

# GM(1,1)模型在洞庭湖区濠口试点血吸虫病发病预测中的应用

赛晓勇<sup>1</sup>, 闫永平<sup>1</sup>, 张治英<sup>1</sup>, 蔡凯平<sup>2</sup>, 李岳生<sup>2</sup>, 周晓农<sup>3</sup>, 夏结来<sup>4</sup>, 徐德忠<sup>1</sup>

**【摘要】 目的** 对洞庭湖区退田还湖地区濠口试点血吸虫病发病进行预测,为国家卫生机构合理分配卫生资源提供决策依据。**方法** 应用 GM(1,1)模型对濠口试点血吸虫病患病率建模并进行残差修正,进行三年预测。**结果** 濠口试点 GM(1,1)模型平均相对误差为 10.67%,模型精度为差;残差 GM(1,1)模型平均相对误差为 4.73%,模型精度为优。残差修正预测模型为  $\hat{x}^{(1)}(k+1) = 1.189\ 654e^{0.272\ 225k} - 0.311\ 351$ ,连续三年预测值分别为 9.12%,9.70%和 10.48%。**结论** 濠口试点残差 GM(1,1)模型预测效果好;血吸虫病发病在未来三年内有缓慢上升的趋势,应加强血吸虫病防治工作。

**【关键词】** 模型;统计学;预测;血吸虫病

**【中图分类号】** R195.1; R532.21 **【文献标识码】** A **【文章编号】** 1008-6013(2005)01-0029-03

**Application of grey model (1,1) in the prediction of schistosomiasis prevalence in Haokou village of Dongting lake** SAI Xiao-yong<sup>1</sup>, YAN Yong-ping<sup>1</sup>, ZHANG Zhi-ying<sup>1</sup>, CAI Kai-ping<sup>2</sup>, LI Yue-sheng<sup>2</sup>, ZHOU Xiao-nong<sup>3</sup>, XIA Jie-lai<sup>4</sup>, XU De-zhong<sup>1</sup>. 1. Department of Epidemiology, Faculty of Preventive Medicine, Fourth Military Medical University, Xi'an 710033, China; 2. Institute of Anti-epidemic of Schistosomiasis, Yueyang 414000, China; 3. Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention, Shanghai 200025, China; 4. Department of Statistics, Faculty of Preventive Medicine, Fourth Military Medical University, Xi'an 710033, China

**【Abstract】 Objective** To forecast the schistosomiasis prevalence in Haokou village of "breaking dikes or opening sluice for waterstore" in Dongting Lake and to provide reference for the government. **Methods** Grey model (1,1) was applied and residual was ameliorated. **Results** Average relative error of schistosomiasis prevalence of grey model (1,1) in Haokou village was 10.67%, model accuracy was not good. Average relative error of ameliorated grey model (1,1) was 4.73%, model accuracy was good. Residual-ameliorated model was  $\hat{x}^{(1)}(k+1) = 1.189\ 654e^{0.272\ 225k} - 0.311\ 351$ . The predicted values of the following three years were 9.12%, 9.70% and 10.48% respectively. **Conclusions** The prediction effect of residual-ameliorated model was good. Schistosomiasis prevalence in Haokou village would increase in the following three years. Schistosomiasis prevention should be reinforced.

**【Key words】** Models; Statistical; Forecasting; Schistosomiasis

(Chin J Dis Control Prev 2005,9(1):29-31)

血吸虫病的传播过程是一个多因素、多层次的复杂系统,既有已为人知的白色信息,又有尚未发现

**【通讯作者】** 徐德忠,电话:029-83374955

的黑色信息,更多的是人们不完全清楚的灰色信息,属于典型的灰色系统。笔者应用 GM(1,1)模型对湖南洞庭湖区退田还湖濠口试点的病情资料进行分析并预测,以期为国家卫生机构合理分配卫生资源提供决策依据。

## 1 材料和方法

**1.1 材料** 收集洞庭湖区退田还湖地区澧县的濠口试点 1998~2003 年连续 6 年的粪检血吸虫卵阳性资料(资料全,随机抽样调查而来),该试点为单退点之一,退人不退田。即洪水期人转移,洪水过后返

**【基金项目】** 国家“十五”科技攻关课题(2001BA705B08)

**【作者单位】** <sup>1</sup>第四军医大学预防医学系流行病学教研室,陕西 西安 710033

<sup>2</sup>湖南省血吸虫病防治研究所,湖南 岳阳 414000

<sup>3</sup>中国疾病预防控制中心寄生虫病预防控制所,上海 200025

<sup>4</sup>第四军医大学预防医学系统计学教研室,陕西 西安 710033

**【作者简介】** 赛晓勇(1974-),男,回族,河南新乡人,讲师,硕士。主要研究方向:流行病学。

回种田。濠口试点常住人口 1 176 人,面积为 297 万 m<sup>2</sup>,为湖南省血吸虫病重灾区监测试点之一。资料由湖南省血防所及小渡口血防站提供。

1.2 方法

1.2.1 模型建立 假定给定时间数据序列 X<sup>(0)</sup> 有 n 个值, X<sup>(0)</sup> = { X<sup>(0)</sup>(k), k = 1, 2, ..., n }, 作相应的 1 阶累加序列 X<sup>(1)</sup> = { X<sup>(1)</sup>(k), k = 1, 2, ..., n }, 则序列 { X<sup>(1)</sup>(k), k = 1, 2, ..., n } 的 GM(1, 1) 模型的白化微分方程为: dX<sup>(1)</sup>/dt + aX<sup>(1)</sup> = μ, 式中: a 为发展灰数; μ 为内生控制灰数。设 â 为待估参数向量。

$$\hat{a} = \begin{pmatrix} a \\ \mu \end{pmatrix} = (B^T B)^{-1} B^T Y_n,$$
$$B = \begin{pmatrix} -\frac{1}{2}[X^{(1)}(1) + X^{(1)}(2)] & 1 \\ -\frac{1}{2}[X^{(1)}(2) + X^{(1)}(3)] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}[X^{(1)}(n-1) + X^{(1)}(n)] & 1 \end{pmatrix},$$
$$Y_n = \begin{pmatrix} X^{(0)}(2) \\ X^{(0)}(3) \\ \vdots \\ X^{(0)}(n) \end{pmatrix},$$
求解微分方程可得预测模型:

$$X^{(1)}(k+1) = [X^{(0)}(1) - \frac{\mu}{a}]e^{-ak} + \frac{\mu}{a} \quad (k=0, 1, 2, \dots, n).$$

1.2.2 模型检验 包括残差检验、关联度检验和后验差检验。残差检验是按预测模型计算 X<sup>(1)</sup>(i), 并将 X<sup>(1)</sup>(i) 累减生成 X<sup>(0)</sup>(i), 然后计算原始序列 X<sup>(0)</sup>(i) 与 X<sup>(0)</sup>(i) 的绝对误差序列及相对误差序列。△<sup>(0)</sup>(i) = |X<sup>(0)</sup>(i) - X<sup>(0)</sup>(i)| (i = 1, 2, ..., n); Φ(i) =  $\frac{\Delta^{(0)}(i)}{X^{(0)}(i)} \times 100\%$  (i = 1, 2, ..., n)。绝对误差越小越好, 相对误差一般认为小于 0.5% 为好。关联度检验是根据 X<sup>(0)</sup>(i) 与原始序列 X<sup>(0)</sup>(i) 的关联系数计算出关联度, 当 ρ = 0.5 时一般认为 r 大于 0.6。后验差检验需计算原始序列的标准差 S<sub>1</sub> =  $\sqrt{\frac{\sum [X^{(0)}(i) - \bar{X}^{(0)}]^2}{n-1}}$  和绝对误差序列的标准差 S<sub>2</sub> =  $\sqrt{\frac{\sum [\Delta^{(0)}(i) - \bar{\Delta}^{(0)}]^2}{n-1}}$ ; 然后计算方差比和小误差概率: C = S<sub>2</sub>/S<sub>1</sub>, P = p { |△<sup>(0)</sup>(i) -  $\bar{\Delta}^{(0)}$ | < 0.674 5 S<sub>1</sub> }, 令 e<sub>i</sub> = |△<sup>(0)</sup>(i) -  $\bar{\Delta}^{(0)}$ |, S<sub>0</sub> = 0.647 5 S<sub>1</sub>, 则 P = p { e<sub>i</sub> < S<sub>0</sub> }。判断标准见表 1, 若残差检验、关联度检验和后验差检验都能通过, 可使用该模型预测, 否则进行残差修正。

表 1 模型检验后精度等级

| Table 1 Accuracy level after model test |                |                |
|---|----------------|----------------|
| Accuracy level                          | C              | P              |
| Better                                  | C < 0.35       | P ≥ 0.95       |
| Good                                    | 0.35 ≤ C < 0.5 | 0.8 ≤ P < 0.95 |
| Qualified                               | 0.5 ≤ C < 0.65 | 0.7 ≤ P < 0.8  |
| Not qualified                           | C ≥ 0.65       | P < 0.7        |

1.2.3 残差修正 设原始序列 X<sup>(0)</sup> 建立的 GM(1, 1) 模型为 X<sup>(1)</sup>(i+1) = [X<sup>(0)</sup>(1) -  $\frac{\mu}{a}$ ]e<sup>-ai</sup> +  $\frac{\mu}{a}$ , 可获得生成序列 X<sup>(1)</sup> 的预测值 X<sup>(1)</sup>, 定义残差为: e<sup>(0)</sup>(j) = X<sup>(1)</sup>(j) - X<sup>(1)</sup>(j), 若取 j = i, i+1, ..., n, 则与 X<sup>(1)</sup> 及 X<sup>(1)</sup> 对应的残差序列为: e<sup>(0)</sup> = { e<sup>(0)</sup>(i), e<sup>(0)</sup>(i+1), ..., e<sup>(0)</sup>(n) }, 经转化 e<sup>(1)</sup> 可建立相应的 GM(1, 1) 模型:  $\hat{e}^{(1)}(k+1) = [e^{(0)}(1) - \frac{\mu_e}{a_e}]e^{-a_e k} + \frac{\mu_e}{a_e}$ ,  $\hat{e}^{(1)}$ (k+1) 的导数加上  $\hat{e}^{(1)}(k+1)$  修正 X<sup>(1)</sup>(k+1) 得到修正模型: X<sup>(1)</sup>(k+1) = [X<sup>(0)</sup>(1) -  $\frac{\mu}{a}$ ]e<sup>-ak</sup> +  $\frac{\mu}{a}$  + δ(k-1)(-a<sub>e</sub>)[e<sup>(0)</sup>(1) -  $\frac{\mu_e}{a_e}$ ]e<sup>-a<sub>e</sub>(k-1)</sup>, 其中 δ(k-1) 为修正系数。最后经残差修正的原始序列预测模型: X<sup>(0)</sup>(k+1) = X<sup>(1)</sup>(k+1) -  $\hat{x}^{(1)}(k)$ , (k = 1, 2, ...)。

2 结果

2.1 建模 收集研究地区发病资料如表 2 示。代入公式可得 a = 0.023 196; μ = 9.027 123。GM(1, 1) 预测模型为 X<sup>(1)</sup>(k+1) = -384.276 538 e<sup>-0.023 196k</sup> + 389.166 538

表 2 澧县濠口试点 1998~2003 年粪检阳性率(%)

| Table 2 Positive rate of fecal test in Haokou village, 1998~2003 |      |      |      |      |      |      |
|--|------|------|------|------|------|------|
|  | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 |
| Positive rate(%)   | 4.89 | 7.72 | 9.52 | 9.76 | 7.28 | 7.80 |

2.2 检验

2.2.1 残差检验 如表 3 示, 绝对误差及相对误差较大, 拟合效果不好, 需修正。

表 3 GM(1, 1) 预测模型绝对误差序列及相对误差

| Table 3 Absolute error and relative error of GM(1, 1) |                |              |                |                   |
|---|----------------|--------------|----------------|-------------------|
| No.   | Observed value | Fitted value | Absolute error | Relative error(%) |
| X(2)  | 7.72           | 8.811108     | -1.091108      | -14.133337        |
| X(3)  | 9.52           | 8.609077     | 0.910923       | 9.568417          |
| X(4)  | 9.76           | 8.411679     | 1.348321       | 13.814623         |
| X(5)  | 7.28           | 8.218807     | -0.938807      | -12.895522        |
| X(6)  | 7.80           | 8.030357     | -0.230357      | -2.953260         |

2.2.2 关联度检验  $r = \frac{1}{n} \sum_{i=1}^n \eta(i) = 0.550 6$ , 可以

认为在  $\rho=0.5$  时  $r>0.6$ 。

**2.2.3 后验差检验**  $C=0.607\ 1$ ,  $P=0.6$ , 未通过, 需残差修正。

**2.3 残差修正** 经两次残差修正,  $\alpha=-0.272\ 225$ ,  $\mu=0.084\ 757$ 。GM(1, 1)修正模型为  $X^{(1)}(k+1)=1.189\ 654e^{0.272\ 225k}-0.311\ 351$ ,  $C=0.249\ 3$ ,  $P=1.0000$ 。

表 4 GM(1,1)残差修正模型绝对误差序列及相对误差

Table 4 Absolute error and relative error of ameliorated GM(1,1)

| No.  | Observed value | Fitted value | Absolute error | Relative error(%) |
|------|----------------|--------------|----------------|-------------------|
| X(2) | 7.72           | 8.039092     | -0.319092      | -4.133261         |
| X(3) | 9.52           | 10.008684    | -0.488684      | -5.133176         |
| X(4) | 9.76           | 9.268647     | 0.4913530      | 5.034307          |
| X(5) | 7.28           | 6.984818     | 0.2951820      | 4.054642          |
| X(6) | 7.80           | 8.215620     | -0.415620      | -5.328396         |

**2.4 预测** 根据修正模型可得 2004 年、2005 年和 2006 年澧县濠口试点连续三年粪检阳性率预测值为 9.12%, 9.70% 和 10.48%。

3 讨论

1998 年我国实行退田还湖以来, 血吸虫病的传播更加复杂<sup>[1]</sup>, 为了阐明退田还湖对血吸虫病传播的影响, 笔者对洞庭湖区退田还湖试点的病情与螺情资料进行分析, 发现了血吸虫病及活螺密度退田还湖前后的变化趋势, 运用时间序列诸方法进行了比较<sup>[2,3]</sup>, 为进一步监测, 建立和完善血吸虫病预测模型系统, 运用了灰色预测。

灰色系统理论于 20 世纪 80 年代初创立, 其核心为灰色动态系统, 现广泛应用于疾病(传染病、寄生虫病, 特别是恶性肿瘤等)的预测<sup>[4~9]</sup>。GM(1, 1)模型是灰色动态模型中应用最广泛的预测模型, 与传统的数理统计模型相比, 有无需典型的概率分布、减少时间序列的随机性、小样本即可计算和计算简便的优点。国内有学者对血吸虫病进行研究并给出了经验公式, 如蔡碧等<sup>[10]</sup>在已有规范化数据的基础上得出人群血吸虫病患病率(%)的预测模型为  $X^{(0)}(k+1)=8.047\ 036e^{-0.089\ 722\ 67k}$ , 又如吴锦华等<sup>[11]</sup>将环境干涉因子引入修正预测结果进行预测。

灰色预测不需要: 原始数据建模、足够的样本和典型的概率分布; 适用性较强、预测性能好。适用于单调性强特别是具有累加型的资料, 其拟合效果与预测效果均较好, 适于近期预测。灰色预测亦有其局限性如未考虑社会因素、不适于长期预测等, 找到完备的公式还有一定距离。

本研究结果显示, 若用单纯的 GM(1, 1)模型, 则精度不高(平均相对误差为 10.67%, 后验差检验显示模型精度为差)。但经过两次残差修正后, 得到的残差 GM(1, 1)模型的精度有所提高(平均相对误差为 4.73%, 后验差检验显示模型精度为优), 因而该模型的预测性能较好。应用该模型预测澧县濠口试点 2004 至 2006 年连续三年粪检阳性率发现有逐年升高的趋势, 提示当地血吸虫病防治工作不容乐观。

【参考文献】

[1] 赛晓勇, 张治英, 徐德忠, 等. 退田还湖对生态环境及对血吸虫病流行的影响 [J]. 中国公共卫生, 2004, 20(2): 237-239.

[2] 赛晓勇, 蔡凯平, 徐德忠, 等. 洞庭湖区退田还湖试点 1990/2002 血吸虫病情与螺情分析 [J]. 第四军医大学学报, 2003, 24(20): 1878-1880.

[3] 赛晓勇, 张治英, 徐德忠, 等. 时间序列分析在洞庭湖区双退试点血吸虫病发病预测中的应用 [J]. 第四军医大学学报, 2003, 24(24): 2297-2300.

[4] 齐显影. GM(1, 1)模型在传染病预测中的应用 [J]. 疾病控制杂志, 1999, 3(3): 235.

[5] 蔡碧, 李建屏, 任先平, 等. 社会经济因素影响血吸虫病流行的灰色关联分析 [J]. 医学与社会, 1999, 12(4): 20-22.

[6] 翁寿清, 徐校平, 阮玉华, 等. 建德市肾综合征出血热 GM(1, 1)模型预测研究 [J]. 中国公共卫生, 1999, 15(7): 645-646.

[7] 段琼红, 聂绍发, 仇成轩, 等. 应用 BP 神经网络预测前列腺癌流行趋势 [J]. 中国公共卫生, 2000, 16(3): 193-195.

[8] 凌莉, 柳青, 骆福添, 等. 社区肿瘤发病和死亡资料趋势预测 [J]. 中国肿瘤, 2002, 11(3): 149-151.

[9] 金永富, 赵玉婉, 裴炯良, 等. 舟山结核病流行趋势的灰色模型分析 [J]. 疾病控制杂志, 2003, 7(5): 472-473.

[10] 蔡碧, 任先平, 李建屏, 等. 血吸虫病预测模型系统的设计 [J]. 预防医学情报杂志, 1999, 15(4): 228-230.

[11] 吴锦华, 蔡碧, 任先平, 等. 血吸虫病疫情的中长期预测. 湖北预防医学杂志 [J], 1999, 10(5): 4-6.

(收稿日期 2004-04-09)  
(刘慧慧校)