Study on Time and Space Prediction Model About Rice Yield in Hei Longjiang Province

Guowei Wang, Hongyan Hu, Hao Zhang, Yu Chen

Study on Time and Space Prediction Model about Rice Yield in Hei Longjiang Province

Guowei Wang[a,*],Hongyan Hu[b],Hao Zhang[c],Yu Chen[d]

College of Information Technology, Jilin Agricultural University,Changchun 130118,China;

[a]41422306@@qq.com,[b]1033823087@qq.com., [c]496837382@qq.com,[d]123929697@qq.com

Abstract: Predicting rice yield plays a significant role in preparing production p lan and relevant decision-making for following year. The paper adopts ARIMA time series algorithm taking rice yield per hectare collected between 1991 and 2 010 from different prefecture-level cities of Heilongjiang Province as the object of time series processing, to analyze the rice yield recorded between 1991 and 2009 of different prefecture-level cities and to build prediction model of rice yie ld. By using the model, the rice yield of different prefecture-level cities in 2010 is predicted. The predicted and actual space distribution of rice yield is obtained using Arcgis software. And by conducting spatial data analysis, the space distribution diagram is mapped out. It is analyzed that the accuracy of predicted model on average attains over 95 percent, featuring good prediction effect.

Keywords: ARIMA model; rice; gis; yield prediction.

1 Introduction

Heilongjiang is the national important commodity grain base; among which rice is one of the three major crops. Predicting rice production in heilongjiang province exerts significant impact on the national macro policy. Foreign production prediction method focuses on statistical dynamics growth simulation model, meteorological yield prediction model and remote sensing technology prediction model; while domestic production prediction is based on the mathematical model; Liu Qianpu uses the space-time regression prediction model to predict the output of grain of henan province and municipality, and Li Bingjn predicts short-term grain output using grey linear regression model; Chen Xiangfang proposes a kind of regression tree based on multivariate time series prediction model, to predict the yield of cucumber; Xu Xingmei proposes the model based on clustering analysis and scheduling algorithm to predict corn production. By removing the noise data and reducing the data dimension [7], the paper builds space-time prediction model of rice yield and the rice yield collected in 2010 of Heilongjiang province was predicted by using this model. Such aspects as yield data processing, selection of predictors, establishment of model and stability of model is noticed during the process of prediction, which has improved the accuracy of the prediction.

To sum up, it has yet reported that the comprehensive utilization of gis and time time series analysis are used for space-time prediction of rice production in heilongjiang province.

2 ARIMA model

The ARIMA model is difference autoregressive moving average model in its full name; ARIMA (p, d, q) is called difference autoregressive moving average model, with AR being autoregression; p autoregression item, MA moving average number, d the difference time conducted when time series become reliable. The so-called ARIMA model refers to the model established by converting non-stationary time series into the stationary time series and then performing regression of the dependent variable lag value and the present value of the random error and lag values, so as to convert the nonstationary series into stationary series

The model is generally referred to as the ARIMA (p, d, q); the model parameter--p, d, q are nonnegative integers, meaning autoregression, the order of integration, and all parts of the moving average model. Baucus ARIMA model serves as an important part of a Jenkins method of time series model.

2. 1 Stationarity test

Usually it is based on the following functions to determine sequences' stationary.

Mean function:

$$\mu_n = EX_n = \int_{-\infty}^{\infty} x dF_n(x) \tag{1}$$

Variance function:

$$DX_n = E(X_n - \mu_n)^2 = \int_{-\infty}^{\infty} (x - \mu_n)^2 dF_n(x) \tag{2}$$

Auto-covariance function:

$$\gamma(n, n+k) = E(X_n - \mu_n)(X_{n+k} - \mu_{n+k}) \tag{3}$$

Autocorrelation function:

$$\rho(n, n+k) = \frac{\gamma(n, n+k)}{\sqrt{DX_n \cdot DX_{n+k}}} \tag{4}$$

2.2 pure randomness test

Null hypothesis: delay between periods no more than m phase sequence values are independent of each other.

$$H_0 : \rho_1 = \rho_2 = \rho_3 = \Lambda = \rho_m = 0, \forall m \geq 1 \tag{5}$$

The test statistic:

$$Q_{LB} = n(n+2)\sum_{k-1}^{m}(\frac{\hat{\rho}_k^{\ 2}}{n-k}) \sim \chi^2(m) \tag{6}$$

$Q_{LB} = n(n+2)\sum_{k-1}^{m}(\frac{\hat{\rho}_k^{\ 2}}{n-k}) \sim \chi^2(m)$, reject the null hypothesis, and consider the sequence as a purely random sequence, can be modeled

$Q_{LB} = n(n+2)\sum_{k-1}^{m}(\frac{\hat{\rho}_k^{\ 2}}{n-k}) \sim \chi^2(m)$, accept the null hypothesis, consider the sequence as pure random sequence and model terminal.

In general, take m = 6, 12, 18.

2. 3 Processing of outliers

If X_{t+1} is an outlier, we can use $\hat{X}_t$ to correction X_{t+1} , $\hat{X}_t = 2X_t - X_{t-1}$ 。 [8]

3 rice yield prediction model

3.1 Data Collection

This paper collected the rice yield data in municipalities of Heilongjiang Province from 1991 to 2010, taking rice yield from 1991 to 2009 as training set, taking rice yield in 2010 as test set.

3.2 Data processing

In the case of Rice yield per unit area of Mudanjiang city, after the treatment of abnormal, getting a visual distribution map of the rice yield, as shown in Figure 1:

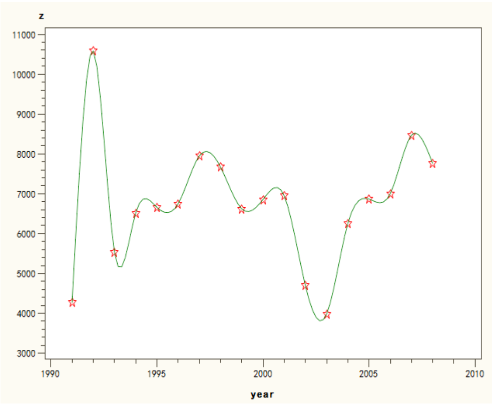

Figure 1. Rice yield distribution maps in Mudanjiang City from 1990 to 2010

Analysis of pictorial diagram, the rice yield doesn't move smoothly enough. After a white noise inspection, we find that the sequence exists in white noise. Quadratic differential on the sequence as shown in figure 2:

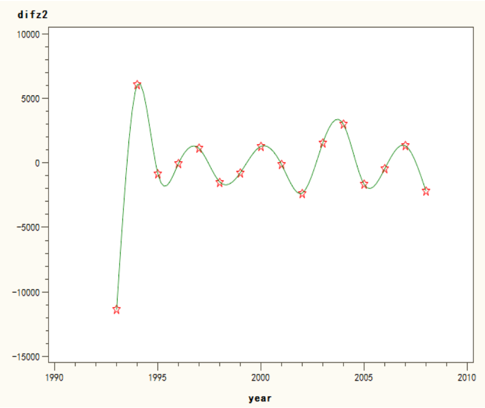

Figure 2. Rice yield distribution after the second differential pictorial diagram

To make the group number of autocorrelation and partial autocorrelation, table 1 for autocorrelation function, table 2 for the partial autocorrelation function:

Table 1. Autocorrelation function

<table>
<tr><th colspan="5">Autocorrelations</th></tr>
<tr><th>Lag</th><th>Covariance</th><th>Correlation</th><th>-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1</th><th>Std Error</th></tr>
<tr><td>0</td><td>12205647</td><td>1.00000</td><td>| |****************|</td><td>0</td></tr>
<tr><td>1</td><td>-4949965</td><td>-0.40555</td><td>| . ****** | . |</td><td>0.250000</td></tr>
<tr><td>2</td><td>-644815</td><td>-0.5283</td><td>| . * | . |</td><td>0.288199</td></tr>
<tr><td>3</td><td>1496268</td><td>0.12259</td><td>| . |** . |</td><td>0.288803</td></tr>
<tr><td>4</td><td>-1120015</td><td>-0.09176</td><td>| . ** | . |</td><td>0.292038</td></tr>
</table>

Table 2. Partial autocorrelation function

<table>
<tr><th colspan="3">Partial Autocorrelations</th></tr>
<tr><th>Lag</th><th>Correlation</th><th>-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1</th></tr>
<tr><td>1</td><td>-0.40555</td><td>| . ******| . |</td></tr>
<tr><td>2</td><td>-0.26007</td><td>| . *****| . |</td></tr>
</table>

3	-0.01268	| . | . |
4	-0.06155	| . * | . |

Pictorial diagram can be seen that the production distribution has no obvious cyclical rice production, and the quadratic differential autocorrelation function and partial autocorrelation function can be seen that rice yields in Mudanjiang city Heilongjiang province in 1991-2008 were stationary series.

3.3 Determine the order number

Based on the BIC criterion

$$BIC(n) = \ln \hat{\sigma}_{\varepsilon}^{2}(n) + \frac{n}{N}\ln N$$

Inside, n is the number of parameters. If an order number n_0' meet

$BIC(n_0') = \min_{1 \le n \le M(N)} BIC(n)$, $M(N)$ is equal to $[\sqrt{N}]$ or $[\frac{N}{10}]$, n_0' is the best order. After the calculation, p=3，q=1. Therefore, rice yield prediction model was ARIMA (3, 2, 1), To examine the Bic sequence analysis by SAS software, as shown in Table 3

Table 3. BIC order determination results map

Minimum Information Criterion						
Lags	*MA0*	*MA1*	*MA2*	*MA3*	*MA4*	*MA5*
AR0	13.1106	13.03749	10.6488 6	-22.9253		
AR1	13.13289	13.08792				
AR2	9.877755	-22.5844	-19.9156	-24.3156		
AR3	-22.8096	-26.8701	-25.0403			
AR4	-25.6454	AR4				
AR5						
Minimum Table Value:BIC(3,1)=-26.8701						

4 yield prediction

According to the rice yield prediction model , To predict rice yield in cities of Heilongjiang province in 2010 . The forecast output and the actual output are shown in Table 4 . The distribution maps of Prediction and actual yield in Mudanjiang city in 2010 is shown in Figure 3.

Table 4. the yield of rice in Heilongjiang Province in 2010 actual yield is compared with the predicted values.

The prefecture level city	Prediction of yield(kg / HA)	The actual yield(kg / HA)
Daqing	9326.15	9094.245
Harbin	8927.8238	8842.127
Hegang	6226.5706	6174.625
Heihe	6672.7103	5387.501
Jixi	7851.6911	7774.634
Jia Musi	8225.7806	7517.753
Mudanjiang	7721.8741	7839.623
Qigihar	7154	6725.376
Shuangyashan	7396.7602	7567.996
Suihua	8857	8959.85
Yichun	8068	8046.453

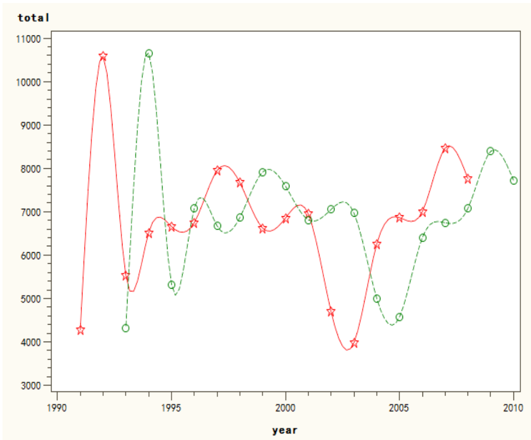

Figure 3. The distribution maps of Prediction and actual yield n Mudanjiang city in 2010

5 Spatial analysis of rice yield

By using ArcGIS software，To establish the spatial distribution map of The yield of rice in Heilongjiang Province in 2010 cities that the actual and the predicted. As shown in Figure 4, Figure 5.

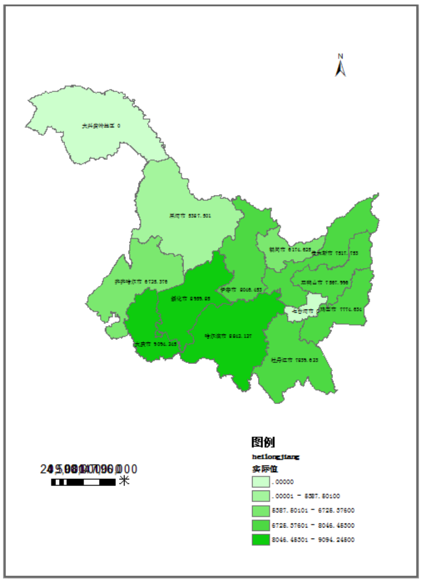

Figure 4. The actual yield distribution in space

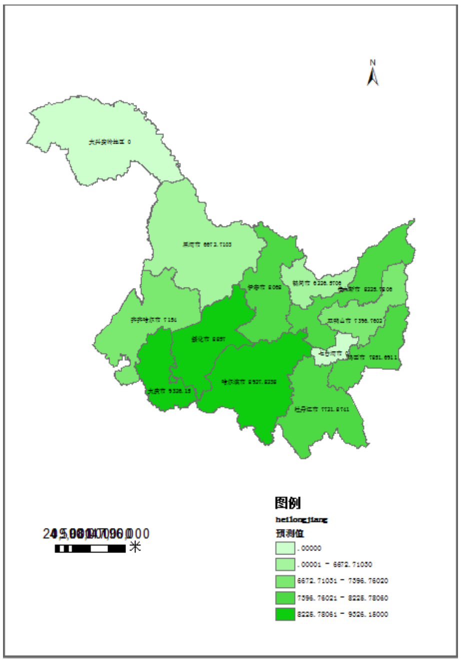

Figure 5. prediction of yield distribution in space

Using the functions of GIS spatial analysis, Error analysis for the actual yield of rice and predicted values in 2010 Heilongjiang province. The spatial distribution map of error is shown in Figure 6.

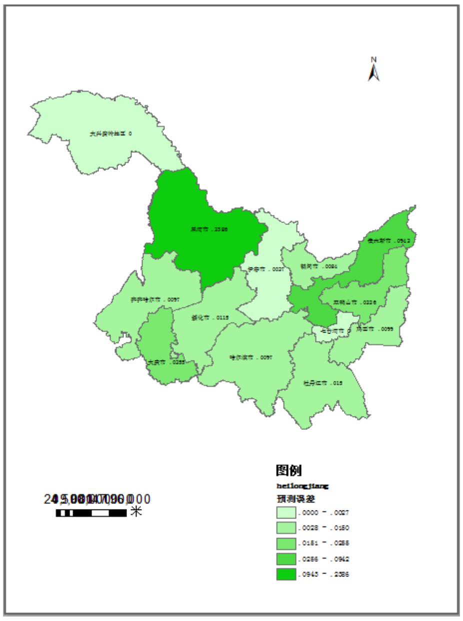

Figure 6 distribution of the prediction error space

According to the above analysis, Getting the prediction error of Rice yield of municipalities in Hei longjiang province. As shown in Table5 .

Table 5. prediction error of rice yield of Heilongjiang province in 2010

cities	Prediction error	cities	Prediction error
Daqing	0.0255	Mudanjiang	0.01502
Harbin	0.009692	Qigihar	0.063732
Hegang	0.008413	Shuangyashan	0.022626
Heihe	0.238554	Suihua	0.011479
Jixi	0.009911	Yichun	0.002678
Jia Musi	0.094181		

It can be seen from table 5 that Yichun has the least prediction error of 0.26 percent, with Heihe having the largest prediction error of 23.86 percent. By carrying out calculation, the prediction error of different cities on average is 4.5%.

6 Conclusions

The paper establishes time prediction model of rice production in Heilongjiang province using ARIMA model and gis technology; By using the model, the prediction value of rice yeild in heilongjiang province in 2010 is obtained; and the actual production, predicted production and predict the spatial distribution of prediction error is acquired. By conducting analysis, the prediction accuracy of the model on average reaches over 95 percent that can be used for rice yield prediction and that provides relevant scientific reference for government sectors' overall planning and decision-making.

However, the prediction error of this model used for prediction of rice yield of Heihe reaches a maximum value of 23.8 percent, indicating the model is not well designed taking all factors into account that further research is still required.

References

1.Wang Huiting [J]. statistics and decision making,prediction of grain yield of mixture time series modelbased on 2013, (12): 23-25.

2. Liu Qinpu. The grain yield of Henan province prediction [J]. spatio-temporal regression model based onAnhui agricultural sciences, 2011,39(28):17631-17633.DOI:10.3969/j.issn.0517-6611.2011.28.189.

3. Li Bingjun; Li Qiufang; Lu Xiuxia. Application of graylinear regression combination model in the prediction of grain yield in Henan province [J]. Henan Agricultural Sciences2009,10:44-47.DOI:10.1016/j.jhin.2010.02.010

4. Li Jinxia. Prediction of grain production in HenanProvince, the grey system model [J]. Journal of Henan University of Technology (Social Science Publishing)based on,2009,5(4):1-3,7.DOI:10.3969 /j.issn.1673-1751.2009.04.001.

5.Chen Xiangfang, Chen Ming, Feng Guofu et al. [J].computer engineering and design model for the prediction of multi variable time series regression tree yield of cucumber,2012,33(1):407-411.DOI:10.3969/j.issn.1000-7024.2012.01.080.

6.Xu Xingmei, Zhou Cuijuan, Chen Guifen et al.Clustering analysis and sequential algorithm of cornproduction model test [J]. Journal of Jilin Agricultural University based on, 2012,34(6):688-691,704.

7.Su Hengqiang, Zhu Chunrao, Wen Changji. Model in corn fertilization forecasting [J]. Journal of Jilin Agricultural University, 2010,32(3):312-315.
8.Xiao Zhihong, Guo Mingyue. Time series analysis and application of SAS [M] Wuhan University press, 2009
9.Ana Stojanovic,Leonid Stoimenov,Aleksandar Dimitrijevic,Slovodanka Djordjevic-Kajan,GinisWeb-The Tool for GIS Applications Development based on Web,10th International Electrotechnical Conference MEleCon 2000,Vol.1,pp331-333
10.ESRI Shapefile Technical Description,GIS by ESRI,1998