more high economic efficiency crops to increase profits, while other crops are subject to constraints to ensure that relatively more efficient crops can be planted relatively more.

2.2. Optimal crop planting scheme based on linear programming

Planning for optimal crop cultivation involves careful matching of different types of plots and crops. We categorize all plots of land into five main types: flat and dry land, terraced land, hillside land, irrigated land, and greenhouses (including ordinary greenhouses and smart greenhouses), each of which has its own specific suitability for crop cultivation. For example, flat dry land, terraced land and hillside land are suitable for growing food crops in one season a year, but not rice; watered land is flexible and can be used to grow one season of rice or two seasons of vegetables; and greenhouses, both ordinary and smart, are suitable for growing vegetables and edible mushrooms in two seasons a year (with the exception of Chinese cabbages, white radishes and carrots). In the following model, we defined decision variables and constructed objective functions and constraints based on sales volume, acreage and cost, and finally solved the optimal planting strategy by maximizing the profit from crop sales.

Define the decision variable,

$$x_{i,j,t} = \begin{cases} 1 \text{ if plot i plants is crop j in season t} \\ 0 \text{ else} \end{cases} \tag{1}$$

where $\mathrm{x}_{\mathrm{i,j,k}}$ denotes the i -th plot; $j \in \{1, \ldots, m\}$ denotes the crop; $t \in \{1, 2\}$ denotes the season, $t = 1$ is the first season and $t = 2$ is the second season.

The objective function is to maximize the total return, which consists of the revenue from the sale of the crop minus the cost of growing it:

$$\max \sum_{\mathrm{i,j,k}} \left(\mathrm{p_j} \cdot \mathrm{v_j} - \mathrm{C_j}\right) \cdot \mathrm{x_{i,j,k}} \tag{2}$$

where p_j is the unit yield of the j-th crop; v_j is the market price of the j-th crop; C_j is the cost of growing the j-th crop; $x_{i,j,k}$ is the i-th area of crop planted in the k-th season of j-th plot. In order to make the constructed linear programming model fit the specific situation of the village, we propose a complete rich set of constraints as follows:

(1) Climate and crop seasonal constraints

As the countryside is located in the mountainous region of northern China, where temperatures are low all year round, most of the arable land is limited to one crop season per year. Therefore, open cropland is constrained to a single cropping season per year, except for two seasons of vegetables that may be grown on watered land. The constraints are as follows:

$$x_{i,j,2}=0,\forall j\in\Pi_{\text{flat and dry land, terraced land, hillside land}},S_j\in\left\{S_{\text{open cropland}}\right\}\quad(3)$$

where Π denotes the type of plot and S_j denotes the area of the j-th plot, the above equation indicates that for flat dry land, terraced land, and hillside type of plots, they can only be planted in the first season and no crop can be planted in the second season.

(2) Plot type and crop suitability constraints

Depending on the type of plot, there exists a specific suitability for crop cultivation. Flat drylands, terraces, hillsides: only food crops can be grown, and rice cannot be grown. Watered land: can grow one season of rice or two seasons of vegetables, of which the first season can grow a variety of vegetables, except cabbage, white radish and carrot, and the second season can only grow cabbage, white radish or carrot.

The constraints are:

$$x_{i,j,2}=0\quad\forall j\notin\mathbb{k}_{cabbage},\mathbb{k}_{whiteradish},\mathbb{k}_{redradish},i\in\Pi_{wateredland}\quad(4)$$

where $\mathbb{k}_{cabbage},\mathbb{k}_{whiteradish},\mathbb{k}_{redradish}$ are the types of produce as cabbage, white radish, and red radish respectively.

Depending on the type of plot, there are specific restrictions on crops and seasons.

Flat drylands, terraces, hillsides

Only single-season grain crops may be grown, not rice. And only the first season's crop can be grown, not the second season:

$$\begin{cases}x_{i,j,1}=0,\forall j\in\mathbb{k}_{\text{rice}}\ \text{if}\ i\in\left\{\Pi_{\text{Flat drylands}},\Pi_{\text{terraces}},\Pi_{\text{hillsides}}\right\}\\ x_{i,j,2}=0,\forall j\ \text{if}\ i\in\left\{\Pi_{\text{Flat drylands}},\Pi_{\text{terraces}},\Pi_{\text{hillsides}}\right\}\end{cases}\quad(5)$$

Watered land

It is possible to grow a single season of rice or two seasons of vegetables.

If you choose to grow rice, you can only grow it for one season and for the first season.

$$\begin{cases}x_{i,j,2}=0\forall j\ \text{if}\ i\in\Pi_{\text{watered land}},j\in\mathbb{k}_{\text{rice}}\\ \sum_{j\in\mathbb{k}_{\text{vegetables}}}x_{i,j,1}+x_{i,j,2}=0\ \text{if}\ \sum_{j\in\mathbb{k}_{\text{rice}}}x_{i,j,1}>0\end{cases}\quad(6)$$

If you grow two seasons of vegetables, you can grow a variety of vegetables in the first season, but not cabbage, white radish or carrot, and in the second season you can only grow one of cabbage, white radish or carrot.

$$\begin{cases}x_{i,j,1}=0,\forall j\in\left\{\mathbb{k}_{\text{cabbage}},\mathbb{k}_{\text{white radish}},\mathbb{k}_{\text{carrot}}\right\}i\in\Pi_{\text{watered land}}\\ x_{i,j,2}=0,\forall j\notin\left\{\mathbb{k}_{\text{cabbage}},\mathbb{k}_{\text{white radish}},\mathbb{k}_{\text{carrot}}\right\}i\in\Pi_{\text{watered land}}\end{cases}\quad(7)$$

General shed

A wide variety of vegetables can be grown in the first season, with the exception of cabbage, white radishes, and carrots:

$$x_{i,j,i}=0\forall j\in\left\{\mathbb{k}_{\text{cabbage}},\mathbb{k}_{\text{white radish}},\mathbb{k}_{\text{carrot}}\right\},i\in\Pi_{\text{General shed}},\quad(8)$$

Only edibles can be grown in the second season:

$$x_{i,j,2}=0\forall j\notin\mathbb{k}_{\text{edibles}},i\in\Pi_{\text{General shed}},\quad(9)$$

Intelligent greenhouses

Smart greenhouses can grow two seasons of vegetables per year, but not cabbages, white radishes or carrots:

$$x_{i,j,i}=0,\forall j\in\left\{\mathbb{k}_{\text{cabbage}},\mathbb{k}_{\text{white radish}},\mathbb{k}_{\text{carrot}}\right\},i\in\Pi_{\text{Intelligent greenhouses}}\quad(10)$$

(3) Limitations of greenhouse cultivation

The growing requirements for ordinary greenhouses and smart greenhouses are different respectively. Ordinary greenhouses can grow a variety of vegetables (except cabbage, white radish and carrot) in the first season, and can only grow edible mushrooms in the second season. Edible fungi can only be planted in the fall and winter. Smart greenhouses can grow two seasons of vegetables per year, except cabbage, white radish and carrot.

The constraints are:

$$\begin{cases}x_{i,j,i}=0,\forall j\in\left\{\mathbb{k}_{\text{cabbage}},\mathbb{k}_{\text{white radish}},\mathbb{k}_{\text{carrot}}\right\},i\in\Pi_{\text{General shed}}\\ x_{i,j,2}=0\forall j\notin\mathbb{k}_{\text{edibles}},i\in\Pi_{\text{General shed}}\\ x_{i,j,i}=0,\forall j\in\left\{\mathbb{k}_{\text{cabbage}},\mathbb{k}_{\text{white radish}},\mathbb{k}_{\text{carrot}}\right\},i\in\Pi_{\text{Intelligent greenhouses}}\end{cases},\quad(11)$$

Crop rotation requirements for legumes

Legumes benefit the soil and are required to be planted on each plot at least once every three years:

$$\sum_{k=1}^{3}\sum_{j\in\mathbf{k}_{\text{legumles}}}x_{i,j,k}\geq 1\quad\forall i\in\left\{\Pi_{\text{each plot}}\right\},\quad(12)$$

This means at least one legume planting in three years within each plot.

Crop heavy cropping restrictions

The same crop cannot be grown consecutively in the same plot or shed each season to avoid yield reductions caused by heavy cropping. The constraints are:

$$x_{i,j,k}\times x_{i,j,k+1}=0\quad\forall i,j,\quad(13)$$

Indicates that the same crop cannot be grown in the same plot for two consecutive seasons.

Limitations of decentralized cultivation

For ease of management and cultivation, each crop should not be spread too thinly each season, i.e., the number of crops planted on a plot should not be too high. The constraints are:

$$\sum_i x_{i,j,k}\leq\mathbb{R}_{\max}\forall i,k.\quad(14)$$

where $\mathbb{R}_{\max}$ indicates the maximum number of crop species that can be grown per plot per season.

Minimum planting area for a single plot

The acreage of each crop in a single plot must be greater than a certain minimum , for example: This constraint ensures that the crop is not planted too small to be manageable.

$$x_{i,j,k} \geq S_{min} \forall i, j, k, \text{ if } x_{i,j,k} > 0. \quad (15)$$

Market demand constraints

The total production of a crop should not exceed the market demand, defining the market demand coefficient for a crop, the relationship between production and demand is:

$$\sum_i x_{i,j,k} \cdot p_j \leq \mathbb{Q}_j \cdot \mathbb{S}_i \forall j. \quad (16)$$

where $\mathbb{Q}_i$ is the market demand coefficient and p_j is the unit yield of crop i. To obtain the appropriate market demand coefficient, corn and soybean demand from 2021 to 2023 were analyzed and their average values were taken to obtain the projected market demand coefficient of 2.3%.

By using these constraints and objective functions to ensure that crop suitability, seasonality and crop rotation requirements are met in different plot types and seasons while maximizing the total return, we obtained the optimal planting strategy and profit obtained for each season between the three years 2024-2026. The optimal planting scheme solved by the linear programming model yields a profit in each quarter, and the total profit obtained in all quarters is ￥7851693.87.

3. A Dynamic Planning Model for Planting Schemes Based on Sample Average Approximation

In order to simulate the impact of market and climate fluctuations on planting strategies, this paper introduces the fluctuating factors of planting cost, selling price, and acreage. We optimize the planting scheme from 2024 to 2030 by dealing with multiple uncertainties through the Sample Average Approximation (SAA) method, which transforms the stochastic crop optimization problem into a deterministic problem to solve. The planting scheme is adjusted year by year through dynamic planning to ensure the robustness of the optimal crop combination.

The sample average approximation algorithm is a method for dealing with optimization problems containing random variables. It transforms the original stochastic optimization problem into a deterministic optimization problem to be solved by taking a set of samples from the probability distribution of the random variable and then using the average of these samples to approximate the expected value of the objective function or constraint in the original problem.

Define the objective function as:

$$\max \mathbb{Z} = \sum_{y=2024}^{2030} \sum_{a=1}^{N_a} \sum_{b=1}^{N_b} \left(\min\left(P_{yab}, \mathbb{T}_{yab}\right) \cdot V_{yab} - C_{yab} \cdot \mathbb{S}_{yab} \right). \quad (17)$$

where y is the year (2024~2030); p is the plot number; c is the crop number; P_{ypc} is the total production of the crop grown on the plot p in year y ; T_{ypc} is the sales volume of the crop c in year y; P_{ypc} is the sales price of the crop y in year c; C_{ypc} is the planting cost of the crop c in year y ; A_{ypc} and is the area of the c crop grown on the plot p.

Probabilistic modeling of variables such as crop prices, yields, and planting costs based on historical data is sampled from known distributions to generate samples. For each sample scenario, we compute its objective function and then construct the sample average objective function:

$$\hat{\mathbb{Z}}(\xi_i) = \frac{1}{N} \sum_{i=1}^{N} \sum_{y=2024}^{2030} \sum_{a=1}^{N_a} \sum_{b=1}^{N_b} \left(\min\left(P_{yab}, \mathbb{T}_{yab}\right) \cdot V_{yab} - C_{yab} \cdot \mathbb{S}_{yab} \right). \quad (18)$$

The main fluctuations considered are as follows:

Crop yields and fluctuations

Each crop's acreage is affected from year to year by weather, climate, and other factors, so we assumed that acreage fluctuates by ±10%:

$$P_{yab} = p_b \cdot \mathbb{S}_{yab} \cdot \left(1 + \varepsilon_y\right), \varepsilon_y \in [-0.1, 0.1]. \quad (19)$$

ε_y is the fluctuation of acreage yield in y-th year, assuming a random variable of [-0.1, 0.1].

Crop sales volumes and fluctuations

The market demand (sales volume) for each crop is also subject to some uncertainty. For food crops, it is assumed that sales volumes will grow at a rate of 5 to 10 percent per year; for other crops, sales volumes will vary within a range of ±5 percent:

$$T_{yab} = T_b \cdot \left(1 + \eta_y\right), \text{ if } b \in \mathbb{k}_{\text{food crops}} \, \eta_y \in [0.05, 0.1]; \text{ else } \eta_y \in [-0.05, 0.1] \quad (20)$$

where T_b is the initial sales volume for the 2023 crop; is the fluctuation in sales volume in year.

The sales price of vegetable crops grows at a rate of 5% per year, whereas the price of edible mushroom crops decreases at a rate of 1% to 5% per year, especially for morel mushrooms, where the price decreases by 5% per year:

$$V_{yab} = V_b \cdot \left(1 + \lambda_b\right)^{y-1}. \quad (21)$$

where V_b is the sales price of the crop in 2023; λ_b is the rate of price increase or decrease for the crop : for vegetables $\lambda_b = 0.05$, for morels $\lambda_b = -0.05$ and for other edibles $\lambda_b \in [-0.05, -0.01]$.

Crop cultivation costs and fluctuations

Annual planting costs will increase with market conditions, assuming a 5% annual increase in planting costs:

$$C_{yab} = C_b \cdot 1.05^{y-1}. \quad (22)$$

where C_b is the per-acre planting cost for the 2023 crop. The optimal acreage is found by solving the sample-averaged objective function to maximize the overall return under multiple market scenarios. We find that the solution does not change much as the sample size increases and hence the solution is more robust. Based on the profit obtained for each quarter of the dynamically planned optimal planting scenario, the total profit obtained for all quarters was￥10,648,019.72.

According to the law of large numbers, as the number of samples increases, the sample mean approximation converges to the true expectation, ensuring the validity of the SAA method.

4. A Comprehensive Optimization Model for Planting Scenarios Considering Substitutability and Complementarity

A comprehensive optimization model for planting scenarios considering substitutability and complementarity

In actual agricultural practice, certain crops have the potential to substitute for each other. This means that one crop can be grown in place of another under certain circumstances or conditions and this substitution does not significantly affect the efficiency, economic benefits or ecological balance of the agricultural output. Spearman's correlation coefficient analysis captures the relationship between crops more accurately, quantifying their complementarities and mutual exclusivities. The correlation matrix provides a scientific basis for crop planting decisions and helps optimize planting combinations to maximize returns. To assess the substitutability of different crops, we utilized the following complementarity coefficients defined below:

$$\alpha_{ij} = \max\left(0, \rho_{ij}\right). \tag{23}$$

If $\rho_{ij} > 0$, then the crops i and j are complementary and the complementarity coefficient is positive, otherwise the complementarity coefficient is 0. Similarly, we define the mutual exclusivity coefficient:

$$\beta_{ij} = \max\left(0, -\rho_{ij}\right). \tag{24}$$

If $\rho_{ij} < 0$, then the crops i and j are mutually exclusive and the coefficient of mutual exclusivity is positive, otherwise the coefficient of mutual exclusivity is 0. Combining the Spearman's correlation coefficient and the gain model discussed earlier, the final optimization model is:

$$\begin{cases} \max \mathbb{Z} = \sum_{i=1}^{n} z_i x_i + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \left(\alpha_{ij} x_i x_j - \beta_{ij} x_i x_j\right) \\ \text{subject to} \sum_{i=1}^{n} x_i \leqslant \mathbb{S}_{\text{total}}, x_i \geqslant 0 \forall i, x_i + x_j \leqslant \mathbb{S}_{\text{limit}} \ \text{if} \ \beta_{ij} > 0 \end{cases} \tag{25}$$

where z_i denotes the basic unit area yield of the crop i and S is the total available cropland area.

In practical applications, the optimal planting strategy can be derived by combining the historical data of crops and dynamically adjusting and to adapt to different planting conditions and market demands. In practice, the optimal planting strategy can be derived by combining historical crop data and dynamically adjusting and adapting to different growing conditions and market demands. The total profit gained for all seasons is ￥12,149,126.90.

5. Conclusions

In this paper, a set of optimized crop planting strategies is proposed for planting planning in the mountainous regions of North China from 2024 to 2030 through cluster analysis, linear programming, the present mean approximation method and Spearman correlation analysis. It is shown that by introducing uncertainty factors and analyzing complementarities among crops, it can effectively improve returns and maintain the robustness of planting strategies under market fluctuations.

References

[1] Wang Xiaoming, Zhang San. Research on ecological optimization of crop cultivation [J]. Agricultural Science, 2019(5): 45-50.

[2] Zhou Xue, Zhao Yongman, Wang Xiangyang. Research on the equipping of agricultural machinery of 105 regiment based on mathematical planning method[J]. Xinjiang Agricultural Mechanization,2018,(01):36-39.

[3] Li Hua. Research on agricultural optimization model based on dynamic planning [D]. Peking University, 2021.

[4] Yang Yijiang,Yang Yiwen. Dynamic planning model and its application in determining the best planting scheme for crops[J]. Guangdong Chemical Industry,2013,40(14):70-71.

[5] Xu Weichao. A review of correlation coefficient research [J]. Journal of Guangdong University of Technology,2012,(3): 12-17.

[6] Zhang Zhengjia, Chi Liang, Yang Haicheng, et al. Analysis of spatial and temporal changes in crop cropping structure in the Yellow Huaihai region[J/OL]. China Agricultural Resources and Zoning,1-10.