COTTON YIELD INTELLIGENT PREDICTION RESEARCH XiaoWen Wei CRI,CAAS China, Anyang

Abstract

See from the long-term trend, the cotton yield level to be subjected to the technique progress push completely. The technique progress is an integration factor. It includes the exaltation of the seed production capability, the growth of the technical abilities and production technique knowledge of people, reasonable utilization of plantation, and the improvement of production condition and implements of production, perfect of plant and protects technique and etc. The technique progressive development track turns and twists, and be influenced by the different action result and diffuse speed of its self in different region, thus it has the different representation in the different region. On the other hand the yield motion between years is mainly affected by climate factor. The weather influence has much complicated and periodic characteristic, it is some frequency compositions or vary swing of period series to add, and also has the region difference. Therefore, adopt the intelligent method, make use of the calculator procedure to build and choose the most valid mathematics model automatically, can rise the apply scope and the estimate accuracies of predict enormously. Through fitted the cotton yield of the China, the United States and world to express that their models' R^2 are all over more than 0.92.

1 The undulating regulation of cotton yield

See from the long-term trend, the cotton yield level to be subjected to the technique progress push completely. The technique progress is an integration factor. It includes the exaltation of the seed production capability, the growth of the technical abilities and production technique knowledge of people, reasonable utilization of plantation, and the improvement of production condition and implements of production, perfect of plant and protects technique and etc. The technique progressive development track turns and twists, and be influenced by the different action result and diffuse speed of its self in different region, thus it has the different representation in the different region. We use the migration arithmetic mean method, Since by Henan Province, Shanxi Province and Xinjiang autonomous region 1949 cotton yield level 5 years migration averaged data mapping, May obviously feel to the different per unit area yield grows or thetechnology advancement tendency (see Figure 1). Therefore, Should full consider in the model establishment process, And finally will form has wide difference the different forecast model.



Figure 1: 5 years migration average yield of Henan Province, Shanxi Province and Xinjiang autonomous region

On the other hand the yield undulates in years, that mainly displays for receives the climate (Wei Xiaowen, 1996) and other stochastic factors influences. Since through the national cotton yield ambulation proceeding simply from 1949, after dissolving short-term motion its fluctuant periodic characteristic is an outstanding coming out then and clearly. However climate and the yield undulation cycle has the suitable complexity, it is not the result of action or influence by sole cyclical, But is the complex compound which affects together by the multi-cycles forms. After research proof, this complex compound manifests specifically for has the different oscillation amplitude certain frequency ingredients superimposition, namely presents has the harmonic component a race frequency spectrum group.

2 The cotton yield predict model

2.1 Build model

The production level looked from the long-term trend, is invited by technique advancement completely. The technology advancement is a synthesis resultants, It including the seed production performance enhancement, the people's produces technical knowledge and the skill accumulation, the reasonable use of land, the improvement of production tools, cultivation management technology factor and so on. But the years' difference of production level mainly influenced by the climate and other stochastic factors. The technology advancement although is one kind of tendency, But its development path is zigzag and not smooth, It is influenced by it self's effect and its propagation rate in production. Therefore the irregularity of technology advancement, had decided the complexity and difficult to separate the action of technology advancement factor from production level.

According to the above, to analyze 1919-1993 year's national cotton yield material preliminary, has discovered the conic relations between technology advancement and the time (year):

 $Y_t=281.3727-9.4370T+0.2325T^2$, among them, Y_t – yield level(kilogram/hectare), T- time. Related coefficient of model R=0.953, the coefficient R²=0.908, F and T examination shows the regression coefficient is remarkable, but D.W. statistics measure indicated the model is self correlation, need to be improved, namely the model exists system error, also can say the separation of technique progresses still not perfect. For this model improvement is:

 $Y_t = 105.3275 - 3.3742T + 0.0889T^2 + 0.6123Y_{t-1}$

Related coefficient R=0.971, coefficient $R^2=0.942$, F and T examination shows the regression coefficient is remarkable. It expresses the production level in addition to relation with time's conic, and to have something to do with the production level in last period. Videlicet, the technique advancement process has the phenomenon of setback or accelerating development. for instance, after a disaster, soil or seed quality descent, plant diseases and insect pests take place etc., cause the technique action descent in next period of production. Contrarily, seed or environment qualities increases, the plant diseases and insect pests alleviates, the actual performance of technique is enlarged in next period.

Through the national and various provincial areas cotton yield forecast analysis discovered the y forecast effect is not identical. Sometimes very is good, but sometimes misses, its reason lies in yield level undulated in years, the main performance that it is influenced by the climate and other stochastic factors.

Therefore, we have joined the cyclical variable in the model. Because of the complexity of climate and yield undulation cycle, namely its cyclical is not the sole cyclical, but is the complex compound which affects together by the multi-cycles forms. This complex compound is the superimposition of same certain frequencies with different oscillation amplitude, namely presents has the harmonic component frequency spectrum group. If we put all the cycles in the model, on the one hand it is too complex and will be reduced own or other parameters significances, on the other hand model's stability will be decreased. Therefore, the key of model construction lies in discovers certain influences or the oscillation amplitude greatest cycle, and put them in the model. The model after cycle improvement its forecast effect had the distinct enhancement.

2.2 The calculation method of the model solution and tools

It is a good choice, to get the result of polynomial model by matrix method on computer.

This research applies a lot of data handles and data construction, but the database language Visaul Fox Pro's data buildup handles ability is stronger, and it can make use of its language to write auto handles procedure. In addition, its face to the object and visualization programmed method are good for making a program with friendly interface, and easy control management etc. As a result this research adopts the Visaul Fox Pro database language, and I use it to write the forecast system independence completely.

Choose the matrix computational method to get solution of a concrete model. Using Visaul Fox Pro data-handling capacity, Makes the database files of primitive observation data (the DBF table), And further defers to the matrix solution method compilation related procedure, By modular idea design relatively independent function block, Enhances the procedure whole legible and the maintainability, it is signality for this processes which needs frequently to transfer, Like matrix multiplication module (process).

Multivariate linear regress model as follows:

 $y_t = b_1 + b_2 x_{t2} + b_3 x_{t3} + ... + b_k x_{tk} + u_t$, t=1,2,...,n

t means observation times or time, n denotes sample capacity. The yt is the observation variable in t time(year), x_{tj} (j=2,3, ..., k) is independent variable. b1 is intercept, b2, b3 ..., b4 are slope parameters. ut is error item.

For each t, vector $X_t=(1, x_{t2}, x_{t3}, \dots, x_{tk})$, set vector $B = (b_1, b_2, b_3, \dots, b_k)$ '. The model then can mean as:

 $Y_t = X_t B + U_t, t=1,2, ..., n$

All variables expressed as follows with matrix:



Multivariate linear regress model with matrix is: $\mathbf{y} = \mathbf{x}\mathbf{b} + \mathbf{u}$ The following matrixes usually apply to calculate regression coefficient, R² coefficient, variable examination, etc:

$$\mathbf{X'X} = \begin{bmatrix} n & \sum X_{t1} & \cdots & \sum X_{tk} \\ \sum X_{t1} & \sum X^2_{t1} & \cdots & \sum X_{t1}X_{tk} \\ \vdots & \vdots & & \vdots \\ \sum X_{tk} & \sum X_{t1}X_{tk} & \cdots & \sum X^2_{tk} \end{bmatrix}$$
$$\mathbf{X'Y} = \begin{bmatrix} \sum Y_t \\ \sum Y_t X_{t1} \\ \vdots \\ \sum Y_t X_{tk} \end{bmatrix}$$

And matrix $(\mathbf{X'X})^{-1}$.

2.3 Model intelligence choice

As a result of various local production and economic environment difference, and its cotton production technical level and so on difference, Will create the forecast model the difference, including variable choice difference in model. And in a same area, its forecast model also has the possibility to be different at indifferent time. In order to enhance the forecast effect most, automatically completes the model using the computer program control the choice to become the best key to the situation which the model constructs. Automatically achieves the model choice goal what principle

using the computer program is? Because the final goal of the model constructs is the realization as far as possible accurate forecast, Therefore in compilation computer program, the final model choice principle is the determinable coefficient maximization. Certainly, In order to enhance the model's accuracy and stability, the model also must follow the related economical principle, and manifests the economical causal relation in variables.

3 Forecasts applications effect appraisal

Using this forecast method to forecast cotton yield in China, the United States and the world, its effect is much good. Because uses the intellectualized method, automatically will produce the different forecast model in different time and different area, Thus enhanced the forecast effect.

3.1 China

 $Y = b_0 + b_1Y_{t-1} + b_2T + b_3T^2 + b_4C_1 + b_5C_2 + b_6C_3 + b_7C_4 + b_8C_5 + e_t$

Y is the dependent variable cotton yield, C expresses the cyclical variable, T (t) is the time (year), E is an error term. C1... C5 indicated the greatest 5 cycles affected, at present they respectively are 14, 16, 17, 15 and 8 years. The model coefficient RS is 0.9676, regression coefficient R is 0.9837. The model forecast the Chinese cotton yield is 1166kg/hectare in 2005.



Figure 2: Cotton yield forecast in China (kg/ha)

3.2 The United States

 $Y = b_0 + b_1Y_{t-1} + b_2T + b_3T^2 + b_4C_1 + b_5C_2 + b_6C_3 + b_7C_4 + b_8C_5 + e_t$

Y is the dependent variable cotton yield, C expresses the cyclical variable, T (t) is the time (year), E is an error term. C1... C5 indicated the greatest 5 cycles affected, at present they respectively are 27,28,26,29 and 30 years. The model coefficient RS is 0.9265, regression coefficient R is 0.9626. The model forecast the Chinese cotton yield is



Figure 3: Cotton yield forecast in the United States (kg/ha)

3.3 The world

 $Y = b_0 + b_1 Y_{t-1} + b_2 T + b_3 T^2 + b_4 C_1 + b_5 C_2 + b_6 C_3 + b_7 C_4 + b_8 C_5 + e_t$

Y is the dependent variable cotton yield, C expresses the cyclical variable, T (t) is the time (year), E is an error term. C1... C5 indicated the greatest 5 cycles affected, at present they respectively are 23,22,21,24 and 20 years. The model coefficient RS is 0.9858, regression coefficient R is 0.9929. The model forecast the Chinese cotton yield is 659kg/hectare in 2005.



Figure 4: The world cotton yield forecast (kg/ha)

4 Conclusions

ICAC (international cotton advisory committee) and international food and agriculture organization through cooperation research, has promoted it perfect findings officially in 1993 -- "the world cotton market model". This model including the cotton harvest area, cotton yield, cotton demand, cotton price, cotton factory expends, the non-cotton fiber expense and so on 7 multinomial mathematical model. Cotton yield model is: $Y = b_0 + b_1P_{t-1} + b_2T + b_3D_n + b_4Y_{t-1} + e_t$.

Y is the dependent variable cotton yield, P expresses cotton's producer price, T is the tendency output, Dn is the hypothesized variable, T is the time (year), E is an error term.

In this research forecast used the model although also is the multinomial model, But its variable and the realization method is completely different. Because the cotton yield forecast adopts the intellectualized method, automatically produces and chooses the most effective model, including cyclical variable introduction, thus enhanced the forecast applicable scope and the forecast precision greatly.

References

FAO, Modelling the World Cotton Market. The World Cotton Market, 1993, 10:25~29
Wei Xiaowen, Chinese cotton production characteristic and production forecast. China Cotton, 1996, (2)
Wei Xiaowen, Study the Forecast of Chinese Cotton Production,
USDA, Cotton: World Markets and trade, 2000-2003, USA