

基于马尔科夫过程修正的湖南洪涝灾变灰预测研究

侯月,廖基定*

(南华大学 数理学院,湖南 衡阳 421001)

摘要:通过对1985~2015年的湖南省受灾率和成灾率的数据分析,研究未来15年湖南省洪涝灾害的灾年预测和发展趋势.利用统计数据建立灰色灾变模型GM(1,1),得到了洪涝灾害有阶段性波动且受灾率与成灾率有明显的同步性的结论.经由马尔科夫过程修正的灰色残差模型,具有更高的准确率,可为中长期预测提供参考.

关键词:洪涝灾害;灰色残差模型;马尔科夫过程;灾年预测

中图分类号:O211

文献标志码:A

Grey Prediction of Flood Disaster in Hunan Based on Markov Chain Correction

HOU Yue, LIAO Ji-ding*

(School of Mathematics and Physics, University of South China, Hengyang, Hunan, 421001, China)

Abstract:Based on the data analysis of the disaster rate and hazard rate in Hunan Province in 1985~2015, this paper researches the disaster prediction and the development trend of flood disaster in Hunan Province in the next 15 years. Using statistical data to establish the grey catastrophe model GM(1,1), the flood disaster has periodic fluctuations and disaster rate and hazard rate have obvious synchronization conclusion. Through the grey residual model of Markov process correction, with higher accuracy, it can provide reference for long-term forecasting.

key words: flood disaster; grey residual model; Markov process; predic disasters

0 引言

目前,洪涝灾害依然是世界上影响最巨大的灾害,在各种自然灾害所造成的损失中占40%,超过一半的全球洪水灾害发生亚洲^[1],据不完全统计,

平均每年遭受洪涝灾害的农田有 7.3×10^6 hm,高达 $150 \times 10^8 \sim 200 \times 10^8$ 元的经济损失,占国民生产总值1%~2%.近年来,很多专家学者根据已有的历史文献对洪涝灾害的发生规律与预测进行了多方面的研究.在2007年,李祥、王心源等提出了关于洪水的

收稿日期:2017-09-29

作者简介:侯月(1992-),女,硕士研究生,主要从事概率论与数理统计等方向的研究.E-mail:1205097803@qq.com.

* 通讯作者:廖基定,E-mail:liaojiding@163.com.

灰马尔可夫预测理论,然后通过对巢湖流域的预测开发出修正模型^[2],对本文模型的建立提供了现实依据.在文献[3]中提出了基于人工神经网络(artificial neural net, ANN)的地理信息系统(geographic information system, GIS)来解决洪水灾害评估与预测问题,结果表明人工神经网络预测模型有较好的预测精度.然而,神经网络算法是一个复杂的非线性化问题,网络训练容易失败.朱育红,吕翠兰等在2016年的第33届中国气象学会年会上对淮河流域的洪水预报进行讨论,利用遥感与GIS结合进行洪涝灾害预警防洪^[4],给本文研究提供了理论依据.

邓聚龙在20世纪70年代末提出灰理论,目前灰色模型GM(1,1)^[5]作为预测模型已在很多领域有很多成就,同时也存在偏差过大的情况.本文根据1985~2015年湖南省重大洪涝灾害的统计数据,采用改进灰色灾变模型建模,并用马尔科夫过程修正来预测未来15年内湖南省重大洪涝灾害发生的趋势.

1 灰色灾变模型

1.1 模型的建立

灰色灾变预测的核心是GM(1,1)模型,它属于单一变量一阶微分方程预测模型,是对无规律的原始数据序列经过数据处理,使其变成有规律的新序列,然后建立模型.GM(1,1)模型更适用于短时小样本预测^[6],湖南省几十年的洪涝灾害发生的数据属于小样本数据,可以用GM(1,1)模型进行洪涝灾害发生的短期灰色灾变预测.

具体建模过程如下^[7]:

设原始数据序列为 $X^{(0)}(i) = \{x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)\}$,规定一个阈值 ξ , $X^{(0)}(i)$ 中大于或小于 ξ 的值为异常值记为 x_{ξ} ,将其组成新的数据列 $X_{\xi}(k) = \{x_{\xi}(1), x_{\xi}(2), x_{\xi}(3), \dots, x_{\xi}(m)\}$, $m \leq n$,称为上限或下限灾变数据列.相对应的时间为灾变点,将其组成灾变时间序列 P .记灾变点为 p ,显然有映射 $P: \{x_{\xi}(k)\} \rightarrow \{p\}$,即 $P(x_{\xi}) = p$.

记 $P^{(0)}(k) = \{p^{(0)}(1), p^{(0)}(2), p^{(0)}(3), \dots, p^{(0)}(m)\}$,将 $P^{(0)}(k)$ 做一次累加,生成1-AGO序列 $P^{(1)}(k) = \{p^{(1)}(1), p^{(1)}(2), p^{(1)}(3), \dots, p^{(1)}(m)\}$,其中, $p^{(1)}(k) = \sum_{i=1}^k p^{(0)}(i)$, $k = 1, 2, 3, \dots, m$.那么,GM(1,1)模型的白化形式为

$$\frac{dp^{(1)}(t)}{dt} + ap^{(1)}(t) = b \quad (1)$$

式中, a, b 均为参数列, a 发展系数, b 是灰作用量.

$$\text{设矩阵 } \mathbf{B} = \begin{bmatrix} -\frac{1}{2}[p^{(1)}(1) + p^{(1)}(2)] & 1 \\ -\frac{1}{2}[p^{(1)}(2) + p^{(1)}(3)] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}[p^{(1)}(m-1) + p^{(1)}(m)] & 1 \end{bmatrix},$$

$$\mathbf{Y} = \begin{bmatrix} p^{(0)}(2) \\ p^{(0)}(3) \\ \vdots \\ p^{(0)}(m) \end{bmatrix}, \text{则参数 } a, b \text{ 可用最小二乘估计得出参数量}$$

$$\hat{\mathbf{a}} = \begin{bmatrix} a \\ b \end{bmatrix} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{Y} \quad (2)$$

于是,模型的离散形式为

$$\hat{p}^{(1)}(k+1) = (p^{(0)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a} \quad (3)$$

最后,将 $\hat{p}^{(1)}(k+1)$ 累减还原灾变时间模型为

$$\hat{p}^{(0)}(k+1) = \hat{p}^{(1)}(k+1) - \hat{p}^{(1)}(k) = (1 - e^a)(p^{(0)}(1) - \frac{b}{a})e^{-ak} \quad (4)$$

其中 $p^{(0)}(1) = \hat{p}^{(1)}(1)$.

1.2 模型检验

本文采用后验差检验法^[7-8]进行模型检验.该方法是以残差为基础并对其概率分布进行检验的,其中小误差概率 P^* 和后验差比值 C 是检验模型可行性的指标.

设残差 $\varepsilon^{(0)}(k) = p^{(0)}(k) - \hat{p}^{(0)}(k)$,则有残差序列 $\varepsilon^{(0)}(k) = \{\varepsilon^{(0)}(1), \varepsilon^{(0)}(2), \varepsilon^{(0)}(3), \dots, \varepsilon^{(0)}(m)\}$.记数据序列 $p^{(0)}(k)$, $k = 1, 2, \dots, m$ 的均值为 $\bar{p}$,即: $\bar{p} = \frac{1}{m} \sum_{k=1}^m p^{(0)}(k)$.同样地,残差序列 $\varepsilon^{(0)}(k)$, $k = 1, 2, \dots, m$ 的均值 $\bar{\varepsilon}$,有 $\bar{\varepsilon} = \frac{1}{m} \sum_{k=1}^m \varepsilon^{(0)}(k)$.设数据序列 $p^{(0)}(k)$ 和残差数列 $\varepsilon^{(0)}(k)$ 的方差分别为 S_1 和 S_2 ,则有

$$S_1 = \sqrt{\frac{1}{m} \sum_{k=1}^m (p^{(0)}(k) - \bar{p})^2},$$

$$S_2 = \sqrt{\frac{1}{m} \sum_{k=1}^m (\varepsilon^{(0)}(k) - \bar{\varepsilon})^2} \quad (5)$$

那么,后验差比值 C 和小概率误差 P^* 可表示为:

$$C = \frac{S_2}{S_1}, P^* = P^* \{ |\varepsilon^{(0)}(k) - \bar{\varepsilon}| < 0.674 5S_1 \}$$

(6)

在检验过程中,要求后验差比值 C 越小越好而小概率误差 P^* 越大越好.因为 P^* 越大,残差和其残差平均值的差小于 $0.674 5S_1$ 的序列点较多,表示预测数值分布较为均匀.当 C 越小时,表明 S_2 越小而 S_1 越大. S_1 越大,表示原始时间序列方差越大,其数据的离散度越大; S_2 越小,表示预测数列残差方差越小,其残差的离散度越小.由此说明, C 很小表示尽管原始时间序列比较离散,但是模型所得预测值与真实值之差并不太离散.

预测等级一般分为 4 个^[9],如表 1 所示.

表 1 后验差检验标准

Table 1 Inspection criteria for posterior variance examination

等级	P^*	C
好(一级)	>0.95	<0.35
合格(二级)	>0.80	<0.45
勉强(三级)	>0.70	<0.50
不合格(四级)	≤ 0.70	≥ 0.65

2 应用

2.1 数据的选取与处理

本文选取的资料大部分在 1985 年到 2015 年《湖南统计年鉴》和《中国统计摘要》中选取所得,同时对于资料中不全的数据,采用了中华人民共和国国家统计局、中华人民共和国水利部、湖南省统计局和统计年鉴分享平台等权威部门网站所公布的数据和其他文献中提供的数据^[10-11].为了便于探求湖南省洪涝灾害的时间分布规律,主要利用农作物受灾面积和成灾面积与当年粮食播种面积的比值,即农作物受灾率和成灾率^[12]作为基础指标,将其标准化后的数据作为指标进行研究,见表 2.其中,定义标准化后的数据为异常指数^[13],则有表达式如式(7)所示:

$$\eta_i = \frac{M_i - \bar{M}}{\delta}$$
 (7)

式中 η_i 代表第 i 年受灾率异常指数(AIA)或成灾率异常指数(AIS), M_i 是第 i 年受灾率或成灾率, $\bar{M}$ 是平均受灾率或成灾率, δ 是成灾率或受灾率的标准差.

表 2 湖南省 1985 年~2015 年洪灾数据

Table 2 Flood disaster data in Hunan province in 1985~2015

年份	受灾面积/ ($\times 10^4$ hm)	成灾面积/ ($\times 10^4$ hm)	播种面积/ ($\times 10^4$ hm)	受灾率	成灾率	受灾异常指数	成灾异常指数
1985	31.00	17.00	516.14	0.060 1	0.032 9	-0.849 7	-0.818 6
1986	36.00	20.00	521.04	0.069 1	0.038 4	-0.801 1	-0.770 8
1987	56.00	20.00	515.10	0.108 7	0.038 8	-0.587 6	-0.767 0
1988	82.00	36.93	519.63	0.157 8	0.071 1	-0.323 2	-0.484 5
1989	47.20	17.67	533.05	0.088 5	0.033 1	-0.696 3	-0.816 7
1990	70.47	23.63	536.56	0.131 3	0.044 0	-0.465 8	-0.721 3
1991	63.00	30.80	536.52	0.117 4	0.057 4	-0.540 7	-0.604 2
1992	71.47	17.53	524.36	0.136 3	0.033 4	-0.439 0	-0.814 2
1993	122.73	81.00	505.05	0.243 0	0.160 4	0.135 9	0.297 8
1994	134.60	84.40	507.74	0.265 1	0.166 2	0.254 9	0.349 0
1995	111.96	74.82	511.56	0.218 9	0.146 3	0.005 8	0.174 1
1996	123.67	77.26	513.39	0.240 9	0.150 5	0.124 5	0.211 2
1997	75.16	46.21	515.53	0.145 8	0.089 6	-0.387 9	-0.321 9
1998	213.40	146.83	507.48	0.420 5	0.289 3	1.092 2	1.427 4
1999	137.48	86.84	513.52	0.267 7	0.169 1	0.269 0	0.374 3
2000	29.28	12.81	502.99	0.058 2	0.025 5	-0.859 7	-0.884 0
2001	42.68	20.70	480.28	0.088 9	0.043 1	-0.694 6	-0.729 5
2002	171.19	95.23	465.26	0.367 9	0.204 7	0.809 0	0.685 9
2003	91.15	53.06	452.98	0.201 2	0.117 1	-0.089 2	-0.081 0
2004	132.56	85.84	508.22	0.260 8	0.168 9	0.231 9	0.372 5

续表

年份	受灾面积/ ($\times 10^4$ hm)	成灾面积/ ($\times 10^4$ hm)	播种面积/ ($\times 10^4$ hm)	受灾率	成灾率	受灾异常指数	成灾异常指数
2005	59.41	38.02	521.52	0.113 9	0.072 9	-0.559 6	-0.468 5
2006	93.90	62.10	529.58	0.177 3	0.117 3	-0.218 1	-0.079 9
2007	117.33	81.44	529.59	0.221 5	0.153 8	0.020 3	0.240 0
2008	543.50	315.21	494.94	1.098 1	0.636 9	4.742 8	4.471 8
2009	107.98	62.09	479.91	0.225 0	0.129 4	0.038 9	0.026 3
2010	201.13	123.67	480.91	0.418 2	0.257 2	1.079 9	1.145 6
2011	68.95	35.67	487.96	0.141 3	0.073 1	-0.412 1	-0.466 7
2012	106.43	60.65	490.80	0.216 9	0.123 6	-0.005 0	-0.024 6
2013	63.42	34.12	493.70	0.128 5	0.069 1	-0.481 3	-0.501 7
2014	104.11	59.86	497.51	0.209 3	0.120 3	-0.045 9	-0.053 1
2015	75.74	41.73	494.47	0.153 2	0.084 4	-0.348 1	-0.367 8

2.2 湖南省洪涝灾害的变化趋势

由表 2 数据做受灾和成灾率异常指数分析图,见图 1、图 2、图 3。

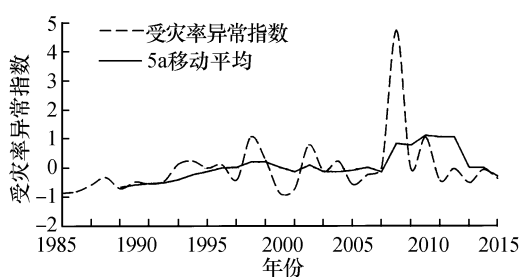


图 1 受灾率异常指数图

Fig.1 The disaster of rate anomaly index

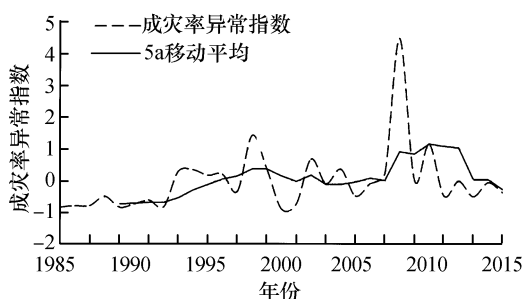


图 2 成灾率异常指数图

Fig.2 The hazard of rate anomaly index

根据图 1 和图 2 湖南省 1985~2015 年洪涝灾害受灾率和成灾率异常指数的时间变化曲线所示,洪涝灾害呈阶段性波动.结合图 1,图 2 显示,1985~1997 年旱灾少而轻属于轻灾期;1998~2002 年旱灾稍重属中度灾期;2003~2007 年属于

第二个轻灾期;2008~2012 年灾情重属于大旱灾.

根据图 3 所示,湖南省洪涝灾害受灾率异常指数与成灾率异常指数随时间的变化有明显的同步性,经分析两者相关系数达到 0.989 1.由此可见,湖南省农业系统防洪抗灾能力较弱.

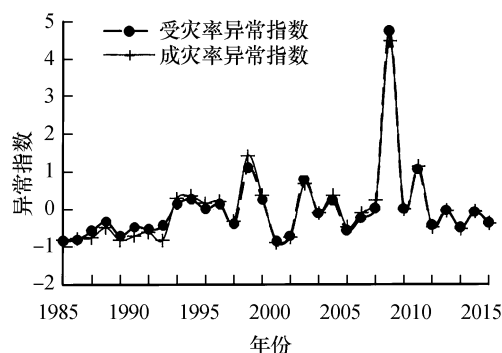


图 3 受灾率和成灾率异常

Fig.3 The disaster rate and hazard rate of anomaly index

2.3 湖南洪涝灾害重灾年份预测

根据表 2 数据,分别取阈值 $\xi > 0$ 和 $\xi > 0.8$ 的受灾率异常值和成灾率异常值序号建立年份序号序列.取 $\xi > 0$ 和 $\xi > 0.8$ 的受灾率异常值,建立年份序号序列 $D_1^{(0)}$ 和 $D_2^{(0)}$.由于在此范围内的受灾率异常值和成灾率异常值年份序号序列一样,因此仍然用受灾率的序列即可.初始序列如下:

$$\begin{cases} D_1^{(0)} = \{9, 10, 11, 12, 14, 15, 18, 20, 23, 24, 25, 26\} \\ D_2^{(0)} = \{14, 18, 24, 26\}, \end{cases} \quad (8)$$

根据 GM(1,1) 的方法,对上述序列建立预测模型:

$$\begin{aligned}\hat{d}_1(k+1) &= 105.241\ 025\ 3(1 - e^{-0.096\ 507\ 629})e^{0.096\ 507\ 629k} \\ \hat{d}_2(k+1) &= 101.25(1 - e^{-0.170\ 961\ 659})e^{0.170\ 961\ 659k}\end{aligned}\quad (9)$$

据公式(9),计算可得未来 15 年湖南省洪涝灾害预测结果,见表 3.

表 3 湖南洪涝灾害在未来 15 年重灾年份预测
Table 3 The forecast of flood disaster in Hunan province in the next 15 years

阈值	预测年份	序号	预测值	误差值/%	相对误差/%
$\xi > 0$	2018	34	33.948 8	0.051 2	0.150 6
	2021	37	37.388 4	-0.388 4	-1.049 8
	2025	41	41.176 5	-0.176 5	-0.430 6
	2029	45	45.348 5	-0.348 5	-0.774 4
$\xi > 0.8$	2016	32	31.527 5	0.472 5	0.150 6
	2028	44	44.379 6	-0.379 8	-1.049 8

通过后验差检验后,两个模型符合标准.并且,发展系数 a 满足 $-a \leq 0.3$ 时^[14],预测精度可达 97% 以上,可用于中长期预测.

3 改进的预测模型的应用

3.1 灰色残差模型

对残差序列 $\varepsilon_1^{(0)}(k) = \{\varepsilon_1^{(0)}(2), \varepsilon_1^{(0)}(3), \dots, \varepsilon_1^{(0)}(n)\}$ 作非负处理^[15],得序列

$$\varepsilon(k) = \{\varepsilon(2), \varepsilon(3), \dots, \varepsilon(n)\} \quad (10)$$

其中 $\varepsilon(k) = |\varepsilon_1^{(0)}(k)|$.对序列(10)的数据进行 GM(1,1,)建模得到模型

$$\varepsilon_1(k+1) = 0.095\ 849\ 207(1 - e^{-0.100\ 759\ 126})e^{0.100\ 759\ 126k} \quad (11)$$

用以修正模型 $\hat{d}_1(k)$,得灰色残差模型 $d_{1\varepsilon}(k+1) = \hat{d}_1(k+1) \pm \varepsilon_1(k+1), k = 2, 3, \dots, n$.

3.2 马尔科夫修正的灰色模型

灰色模型预测连续稳定的数据较为准确,否

则其预测精度会降低,而马尔科夫链理论大多用于分析有较大波动的数据.因此,将两者结合互补缺漏^[15].为了可以更准确地预测未来湖南洪涝灾害的趋势,引入马尔科夫过程进行改进,以 $D_1^{(0)}$ 序列模型为例:1)确定残差状态:残差为正表示 1 状态,为负表示 2 状态;2)根据表 4 数据得出状态转移概率矩阵 $P' = \begin{bmatrix} 3/4 & 1/4 \\ 1/5 & 4/5 \end{bmatrix}$;3)根据表 4 知初始状态向量 $V(0) = [0, 1]$,由状态转移公式 $V(t) = V(0) \cdot P'^t$,则有当 $t = 1$ 时, $V(1) = [0, 1] \begin{bmatrix} 3/4 & 1/4 \\ 1/5 & 4/5 \end{bmatrix} = [1/5, 4/5]$,取概率大的状态,即下一个值的残差为负.若有概率相同的情况,一般取上次的符号.

表 4 灰色预测模型分析表

Table 4 Grey prediction model analysis

序号	原始值	预测值	残差	状态
1	9			
2	10	10.662 8	-0.662 8	2
3	11	11.743 1	-0.743 1	2
4	12	12.932 9	-0.932 9	2
5	14	14.243 3	-0.243 3	2
6	15	15.686 4	-0.686 4	2
7	18	17.275 7	0.724 3	1
8	20	19.026 0	0.974 0	1
9	23	20.953 7	2.046 3	1
10	24	23.076 7	0.923 3	1
11	25	25.414 8	-0.414 8	2

以上数据分析得出修正后的预测结果,如表 5 所示.修正前的相对误差约为 -7.652 9%,经过修正后的预测值约为 -7.534 5%,比修正前的预测值精确了 0.118 4 个百分点.因此,经马尔科夫链修正的灰色模型更具准确性.

表 5 修正后的预测结果

Table 5 Revised prediction results

序号	实际值	灰色 GM(1,1) 预测值	残差值 GM(1,1) 预测值	残差修正符号 修正后结果	修正后结果	修正前相对误差/%	修正后相对误差/%
12	26	27.989 7	0.030 8	-	27.959 0	-7.652 9	-7.534 5

4 结 论

经研究表明,灰色 GM(1,1) 可以预测湖南洪

涝灾害在未来 15 年的灾害发生时间,且其结果属于一级,效果较好.经残差和马尔科夫改进后的模型,精度提高,得出较好的结果.相较于 2008 年的

湖南省特大洪涝灾害,未来 15 年中不会有比其更重的灾情,但灾害发生受多种因素影响,本文所预测结果仅可以提供一定的理论参考.

参考文献:

- [1] TINGSANCHALI T. Urban flood disaster management [J]. Procedia engineering, 2012, 32(3): 25-37.
- [2] LI X, WANG X, SHAO W, et al. Forecast of flood in Chaohu lake basin of China based on Grey-Markov theory [J]. Chinese geographical science, 2007, 17(1): 64-68.
- [3] ZHANG J, WU B, LI Z, et al. Intelligent geographic information systems for flood disaster evaluation and prediction [J]. International journal of advancements in computing technology, 2013, 5(2): 579-586.
- [4] 朱育红, 吕翠兰, 陈云虎, 等. 浅谈淮河流域洪水预报及洪水资源化 [C] // 第 33 届中国气象学会年会 S9 水文气象灾害预报预警. 西安: 中国气象学会, 2016: 306-307.
- [5] 邓聚龙. 灰理论基础 [M]. 武汉: 华中科技大学出版社, 2002: 221-233.
- [6] SUN B J, YIN W S. A modified grey model and its application in short time traffic flow [J]. Mathematics in practice and theory, 2016, 46(23): 201.
- [7] 夏昌浩, 曹瑾, 张密, 等. 电力负荷趋势外推预测算例分析与模型检验 [J]. 中国科技信息, 2016(21): 90-92.
- [8] 陈红保. 基于灰色模型的电网负荷预测方法研究 [D]. 保定: 华北电力大学, 2016.
- [9] 宁宣熙, 刘思峰. 管理预测与决策方法 [M]. 2 版. 北京: 科学出版社, 2009: 113-145.
- [10] 龚日朝, 刘礼仁, 张钰玲, 等. 湖南洪涝灾害波动特征及趋势预测研究 [J]. 热带地理, 2010, 30(3): 284-288.
- [11] 闫淑春. 2008 年全国洪涝灾情 [J]. 中国防汛抗旱, 2009, 20(1): 60-67.
- [12] 杨方, 李茂松, 王春艳, 等. 全国及区域尺度上农业旱灾受灾率分级研究 [J]. 灾害学, 2014, 29(4): 209-214.
- [13] 俞树荣, 韩竣羽, 李淑欣. 利用灰色马尔科夫模型预测腐蚀管道寿命 [J]. 机械强度, 2016, 38(4): 851.
- [14] 张剑明, 居晶琳, 黎祖贤, 等. 湖南省农业旱灾的年际变化及重灾年份预测 [J]. 自然灾害学报, 2011, 20(5): 105-110.
- [15] 陈勤. 基于马尔科夫灰色残差 GM(1,1) 模型的火灾损失预测 [J]. 工业安全与环保, 2015, 2015, 41(1): 70-72.