Journal of Computational Science & Engineering 23 (2016) 703-706

Available at http://www.asocse.org ISSN 1710-4068

Mathematical Prediction of Flood Loss in Shenzhen City

Zhengyu Li[1], Hejun Xu[2], Shuhui Fang[3]

[1]School of Information and Control Engineering, Shenyang Jianzhu University, Liaoning, China

[2]School of Civil Engineering, Harbin Institute of Technology, Harbin, China

[3]School of Municipal and Environmental Engineering, Shenyang Jianzhu University, Shenyang, China

Abstract: **Firstly, the flood disaster data of Shenzhen City from 1980 to 2007 were analyzed, and the gray area prediction model was used to forecast the disaster area in 2020, 2030 and 2040, showing a trend of increasing year by year. After that, the ARMA time series model was used to analyze the data. After analyzing and testing, the trend of the runoff of Shenzhen in the future is obtained. Combined with the above two points, draw the future development trend of Shenzhen flood damage.**

Keywords: Shenzhen flood loss; Grey forecasting; ARMA model

Received 12 February 2016, Revised 23 March 2016, Accepted 25 March 2016

*Corresponding Author: Hejun Xu, 24396725@qq.com

1. Introduction

Recently, foreign studies have forecast Shenzhen as the most affected by flood damage in the city, which also led to a series of discussions. The majority of domestic scholars questioned which triggered a series of discussions. Based on this problem, this paper builds mathematic model based on the statistical data of disaster and runoff in Shenzhen, and predicts the development trend of future flood losses in Shenzhen.

2. Methods

2.1. Roughing index

First of all, to the affected the statistical data of 1980-2007 in Shenzhen city are arranged, then carries on the screening of rough set, and then to discrete data (Table 1) [1].

Perform the following discrete processing of each attribute in the table:

The affected area (1.<=50 2.51~100 3.101~150 4.151~200 5.>=200);

The disaster area (1.<=30 2.31~60 3.61~90 4.91~120 5.>=120);

Sown area (1.<=480 2.481~500 3.500~520 4.>=520);

To simplify the data after the table 2 below:

It assumes that a a= affected area, b= disaster area, c= crops, d= disaster degree（Table 2）.

Index selection for the 28 domain, draw:

rC(d)=1,
rC(d)-r{b,c}(d)=0.64,
rC(d)-r{a,c}(d)=0
rC(d)-r{a,b}(d)=0.79

Thus, the attribute C is biggest, and the B property of the weight is negligible. The system normalized after treatment, a, B, C attribute weights were 0.64, 0, 0.79. Thus, the flood of the Shenzhen City, the serious degree of influence factors is the crop planting area.

Table 1 1980-2007 Shenzhen city disaster statistics

Coding	Year	Disaster area	Disastered area	area
X1	1980	53.87	30.33	545.13
X2	1981	57.74	24.87	542.01
X3	1982	60.12	27.05	540.34
X4	1983	80.2	47.3	542.32
X5	1984	64.9	31.4	539.03
X6	1985	31	17	516.14
X7	1986	36	20	521
X8	1987	56	20	515.09
X9	1988	74.46	40.1	519.63
X10	1989	76.84	42.48	536.56
X11	1990	70.47	23.63	536.52
X12	1991	63	30.8	524.32
X13	1992	71.47	17.53	468.84
X14	1993	122.73	81	507.74
X15	1994	134.6	84.4	511.56
X16	1995	111.96	74.82	513.39
X17	1996	123.67	77.26	515.53
X18	1997	75.16	46.21	507.48
X19	1998	213.4	146.83	513.52
X20	1999	137.48	86.84	502.99
X21	2000	29.28	12.81	480.28
X22	2001	42.68	20.7	465.26
X23	2002	171.19	95.23	452.98
X24	2003	91.15	53.06	508.22
X25	2004	132.56	85.84	521.52
X26	2005	59.41	38.02	529.58
X27	2006	93.9	62.1	529.59
X28	2007	117.33	81.44	529.59

2016 © The American Computational Science Society. All rights reserved.

Table 2 Rough statistics

DomainU	a	b	c	d	DomainU	a	b	c	d	DomainU	a	b	c	d	DomainU	a	b	c	d
X1	2	2	4	1	X8	2	1	3	2	X15	3	3	3	3	X22	2	1	1	1
X2	2	1	4	2	X9	2	2	3	2	X16	3	3	3	3	X23	4	4	1	4
X3	2	1	4	2	X10	2	2	4	2	X17	3	3	3	3	X24	2	2	3	2
X4	2	2	4	2	X11	2	1	4	2	X18	2	2	3	2	X25	4	3	4	3
X5	2	2	4	2	X12	2	2	4	2	X19	5	5	3	5	X26	2	2	4	2
X6	1	1	3	1	X13	2	1	1	2	X20	3	3	3	3	X27	2	3	4	2
X7	1	1	4	1	X14	3	3	3	3	X21	2	1	2	1	X28	3	3	4	3

2.2. Gray theoretical predictions

Time series of gray GM (1,1)[2] model to predict the time taken for the study from 1980 to 2005 series, a five-point data selection。Modeling method is the raw data Yt (0) randomness to be desalinated，Do a cumulative generation of the original data , metering successive superposition and generating sequence .

Providing：

$$Y_t^{(0)} = \sum_{i=1}^{t} Y_t^0 \quad (1)$$

$$Z_t^{(1)} = (Y_t^{(1)} + Y_{t-1}^{(1)})/2 \quad (2)$$

$Y_t^{(0)}$ is the original sequence in the T Data， $Y_t^{(1)}$ is the cumulative number of columns in the tth generate data , $Z_t^{(1)}$ is the mean number of columns in the t-th data (from start to t = 2:00).

With the above formula, the results are in the table below. Temporal distance is 5 years, so here take the t-th data represents the first 5t years. The results are described in table 3.

On the cumulative number of columns were generated by an exponential curve fitting approach. Function of the form:

$$Y_t^{(0)} = [Y_t^{(1)} - \frac{b}{a}]e^{-a(t-1)} + \frac{b}{a} \quad (3)$$

In this section, is forecast year of affected area； is original year of affected area。a、b are parameter(Table 3).

AT=（a,b），A=(BTB)-BTY。The use of accumulated generating series and the mean number of columns, using the least squares method the parameters a, b.

Once accumulated above, draw：

$$B = \begin{bmatrix} -Z_2^{(1)} & 1 \\ -Z_3^{(1)} & 1 \\ -Z_4^{(1)} & 1 \\ -Z_5^{(1)} & 1 \\ -Z_6^{(1)} & 1 \end{bmatrix} = \begin{bmatrix} -69.37 & 1 \\ -120.11 & 1 \\ -211.3 & 1 \\ -281.94 & 1 \\ -326.29 & 1 \end{bmatrix}$$

$$Y = \begin{bmatrix} Y_2^{(0)} \\ Y_3^{(0)} \\ Y_4^{(0)} \\ Y_5^{(0)} \\ Y_6^{(0)} \end{bmatrix} = \begin{bmatrix} 31 \\ 70.47 \\ 111.96 \\ 29.28 \\ 59.41 \end{bmatrix}$$

$$B^TB = \begin{bmatrix} 249841.6 & -1009.01 \\ -1009.01 & 5 \end{bmatrix} \quad (4)$$

Providing D= B^TB ， $(B^TB)^{-1}=D^{-1}$.Matrix inverse matrix calculation formula is$D^{-1}=D^*/|D|$， D* is the adjoint matrix of D. Obtainion：

$$D^{-1} = \begin{bmatrix} 0.000002 & -0.0043 \\ -0.0043 & 1.081 \end{bmatrix}$$

$$A = (B^T B)^{-1} B^T Y = \begin{bmatrix} -2.466 \\ 592.81 \end{bmatrix} \quad (5)$$

Parameters a = - 1.4224, the parameter b = 592.822. The parameters a, b into the obtained：

$Y_t^{(1)}$ =460.48e1.223（t-1）-416.51

So, T is the first year of the affected area is forecast：

$$Vt = Y_t^{(1)} - Y_{t-1}^{(1)} \quad (6)$$

The affected area for 2020 predicts V_{17}:

$$Vt = Y_{17}^{(1)} - Y_6^{(1)} \quad (7)$$

Calculated, in 2020 the affected area is 105.87; similarly calculate the affected area 2030 V_{27} = 119.59, 2040 years affected area V_{37} = 139.59. It shows an increasing trend.

Table 3 Lists the mean number

number	year	origin$Yt^{(0)}$	one generated cumulative value$Yt^{(1)}$	average$Zt^{(1)}$
1	1980	53.87	53.87	
2	1985	31	84.87	69.37
3	1990	70.47	155.34	120.11
4	1995	111.96	267.3	211.3
5	2000	29.28	296.58	281.94
6	2005	59.411	355.99	326.29

2016 © The American Computational Science Society. All rights reserved.

3. Use of runoff data to predict future trends in disaster

In recent years the use of Shenzhen, the largest and most trails traffic data [1], the establishment of ARMA model analyze, and then predict the future trend of development in Shenzhen floods.

3.1. Stability test of the original data

Because only involves a set of measured data (runoff data), the data does not need to assume that distribution. Therefore, the choice test data run stability test method [3]

In order to keep the sequence of the original premise swim with the same scale is defined as a sequence of symbols, such symbols can put observations into opposing categories. For example the value of the observation sequence is x_i, mean of x, by the symbol "+" indicates $x_i \geq x$, and "-" indicates $x_i \leq x$.

Processing the raw data in this way, the results obtained were in table 4.

The "+" and "-" a total of 55, divided into 21 run. We use a sample statistics:

= Total number of one symbol appears

=Total number of another symbol appears

r= The total number of run [4] (Table 4)

Among, r as the test volume. When N1 or N2 exceeds 15 can be approximated by a normal distribution, which can be used to determine the normal distribution table test accepted domain and negative domains. At this point the selected statistical standards:

$$Z = \frac{\text{run numbers} - \text{run of the expected number}}{\text{run standard deviation}} = \frac{r - u_r}{\sigma_r} \quad (8)$$

In the section

$$u_r = \frac{2N_1N_2}{N} + 1;\ \sigma_r = \left[\frac{2N_1N_2(2N_1N_2 - N)}{N^2(N-1)}\right]^{1/2}$$

$$N=N_1+N_2 \quad (9)$$

For the significance level of α=0.05，If $|Z| \leq 1.96$，we can accept the null hypothesis, or to refuse. According to the above equation, derived raw data | Z | = 5.75, is the original non-stationary sequence.

At the same time, because of the ARMA model data should be smooth and of zero mean (ie eliminate the impact of causal factors), so for non-stationary time series data must first be smoothing, zero-mean process [5].

Reference signal processing theory of smoothing, the original time series data for five years $\{X_t\}$ smoothing obtain a sequence $\{n_t\}$, the formula is:

$$n_t = \frac{x_t + x_{t-1} + x_{t-2} + x_{t-3} + x_{t-4}}{5} \quad (10)$$

Then,Making a natural logarithmic transformation to $\{n_t\}$，$y_t = \ln n_t$ getting sequence $\{y_t\}$，Sequence $\{y_t\}$ for zero mean treatment,the formula is: $z_t = y_t - \bar{y}_t$，in this section:

$$\bar{y}_t = \frac{1}{n}\sum_{t=1}^{n} y_t \quad (11)$$

For z_t second-order difference stationary process: $\nabla^2 Z_t = S_t$（$\nabla = 1 - B$ is a differential operator），S_t sequences obtained were smooth run test, the results show a relatively stable S_t zero-mean time series, meet the conditions fit ARMA (p, q) model.

Table 4 ARMA model data stationary test table

-5274	-7384	-1964	1526	-5594	5626	-9874	-2554	6726	4026
—	—	—	+	—	+	—	—	+	+
-1704	-3164	8626	3826	426	5326	-7084	7226	-2884	-474
—	—	+	+	+	+	—	+	—	—
5626	2426	9126	-7724	526	8426	-4634	-3644	-3974	-3674
+	+	+	—	+	+	—	—	—	—
1226	12126	4026	-3144	-2154	4726	4726	……	-5854	-8824
+	+	+	—	—	+	+	……	—	—

3.2. ARMA model established

Use ARMA model PANDIT-WU [6] algorithm, establish runoff prediction equation. Model thinking: first for ARMA (2,1) model fitting, and gradually increase the number of model order, until on higher-order ARMA (2n + 2,2n + 1), were fitted. Then after F test for each model. If F_0 <F_0.05 (taking the significance level) is that the ARMA (2n, 2n-1) is more reasonable; if think ARMA (2n + 2,2n + 1) is more reasonable, then continue to rise until the residual sum of squares and the order is not significantly reduced (ie, receiving the null hypothesis) so far.。

Using PANDIT-WU,sequence from lower order to higher-order one by one fitting ARMA (p, q) model:

$$x_t - \varphi_1 x_{t-1} - \cdots - \varphi_p x_{p-1} = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \quad (12)$$

In the section, $\varphi_1, \varphi_2 \ldots \varphi_p \ldots \varphi_n$ and $\theta_1, \theta_2 \ldots \theta_q \ldots \theta_n$ are real。p，q are nutual number.

$\{a_n\}$ is white noise sequence，getting maximum runoff ARMA（2，1）model and minimum runoff ARMA（5，4）model as follow:

Max: $x_t = 0.0952x_{t-1} + 0.0345x_{t-2} + a_t - 0.670a_{t-1}$

2016 © The American Computational Science Society. All rights reserved.

Min:

$$x_t = 0.487x_{t-1} + 1.001x_{t-2} - 0.439x_{t-3} - 0.0412x_{t-4} + a_t - 0.189a_{t-1} - 0.822a_{t-2} + 0.211a_{t-3}$$

4. Conclusion

In this paper, the disaster statistics and runoff statistical data through the city of Shenzhen, the corresponding gray prediction model is established and the ARMA time series model, respectively, to the disaster situation Shenzhen city in the future is forecasted, the following conclusions are obtained: On the flood disaster in Shenzhen City, the serious degree of influence factors is the crop planting area, Shenzhen City, the total affected area showed a rising trend year by year. In Shenzhen City, the biggest future runoff will continue at high level fluctuations, the annual flood control work is still very grim.

References

[1] Guangdong Provincial Bureau of statistics. Statistical Yearbook Guangdong Guangdong (2013)1980-2012.

[2] S.J. Zhang, X. Qi. Time series analysis simple tutorial, (2003) 15-46.

[3] Q.Y. Jiang, A mathematical model, (2003) 45-69.

[4] S.X. Guo, Y.Y. Huang, H.J. Xu. University Physics Experiment, 28 (2015) 51-53.

[5] R.B. Wang. Shenzhen city water Zhi Guangdong, (2001) 17-65.

[6] H.J. Xu. Industrial Construction 45 (2015) 589-591.

2016 © The American Computational Science Society. All rights reserved.