Scholars Journal of Physics, Mathematics and Statistics OPEN ACCESS
Abbreviated Key Title: Sch J Phys Math Stat
ISSN 2393-8056 (Print) | ISSN 2393-8064 (Online)
Journal homepage: http://saspjournals.com/sjpms

Forecasting of Corn Yeild Based on Gray-Weighted Markov Model

Song Qian-hong*, Yang Hong, Dai Dong-yan, Wang Kun

School of Arts and Science, Heilongjiang Bayi Agriculture University, Daqing163319, China

DOI: 10.21276/sjpms.2019.6.6.1 | **Received:** 05.06.2019 | **Accepted:** 15.06.2019 | **Published:** 30.06.2019

***Corresponding author:** Song Qian-hong

Abstract **Review Article**

The yield of corn was affected by a variety of factors, and show the random fluctuation characteristics, so the method of GM (1, 1) and weighted markov chain was proposed and used to predict the yield of corn. The method to predict the yield of corn in 2015 based on the data of corn yield over the years in Daqing City and its validity was verified.

Key words: GM (1, 1) weighted Markov chain；transfer matrix.

INTRODUCTION

As an important food crop, feed crop and economic crop, corn occupies an increasingly important position in China's agricultural production and national economic development. In 2012, corn surpassed rice to become the largest grain crop variety in China. It is not only an important food source for residents, but also an important raw material for feed industry, food industry, chemical industry, fuel, medicine and other industries. It can be seen that the level of corn yield directly affects the economic benefits of a series of related industries. Accurate judgment of future corn production situation and yield can provide decision basis for the government to regulate agricultural production and provide effective guidance for farmers to properly arrange corn production.

There are three methods commonly used to predict grain yield in the world: Weather yield prediction method Remote sensing technology, Statistical dynamics growth simulation method. All the three prediction methods have a lead time of about 2 months, and the prediction error is usually between 5% and 10%, and the prediction accuracy is poor[1-3]. In fact, the Weather yield prediction method only takes into account the effect of meteorological factors on grain yield under the assumption that the conditions of grain economy are stable and unchanged for a long time. In fact, grain yield is greatly influenced by factors such as market and management. In addition, weather forecasts can be difficult to predict accurately for more than one month. Remote sensing technology is applied to the prediction of grain yield, mainly using electromagnetic wave reflecting different wavelength to predict grain yield, but the prediction accuracy is not high, because the growing period of ground objects is not up to a certain stage, and the relevant information cannot be accurately captured by remote sensing technology. In addition, remote sensing technology requires a large amount of capital cost. Statistical dynamics growth simulation method is to simulate the growth principle of plants and use environmental factors, such as temperature, sunshine, CO2 concentration and other factors to affect crops. However, this method is not verified by large range of data and is only theoretically feasible in a small range.

At present, domestic forecasting methods of grain yield mainly include input-output model [4], BP neural network method [5], regression model method [6], ARMA model [7], GM (1,1) forecasting method[8] and so on.

Input-output model food production forecast accuracy is higher, but you need to consider factors such as social economy, natural conditions and production conditions, also need to consider the problem of diminishing marginal returns, not only consider the relationship between the traffic, still need to consider the relationship between the stock and flow, need to collect a large amount of data can be input and output analysis table, larger workload.

The prediction accuracy of BP neural network method is relatively high. Its advantage is that it does not need to establish the mathematical model of explanatory variable and explained variable. The disadvantage is that the level and number of hidden layers are determined and the learning rate are all set artificially, which has great randomness, time-consuming work and cannot solve the network oscillation problem caused by the network structure.

The advantage of regression model is that it can clearly understand the relationship between variables. The disadvantage of regression model is that it needs to collect more food variable factors and more sample data to improve the accuracy of prediction.

The advantage of ARMA model is that it only considers the time factor and does not consider other factors affecting the total grain output, so it is suitable for short-term prediction.

In order to improve the prediction accuracy, GM (1, 1) - weighted Markov chain prediction model is established by combining GM (1, 1) with weighted Markov chain model.

Establishment of Grey Weighted Markov Chain Model

Establishment of Grey GM (1, 1) Prediction Model

The grey system theory was put forward and developed by Professor Deng Julong in 1982. It has unique effects on the analysis and modeling of systems with short time series, less statistical data and incomplete information. Grey GM (1, 1) prediction model has been widely used in prediction as a grey system theory. The grey GM (1,1) prediction model is a prediction method that establishes a mathematical model and makes predictions through a small amount of incomplete information. Its calculation steps are as follows [9]:

Step1 Data processing: Set the original data sequence $x^{(0)}=\left\{x_1{}^{(0)},x_2{}^{(0)},\cdots,x_n{}^{(0)}\right\}$, the original data sequence is accumulated once to generate a new sequence $x^{(1)}=\left\{x_1{}^{(1)},x_2{}^{(1)},\cdots,x_n{}^{(1)}\right\}$, within $x_k{}^{(1)}=\sum_{j=1}^{k}x_j{}^{(0)},k=1,\cdots,n$;

Step2 Establishing GM (1, 1) Model: Based on the sequence $x^{(1)}=\left\{x_1{}^{(1)},x_2{}^{(1)},\cdots,x_n{}^{(1)}\right\}$, Establish grey generation model $\frac{\mathrm{d}x^{(1)}}{\mathrm{d}t}+ax^{(1)}=u$,

Within a : the development coefficient, u : control coefficient

Step3 the Least Square Method to Solve Parameters: Construct the summation matrix

$$B=B_{(n-1)\times 2}=\begin{pmatrix} -\frac{1}{2}(x_1{}^{(1)}+x_2{}^{(1)}),1 \\ -\frac{1}{2}(x_2{}^{(1)}+x_3{}^{(1)}),1 \\ \cdots \\ -\frac{1}{2}(x_{n-1}{}^{(1)}+x_n{}^{(1)}),1 \end{pmatrix},$$

Constant term vector $y_n=\left(x_2{}^{(0)},x_3{}^{(0)},\cdots,x_n{}^{(0)}\right)^T$, calculated parameters $\begin{pmatrix}\hat{a}\\ \hat{u}\end{pmatrix}=\left(B^TB\right)^{-1}B^Ty_n$;

Step4 Construct the response equation: $\hat{x}_{k+1}{}^{(1)}=\left(x_1{}^{(1)}-\frac{u}{a}\right)e^{-ak}+\frac{u}{a}$;

Step5 Cumulative reduction, get the predicted value:

$$\hat{x}_{k+1}{}^{(0)}=\hat{x}_{k+1}{}^{(1)}-\hat{x}_k{}^{(1)}=(1-e^a)\left(x_1{}^{(0)}-\frac{u}{a}\right)e^{-ak},k=1,2,\cdots,n-1,$$

let $Y(k)=\frac{x_k{}^{(0)}}{\hat{x}_k{}^{(0)}}$ be the grey fitting precision index.

Weighted markov chain method for prediction of grey fitting precision index

Markov process is a stochastic process with no aftereffect.The objects of markov chain prediction are dynamic systems with random changes.It predicts the future development of the system according to the transition probability between states. The calculations steps are as follows:

Step1 State division: Divide the data reasonably according to the actual situation,specific methods include sample mean square deviation method and cluster analysis method.The traditional mean-mean-variance classification method requires a large amount of data, which conforms to the characteristics of statistics.The data of this study are few and the category status is not clear.It is advisable to use clustering classification to establish the number of classes and classification interval.Therefore, Q-type clustering analysis is adopted for grey fitting precision indexes.It uses the sum of squares of dispersion and the Euclidean square distance to evaluate between classes, which is more comprehensive and more reasonable than the traditional classification method. Each state is represented as $E_i \in [y_{1i}, y_{2i}]$;

Step2 Establish the transition probability matrix: According to the statistics of the results obtained in Step1, the transition probability matrix of Markov chains with different delay times (step sizes) can be obtained.The formula of calculating the state transition probability matrix is $P_{ij} = N_{ij} / N_i$, $i, j = 1, 2, \cdots, s$, P_{ij} is the probability of going from state E_i to state E_j in one step, N_i is the number of occurrences of state E_i, N_{ij} is the number of times that state E_i transits to state E_j in one step.Thus, the one-step state transition probability matrix is obtained as P, then the n-step transition probability matrix is $P^{(n)} = P^n$;

Step3 Calculate the autocorrelation coefficients of each order:

$$r_k = \sum_{l=1}^{n-k} (Y(l) - \overline{Y})(Y(l+k) - \overline{Y})$$

, $k \in E$

E Indicates the state space; r_k represents the k -th order autocorrelation coefficient; $Y(l)$ indicates the gray fitting accuracy index of the l th year; $\overline{Y}$ indicates the mean value of the gray fitting accuracy index; n indicates the length of the gray fitting accuracy index sequence. Normalize the autocorrelation coefficients of each order, $w_k = |r_k| / \sum_{k \in E} |r_k|$ as the weight of various time-delay (step) Markov chains;

Step4 Predict the state probability of the gray fitting accuracy index for the year

$$P_i^{(k)},\ i \in E,\ k = 1, 2, \cdots, m$$

The weighted summation of each prediction probability in the same state is used as the prediction probability that the gray fitting accuracy index is in this state.

$$P_i = \sum_{k=1}^{m} w_k P_i^{(k)}, i \in E$$

The i of $\max\{P_i,\ i \in E\}$ is the predicted state of the gray fitting accuracy index for the year;

Step5 Gray fitting accuracy index prediction value: After the state of the gray fitting accuracy index is determined, the state probability linear interpolation method is used to calculate the specific gray fitting precision index value.

$$\hat{Y}(n+1) = y_{1i} \times \frac{p_{i-1}}{p_{i-1} + p_{i+1}} + y_{2i} \times \frac{p_{i+1}}{p_{i-1} + p_{i+1}}$$

Prediction of Maize Yield in Daqing City

The corn yield of Daqing City, Heilongjiang Province from 1995 to 2014 was selected as the raw data (Table 1) for modeling and prediction.

Establishment of Grey GM (1, 1) Prediction Model

According to the above GM (1, 1) prediction step and by means of MATLAB software, it is concluded that:

$\hat{a} = -0.1204$, $\hat{u} = 42.9467$

Then get the predicted value

$$\hat{x}_{k+1}^{(0)} = \hat{x}_{k+1}^{(1)} - \hat{x}_k^{(1)} = (1 - e^a)\left(x_1^{(0)} - \frac{u}{a}\right)e^{-ak} = 50.82e^{0.1204k}, k = 1, 2, \cdots, n-1,$$

The specific results are shown in Table 1.

Table-1: Maize production data of Daqing City from 1995 to 2014 (unit: 10,000 tons)

1995	91.44	91.44	1	1
1996	133.72	57.32	2.33	3
1997	130.17	64.66	2.01	3
1998	106.97	72.93	1.47	2
1999	140.82	82.26	1.71	3
2000	42.37	92.79	0.46	1
2001	67.52	104.66	0.65	1
2002	112.11	118.05	0.95	1

2003	82.65	133.15	0.62	1
2004	124.15	150.19	0.83	1
2005	165.46	169.41	0.98	1
2006	177.57	191.08	0.93	1
2007	160.67	215.53	0.75	1
2008	242.8	243.10	1	1
2009	310.56	274.21	1.13	1
2010	416.18	309.29	1.35	2
2011	458.68	348.87	1.32	2
2012	540.38	393.50	1.37	2
2013	431.16	443.85	0.97	1
2014	451.17	500.64	0.90	1
2015	454.60	564.70		

Weighted markov chain method for prediction of grey fitting precision index

Step1 State division

The state is divided by hierarchical clustering. The results are shown in Table 1. The state interval is shown in Table 2.

Table-2: Gray fitting precision index value state division table

	E1	E2	E3
State interval	0.46~1.23	1.23~1.59	1.59~2.33

Step2 Establish the transition probability matrix

According to the corresponding states of each year in Table 1 and the calculation of the transition probability matrix, the transition probability matrices of each step are obtained as follows:

$$\boldsymbol{P}^{(1)} = \begin{pmatrix} 0.83 & 0.08 & 0.08 \\ 0.25 & 0.5 & 0.25 \\ 0.33 & 0.33 & 0.33 \end{pmatrix}, \boldsymbol{P}^{(2)} = \begin{pmatrix} 0.74 & 0.14 & 0.12 \\ 0.42 & 0.35 & 0.23 \\ 0.47 & 0.31 & 0.22 \end{pmatrix}, \boldsymbol{P}^{(3)} = \begin{pmatrix} 0.69 & 0.17 & 0.14 \\ 0.51 & 0.29 & 0.20 \\ 0.54 & 0.27 & 0.19 \end{pmatrix}.$$

Step3 Calculate the autocorrelation coefficients of various orders and the Markov chain weights of various step sizes.

Table-3: Various order autocorrelation coefficients and Markov chain weights of various step sizes

project	k				
	1	2	3		
r_k	0.46	0.21	0.09		
w_k	0.61	0.28	0.12		

Step4 Predict the status of the gray fitting accuracy indicator

According to the gray fitting accuracy index and the corresponding state transition probability matrix of corn production in Daqing City, Heilongjiang Province from 1995 to 2014, the state of gray fitting precision index of corn production in Daqing City, Heilongjiang Province was predicted.

Table -4: 2015 gray fitting accuracy index prediction table

Initial year	**status**	**time lag/year**	w_k	**State probability**			**Probability source**
				State 1	State 2	State 3	
2014	1	1	0.61	0.83	0.08	0.08	$P^{(1)}$
2013	1	2	0.28	0.74	0.14	0.12	$P^{(2)}$
2012	2	3	0.12	0.51	0.29	0.20	$P^{(3)}$
2015				0.78	0.12	0.11	

According to table 4, $\max\{P_i,\ i \in E\} = 0.78$ and at this point $i = 1$. It can be seen that the possible state of grey fitting accuracy index in 2015 is 1.

Step5 Gray fitting accuracy index prediction value

It can be known from the above calculation results that the possible state of grey fitting accuracy index in 2015 is E_1 and its adjacent states are E_2 and E_3. The specific index value is calculated by interpolation in the interval [a,b], and

$$\hat{Y}(2015) = 0.46 \times \frac{0.12}{0.12+0.11} + 1.23 \times \frac{0.11}{0.12+0.11} = 0.83$$

Data restore, get $x_{2015}^{(0)} = \hat{Y}(2015) \cdot \hat{x}_{2015}^{(0)} = 0.83 \times 564.70 = 468.70$.

It can be seen from table 1 that the prediction accuracy of gray weighted markov model is much higher than that of gray GM (1, 1) prediction model.

CONCLUSION

At present, there are many methods to predict grain yield at home and abroad, but the accuracy of prediction needs to be improved. The GM (1, 1) -weighted markov chain prediction method proposed in this paper has the following main characteristics:

Applying ordered clustering to determine the grading standard, the data structure of the gray fitting accuracy index can be more fully considered, and the state of the division is more reasonable.

According to the weighted markov probability prediction table, the predicted value of grey precision index is obtained by linear interpolation method.

The research ideas in this paper are slightly inferred, and the state interval of the gray fitting precision index can be obtained, instead of the specific value. Under the premise that the actual work needs can be fully satisfied, the reliability of the forecast will also increase.

REFERENCES

1. LI Fu-qin, Sui Hong-zhi. Research Progress in Crop Yield Estimation by Remote Sensing. Progress in Earth Science. 1992, 7(3):30-36.
2. Hayes MJ, Decker WL. Using Noaa Avhrr data to estimate maize production in the united States corn belt.International journal of remote sensing. 1996, 17(16):3189-3200.
3. Chen Xi-kang, Yang Cui-hong. Characteristics of Agricultural Complex Giant System and Prediction of National Grain Production. System Engineering Theory and Practice. 2002, 22(6):108-112.
4. Wang Yu-qian. Study on the spatial and temporal pattern of grain crop input and output and its influencing factors in hebei province. Master Degree Thesis of Hebei University of Science and Technology. 2013.
5. SU Bo, LIU Lu, Yang Fang-ting. Comparative Study on Grain Yield Prediction of GM (1,N) Gray System and BP Neural Network Method. Journal of China Agricultural University. 2006,11(4):99-104.
6. LIU Dong, Bai Xue-feng, MENG Jun. Multivariate Linear Regression Forecast Model of China's Grain Output Based on Forward Selection of Variables. Journal of Northeast Agricultural University.2010, 41(10):124-128.
7. Chen Xiao-xiao, Meng Zhao-wei. Research on Forecast of Shandong Grain Output Based on ARMA Model. Modern Business Trade Industry. 2015(7):31-33.
8. Chen Hua-xi. Application of Grey GM (1, 1) Model in Grain Yield Forecast of Huainan City. Journal of Hunan Institute of Engineering. 2018, 28(1):49-52.
9. Liu Si-feng, Dang Yao-guo, Fang Zhi-geng. Grey system theory and its application. BEI Jing: Science Press. 2004.

1. Daqing guides the project NO.zd-2016-101
2. Cultivation project in Heilongjiang Bayi Agricultural University NO.XA2017-03
3. Daqing guides the project NO.zd-2016-101 zd-2016-091