ISSN : 0974 - 7435 Volume 10 Issue 2

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(2), 2014 [055-060]

Research and application on performance prediction of the olympic men's sprint based on GM (1, 1) model

Ling Xu[1]*, Deding Tang[2], Mingming Xing[2], Yufei Jin[2]

[1]Institute of Physical Education, Anhui Polytechnic University, Wuhu 241000, (CHINA)

[2]Department of Physical Education, Maanshan Teacher's College, Maanshan 243000, (CHINA)

ABSTRACT

This paper takes the Olympic men's 100m sprint champion performance in recent six sessions as the original data, and uses $GM(1,1)$ prediction model in gray system theory to predict men's 100m sprint champion performance in the 31st Olympic Games. The results can be obtained by using MATLAB curve fitting $\hat{x}^{(1)}(k+1) = -1622.73e^{0.006117\,k} + 1632.69$; the predicted score is 9.59 seconds. The results show that: the score of men's 100m sprint champion in 31 Olympic will break the Olympic record. The findings have some significance to explore the athletics development law of track and field and to further promote the application of gray system theory in sports.

KEYWORDS

Men's 100m;
GM (1, 1);
Prediction model;
Gray theory.

INTRODUCTION

Men's 100m sprint is most intense in track and field competition, also is the most exciting project. The achievements of our athletes on this project have a big gap with that of outstanding foreign players, especially in the past two Olympic Games athlete Usain Bolt of Jamaica continuously refreshed the Olympic record and let the world remember his name.

Since developed sprint countries has greater advantage in terms of training concepts, training methods, training means and sports technology, the world sprint goes into a period of rapid development, mainly in: the world record is broken again and again, the new record cycle shortens, world sprint enters into a stage of rapid development. How the competitive level of the 31th Olympic Games men's 100m project develop, whether the record will be broken frequently in accordance with the law of athletics development should be a matter of concern for sprint. Therefore, it is necessary to predict the champion score of the 31th Olympic men's 100m project and provide reference for China to grasp the developmental level of the world men's 100m athletic ability.

After the screening of a six-dimensional priority model $GM(1,1)$, the prediction accuracy of the six-dimensional model $GM(1,1)$ is higher, so it takes the Olympic men's 100m sprint champion performance in recent six sessions as the data to establish the $GM(1,1)$

prediction models of the 31th Olympic men's 100m sprint and predict, which has important practical significance to grasp the development law of swimming athletic ability and to further promote the application of gray $GM(1,1)$ model in sports performance prediction.

ESTABLISHMENT AND TEST OF GRAY MODEL FOR SPRINT PERFORMANCE PREDICTION

Establishment of gray model $GM(1,1)$

Definition 1Suppose the sequence

$$X^{(0)}=\left(x^{(0)}(1),x^{(0)}(2);\cdots x^{(0)}(n)\right)$$
$$X^{(1)}=\left(x^{(1)}(1),x^{(1)}(2);\cdots x^{(1)}(n)\right)$$
$$Z^{(1)}=\left(z^{(1)}(1),z^{(1)}(2);\cdots z^{(1)}(n)\right)$$

Equation $x^{(0)}(k)+az^{(1)}(k)=b$ is gray differential equation, a is called the development gray number, and b is the endogenous control gray number.
Wherein:

$x^{(0)}(k),\ k=1,2,\cdots,n$ is the original series

$x^{(1)}(k)=\sum_{i=1}^{k}x^{(0)}(i), k=1,2,\cdots,n$ $X^{(1)}$ is the 1-AGO sequence of $X^{(0)}$

$z^{(1)}(\mathrm{k})=0.5x^{(1)}(k)+0.5x^{(1)}(k-1)$ $Z^{(1)}$ is the near raw sequence of $X^{(1)}$

The above differential equations satisfy the three conditions of gray differential equations.

Definition 1 Suppose the sequence $X^{(0)}=\left(x^{(0)}(1),x^{(0)}(2);\cdots x^{(0)}(n)\right)$ and $X^{(0)}$ is the sequence non-negative.

$$X^{(1)}=\left(x^{(1)}(1),x^{(1)}(2);\cdots x^{(1)}(n)\right)$$
$$Z^{(1)}=\left(z^{(1)}(1),z^{(1)}(2);\cdots z^{(1)}(n)\right)$$

Grey differential equation: $x^{(0)}(k)+az^{(1)}(k)=b$

Where in $x^{(0)}(k),\ k=1,2,\cdots,n$ is the original series.

$x^{(1)}(k)=\sum_{i=1}^{k}x^{(0)}(i), k=1,2,\cdots,n$ $X^{(1)}$ is the 1-AGO sequence of $X^{(0)}$

$z^{(1)}(\mathrm{k})=0.5x^{(1)}(k)+0.5x^{(1)}(k-1)$ $Z^{(1)}$ is the near raw sequence of $X^{(1)}$.

If $\hat{\alpha}=(a,b)^T$ is the parameter column, order:

$$Y=\begin{bmatrix}x^{(0)}(2)\\x^{(0)}(2)\\\cdots\\x^{(0)}(2)\end{bmatrix},\quad B=\begin{bmatrix}-z^{(1)}(2) & 1\\-z^{(1)}(3) & 1\\\cdots & \cdots\\-z^{(1)}(n) & 1\end{bmatrix}$$

The least-squares parameter estimation columns of gray differential equations $x^{(0)}(k)+az^{(1)}(k)=b$ satisfy: $\hat{\alpha}=\left(B^TB\right)^{-1}B^TY$

Definition 2: $X^{(0)}=\left(x^{(0)}(1),x^{(0)}(2);\cdots x^{(0)}(n)\right)$ Is non-negative sequence, $X^{(1)}=\left(x^{(1)}(1),x^{(1)}(2);\cdots x^{(1)}(n)\right)$ is the 1-AGO sequence of $X^{(0)}$, $Z^{(1)}$ is the near raw sequence of $X^{(1)}$,

$\frac{dx^{(1)}}{dt}+ax^{(1)}=b$ is the winterization equation of gray differential equation $x^{(0)}(k)+az^{(1)}(k)=b$, which is also called shadow equation.

Definition 3: The relational expression of $B,Y,\hat{a}$ is $\hat{\alpha}=\left(B^TB\right)^{-1}B^TY$:

1) The solution of winterization equation $\frac{dx^{(1)}}{dt}+ax^{(1)}=b$ is also called the time response function

$$x^{(1)}t=\left(x^{(1)}(0)-\frac{b}{a}\right)e^{-at}+\frac{b}{a}$$

2) The time response sequence of GM (1,1) gray differential equation $x^{(0)}(k)+az^{(1)}(k)=b$ is

$$\hat{x}^{(1)}(k+1)=\left[x^{(1)}(0)-\frac{b}{a}\right]e^{-ak}+\frac{b}{a},$$

$k=1,2,\cdots,n$

3) Set $x^{(1)}(0)=x^{(0)}(1)$then

$$\hat{x}^{(1)}(k+1)=\left[x^{(0)}-\frac{b}{a}\right]e^{-ak}+\frac{b}{a},$$

$k=1,2,\cdots,n$

4) Reducing Value

$$\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k)$$

Description of the problem related to the modeling: the data of the original sequence $X^{(0)}$ does not have to be used for modeling completely, the different trade-offs of the original data can get the different models, namely a and b is different.

When choosing the used data in the established mathematical model, we must ensure that the sequence time produced by the model building is equidistant with continuity, and no breakpoint occurs. When generally establishing the mathematical model based on the data sequence, the data used should be the most recent data, and the data needs to be adjacent. When new data are generated, there are usually two methods that can be used to deal with: the first is adding the new data to the original data sequence, and re-estimate the value of the parameter; the second is to weed out the most old and ancient set of data in the original data, and add on the latest obtained data sets, we must ensure that data the elements number of the formed sequence is equal to that of the original data, and then re-estimate the parameters.

Model test

The rest of model $GM(1,1)$ is divided into three areas: residual test; association degree test; posterior error test.

Residual test

The residual error test refers to the checking and validation on the errors of the simulated number generated by the established model and the actual value one by one.

The absolute residual sequence $\Delta^{(0)}=\{\Delta^{(0)}(i),i=1,2,\cdots,n\},\Delta^{(0)}(i)=\left|x^{(0)}(i)-\hat{x}^{(0)}(i)\right|$

And the relative residual sequence $\phi=\{\phi_i,i=1,2,\cdots,n\}$ $\phi_i\left[\frac{\Delta^{(0)}(i)}{x^{(0)}(i)}\right]\%$

And calculate the average relative residual

$$\bar{\phi}=\frac{1}{n}\sum_{i=1}^{n}\phi_i$$

If α is known, when $\bar{\phi}<\alpha$ and $\phi_n<\alpha$, the mathematical model is called the residual qualified mathematical model.

Two criteria of the models' residual test:

(1) When α is known and $\bar{\Delta}<$, means that the residual error of the mathematical model is qualified (General α takes $0.01-0.05$)

(2) If the generated average value by the relative error of the established mathematical model is in the range of 1%, at this time the level of the mathematical model is the first level. When the mean value of the model's relative error is greater than 1% and less than or equal to 5%, the accuracy level of the model is the second level. When the mean value of relative error is greater than 5% and less than or equal to 10%, the accuracy level of the model is the third level. When the mean value of relative error is greater than 10% and less than or equal to 20%, the accuracy level of the model is the fourth level.

Association degree test

Association degree test generally refers to: conduct curve fitting on the data generated by modeling, then conduct similarity comparison and testing with the fitting curve of the data sequence generated by the original model. In accordance with the mentioned correlation calculation pattern, the correlation coefficient of $\hat{x}^{(0)}(i)$ and the original data sequence $x^{(0)}(i)$ is obtained, then correlation degree is calculated according to the correlation coefficient; it is generally believed that, if the correlation degree> 0.6, it is considered to be satisfactory, otherwise it is considered to be unsatisfactory.

Association degree test means conducting test by observing the similarity level of the model value curve and the modeling sequence curve and calculating the results by the previously described correlation calculation method.

Posterior error test

Posterior error test generally refers to test on the

FULL PAPER

statistical properties of the residual distribution.

(1) Calculate the average value of the original sequence:

$$\bar{x}^{(0)} = \frac{1}{n}\sum_{i=1}^{n} x^{(0)}(i)$$

(2) Calculate the mean square error of the original series $X^{(0)}$:

$$S_1 = \left(\frac{\sum_{i=1}^{n}\left[x^{(0)}(i) - \bar{x}^{(0)}\right]^2}{n-1} \right)^{1/2}$$

(3) Calculate the mean value of the residuals:

$$\bar{\Delta} = \frac{1}{n}\sum_{i=1}^{n} \Delta^{(0)}(i)$$

(4) Calculate the mean square error of the residuals:

$$S_2 = \left(\frac{\sum_{i=0}^{n}\left[\Delta^{(0)}(i) - \bar{\Delta}^{(0)}\right]^2}{n-1} \right)^{1/2}$$

(5) Calculate the variance ratio C:

$$C = \frac{S_1}{S_2}$$

(6) Calculate the small residual probability:

$$P = p\left\{\left|\Delta^{(0)}(i) - \bar{\Delta}\right| < 0.6745 S_1\right\}$$

Set $S_0 = 0.6745 S_1$, $e_i = \left|\Delta^{(0)}(i) - \bar{\Delta}\right|$, that is $P = p\{e_i < S_0\}$

For given $C_0 > 0$, when $C < C_0$, the model is called the qualified model of mean square error ratio, as shown in TABLE 1. For given $P_0 > 0$, when $P < P_0$, the model is called the qualified model of small residual probability.

TABLE 1 : The reference table of posterior residual test

P	C	Model accuracy	P	C	Model accuracy
>0.95	<0.35	excellent	>0.70	<0.65	Barely qualified
>0.80	<0.5	qualified	<0.70	>0.65	disqualified

If the relative residuals, correlation degree and posterior test are within the allowable range, it can be predicted by the model built; otherwise it should be corrected with residuals.

THE GRAY $GM(1,1)$ MODEL SOLUTION OF PERFORMANCE PREDICTION PROBLEM

Take the champion score of men's 100m sprint in the 25th, 26th, 27th, 28th, 29th and 30th Summer Olympics as the original data to establish the GM (1, 1) prediction model.

The champion score of men's 100m sprint in the recent 6 Summer Olympic Games is shown in TABLE 2 below (the results take "s" as a unit to be easy to calculate), establish the gray prediction model.

TABLE 2 : The champion performance statistical tables of men's 100m sprint in the recent 6 summer olympic games

All previous sessions (session)	25	26	27	28	29	30
performance (s)	9.96	9.84	9.87	9.85	9.69	9.63

The model solution is as follows:

Suppose: $X^{(0)}(k) = \{9.96, 9.84, 9.87, 9.85, 9.69, 9.63\}$

Step 1 Construct the accumulated generated sequence

$$X^{(1)}(k) = \{9.96, 19.8, 29.67, 39.52, 49.21, 58.84\}$$

Step 2 Construct data matrix B and data vector Y_n

$$B = \begin{bmatrix} -\frac{1}{2}\left[x^{(1)}(1) + x^{(1)}(2)\right] & 1 \\ -\frac{1}{2}\left[x^{(1)}(2) + x^{(1)}(3)\right] & 1 \\ -\frac{1}{2}\left[x^{(1)}(3) + x^{(1)}(4)\right] & 1 \\ -\frac{1}{2}\left[x^{(1)}(4) + x^{(1)}(5)\right] & 1 \\ -\frac{1}{2}\left[x^{(1)}(5) + x^{(1)}(6)\right] & 1 \end{bmatrix} = \begin{bmatrix} -14.88 & 1 \\ -24.74 & 1 \\ -34.60 & 1 \\ -44.37 & 1 \\ -54.03 & 1 \end{bmatrix},$$

$$Y_n = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ x^{(0)}(4) \\ x^{(0)}(5) \\ x^{(0)}(6) \end{bmatrix} = \begin{bmatrix} 9.84 \\ 9.87 \\ 9.85 \\ 9.69 \\ 9.63 \end{bmatrix}$$

Step 3 Calculate

$$\hat{\alpha} = \begin{bmatrix} a \\ b \end{bmatrix} = \left(B^T B\right)^{-1} B^T Y_n$$

$$\hat{\alpha} = \left(B^T B\right)^{-1} B^T Y_n \ \hat{\alpha} = \begin{bmatrix} 0.006117 \\ 9.987170 \end{bmatrix}$$

Step 4 Obtain the prediction model

$$\frac{dx^{(1)}}{dt} + 0.006117 x^{(1)} = 9.987170$$

$$\hat{x}^{(1)}(k+1) = -1622.73 e^{0.006117\,k} + 1632.69$$

$$\left(x^{(0)}(1) = 9.96\ ; \quad \frac{b}{a} = 1632.69 \right)$$

Step 5 Residual test

(1) According to the forecast formula, calculate $\hat{X}^{(1)}(k)$, get:

$$\hat{X}^{(1)}(k) = \{9.96\ 19.86, 29.7, 39.48, 49.2, 58.87, 68.47\}$$

$$(k = 0, 1, \cdots, 6)$$

(2) Regressive generated sequences $\hat{X}^{(0)}(k), k = 0, 1, \cdots, 6$

$$\hat{X}^{(0)}(k) = \{9.96, 9.90, 9.84, 9.78, 9.72, 9.67, 9.60\}$$

Original sequence: $X^{(0)}(k) = \{9.96, 9.84, 9.87, 9.85, 9.69, 9.63\}$

(3) Calculate absolute residual and relative residual sequences

Absolute residual sequence: $\Delta^{(0)} = \{0, 0.06\ 0.03, 0.07, 0.02, 0.02\}$

Relative residual sequence: $\phi = \{0, 0.60\%\ 0.30\%\ 0.71\%\ 0.21\%\ 0.21\%\}$

The relative residual is less than 1%, and the model accuracy is high.

(4) Calculate the correlation coefficient

$$\eta(k) = \frac{\min\{\Delta(k)\} + P \max\{\Delta(k)\}}{\Delta^{(0)}(k) + P \max\{\Delta(k)\}} (k = 1, ..., 6, P = 0.5)$$

Get:

$$\eta(k) = \{1, 0.37, 0.54\ 0.33\ 0.64, 0.64\}$$

(5) Calculate the correlation degree:

Correlation degree:

$$r_i = \frac{1}{6} \sum_{k=1}^{6} \eta_i(k) = 0.601$$

When r = 0.601, it satisfies the p = 0.5test criterion r> 0.6.

Step 6: Posterior residual test:

(1) Calculate:

$$x_0 = \frac{1}{6}[9.96 + 9.84 + 9.87 + 9.85 + 9.69 + 9.63] = 9.81$$

(2) Calculate the mean square error of sequence $x^{(0)}$:

$$S_1 = \left(\frac{\sum [x^{(0)}(k) - \bar{x}]^2}{n-1}\right)^{\frac{1}{2}} = 0.0474$$

(3) Calculate the mean value of residuals:

$$\bar{\Delta} = \frac{1}{6}[\Delta(k)] = 0.03$$

(4) Calculate the mean square error of residuals:

$$S_2 = \left(\frac{\sum [\Delta(k) - \bar{\Delta}]^2}{n-1}\right)^{\frac{1}{2}} = 0.122746$$

(5) Calculate C:

$$C = \frac{S_2}{S_1} = \frac{0.0059}{0.54522} = 0.3862$$

(6) Calculate the small residual probability:

$$S_0 = 0.6745 \times 0.0474 = 0.032$$

$$e_k = \left| \Delta(k) - \bar{\Delta} \right| = \{0, 0.03, 0, 0.04, 0.01, 0.01,\}$$

All e_i are less than S_0, the small residual probability $P\{e_i < S_0\} = 83.33\%$, moreover $C = 0.3862 < 0.5$ thus the model is qualified.

Through expression $\hat{x}^{(1)}(k+1) = -1622.73 e^{0.006117\,k} + 1632.69$, it can be obtained that the score of men's 100m champion in the thirty-first Olympic is 9.59s, which will break the Olympic record.

The development trends of the Olympic 100m sprint in recent years is shown in Figure 1 below.

It can be obtained from the figure that the champion achievement of the Olympic men's 100m sprint has some volatility, but the fluctuation band is very narrow; but the champion performance of recent sessions are relatively stable, there will not be a lot of deviation, the 100m sprint of the 31 Olympic Games may break record.

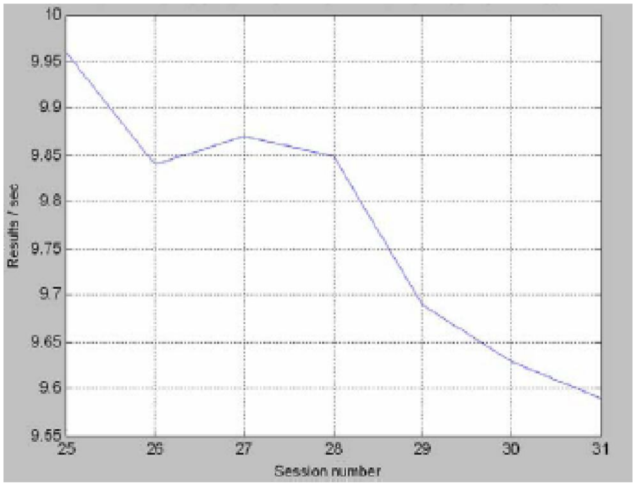

Figure 1 : The performance trends of the Olympic 100m sprint champion in recent years

CONCLUSIONS

The advantages of gray $GM(1,1)$ prediction model are applied to the complex study mechanism, more hierarchic and situation that is difficult to establish accurate models for quantitative measurement. Since the mathematical methods used in the theory is non-statistical method, when the system data is less or the condition does not meet the statistical requirements, this method is more practical, which is in line with the characteristic of the performance prediction after the 30th Olympic Games; and because there are many factors affecting 100m sprint performance, study factors are more complex, it closer to the actual result by using gray prediction, so using the gray system theory to predict the score just can take advantage of the gray system theory. Whether a model was able to predict depends on whether it passes the test, in this paper, the gray model passes through the residual test, correlation degree test and posterior error test, so the $GM(1,1)$ model is suitable for predicting the performance.

REFERENCES

[1] Zhu Hong-Bing, Liu Jian-Tong, Wang Gang; Gray Forecast of GM (1, 1) Model and Its Application in Forecast of Sports Performance [J]. Journal of Capital College of Physical Education, **15(1)**, 118-121 **(2003)**.

[2] Wang Wei; Application Research about System Theory and Ma erkefu Chain Used in Predicting Results of Track and Field [J]. Journal of Nanjing Institute of Physical Education, **18(6)**, 85-87 **(2004)**.

[3] Zhao Yun Hong, Zhou Yao; Equal Dimensions Gray Number Pattern and Its Forecast of World and National Records in Women 100 m Hurdles [J]. China Sport Science and Technology, **38(2)**, 39-40 **(2002)**.

[4] Hen Liang, Sai Qing-Bin; An Analysis on the Developing Trend of the Performance of Track and Field in National Sports Games and Its Study with Grey Prediction[J]. Journal of Capital College of Physical Education, **18(5)**, 85-88 **(2006)**.

[5] Song Ai-Ling, Chen Kai; Gray Model and Forecast of Hurdles the Best Results in the World and Chinese Liuxiang[J]. Journal of Capital College of Physical Education, **15(4)**, 68-69,116 **(2003)**.

[6] Chen Liang, Tian De-Bao; The Application Research of the Reconstructed GM (1, 1) Model of Background Value in Sports [J]. Journal of Capital College of Physical Education, **18(6)**, 127-128 **(2006)**.

[7] Tong Liping, Yuan Jianguo; GM (1, 1) Model of International and National Records in Women's 400 m Hurdling and Calculation [J]. Journal of Shanghai Physical Education Institute, **25(2)**, 44-47 **(2001)**.

[8] Liu Jia-Jin; Research on the New Methods of Gray Envelope in Sports Field [J]. Journal of Guangzhou Physical Education Institute, **26(2)**, 54-56 **(2006)**.

[9] Liu Jia-Jin et al.; Applied Study of GM (1, 1) Models in Sports Forecast [J]. Zhejiang Sport Science, **21(1)**, 60-64 **(1999)**.