血吸虫病蔓延周期的灰色预测模型°

陈业华[1]　高遵海[1]　周启予[2]

[1](湖北荆州师范高等专科学校数学系, 湖北 荆州 434100)
[2](湖北荆州血防研究所, 湖北 荆州 434100)

摘要　对血吸虫病蔓延高峰期的预测具有深远的意义。本文建立一种新的灰色预测模型, 对该病蔓延发展的高峰期进行预测, 通过实例计算分析, 预测结果正确, 模型拟合精度为优。

关键词　灰色模型　差分格式　阈值

The Gray Prediction Model for the Spread Period of Schistosomiasis

Chen Yehua[1]　Gao Zunhai[1]　Zhou Qiyu[2]

[1](Hubei Jingzhou Teacher's College, Jingzhou Hubei 434100)
[2](Hubei Jingzhou Schistosomiasis Prevention Institute, Jingzhou Hubei 434100)

Abstract　It is of great importance to predict the spreading peak of schistosomiasis. This paper establishes a difference gray prediction model, which forecasts the peak peroid of schistosomiasis growing and spreading. A practical example is caculated and analyzed, which indicates that the prediction result is correct and the precision grade of the model is excellent.

Keywords　grey model; difference scheme; valve value

1　前言

自 1956 年毛主席代表党和国家发出"一定要根治血吸虫病"的号召以来, 全国各大疫区已收到很好的效果, 但该病非常顽固, 容易复发, 特别是通过人—畜—虫—螺—虫—人的恶性循环, 某些地区疫情仍然非常严重, 给政府带来很大的负担, 给人民带来很大的痛苦。根治血吸虫病再来一次过去的"人民战争", 目前是不可能的, 必须采取科学的预测, 合理的投资, 准确的防治, 才能达到最少消耗的防治和根除该病的目的。血吸虫病的蔓延周期是一种非线性模糊概念。病源的主要宿体是钉螺。国家每年用于灭螺的投资总额达 1 亿多元, 有的地区投资带有盲目性, 浪费人力、物力、财力, 造成很大的损失。若能预测到某地区该病蔓延的高峰期, 然后合理的分配投资, 将减少不必要的损失, 达到准确防治的目的。本文建立一种新的灰色预测差分模型, 能较为准确地预测疫病蔓延的高峰期, 为政府部门对血防的投资提供依据。实例计算分析, 模型拟合效果良好, 预测结果正确。

2　灰色系统的双向差分法

设原始数列 $x^{(0)}(t)=\{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)\}$ 的 1-AGO 生成数列为 $x^{(1)}(t)$, 其灰色模型为

$$dx^{(1)}(t)/dt = a + bx^{(1)}(t) \tag{1}$$

其中 a, b 为预测估计参数。现选择等时间隔取样, 令 $\Delta t = t+1-t$, 将方程(1)表示为如下差分格式

° 本文于 1997 年 6 月 23 日收到

$$\Delta x^{(1)}(t) = (a + bx^{(1)}(t))\Delta t \tag{2}$$

由Burg 时间序列估计原理[2], 将数据倒序使用一次生成以提高信息的利用率, 考虑式(2)的向前向后差分

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

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

由此而得到方程(1)的向前向后差分格式为

$$\alpha_l(K) = \Delta_L x^{(1)}(k) - bx^{(1)}(k) - a \tag{3}$$

$$\alpha_r(k) = \Delta_r x^{(1)}(k) - bx^{(1)}(k) - a \tag{4}$$

其中 $k = 2, 3, \cdots, n-1$。由于预测模型拟合精度, 决定于双向差分平方总和达到最小时 a, b 的值, 因此, 取式(3), (4)的平方总和 S, 并使

$$S = \sum_{k=2}^{n-1} (\alpha_l^2(k) + \alpha_r^2(k)) = \min$$

这里 $x^{(0)}(1) = x^{(1)}(1) = 1$ 为初始值。由拉格朗日乘数法, 令 $\frac{\partial s}{\partial a} = 0$, $\frac{\partial s}{\partial b} = 0$, 并联立即得方程

$$\begin{cases} \sum_{k=2}^{n-1} x^{(0)}(k) + \sum_{k=2}^{n-1} x^{(0)}(k+1) = 2a(n-2) + 2b\sum_{k=2}^{n-1} x^{(1)}(k) \\ \sum_{k=2}^{n-1} x^{(0)}(k)x^{(1)}(k) + \sum_{k=2}^{n-1} x^{(0)}(k+1)x^{(1)}(k) = 2a\sum_{k=2}^{n-1} x^{(1)}(k) + 2b\sum_{k=2}^{n-1} (x^{(1)}(k))^2 \end{cases} \tag{5}$$

用 $n-2\,(n>2)$ 分别同时除方程(5)的两边, 得

$$\begin{cases} aX_1 + bY_1 = C_1 \\ aY_1 + bX_2 = C_2 \end{cases} \tag{6}$$

其中

$$\begin{cases} C_1 = \frac{1}{2(n-2)}\left(\sum_{k=2}^{n-1} x^{(0)}(k) + \sum_{k=2}^{n-1} x^{(0)}(k+1)\right) \\ C_2 = \frac{1}{2(n-2)}\left(\sum_{k=2}^{n-1} x^{(0)}(k)x^{(1)}(k) + \sum_{k=2}^{n-1} x^{(0)}(k+1)x^{(1)}(k)\right) \\ X_1 = \frac{1}{n-2}\sum_{k=2}^{n-1} 1^2 = 1, \quad X_2 = \frac{1}{n-2}\sum_{k=2}^{n-1} (x^{(1)}(k))^2 \\ Y_1 = \frac{1}{n-2}\sum_{k=2}^{n-1} x^{(1)}(k) \end{cases} \tag{7}$$

记 $\hat{A} = (\hat{a}, \hat{b})^T$, $C = (c_1, c_2)^T$, $B = \begin{pmatrix} X_1 & Y_1 \\ Y_1 & X_2 \end{pmatrix}$, 则方程(6)可写成 $B\hat{A} = C$。由最小二乘估计原理, 有

$$\hat{A} = B^{-1}C = (B^TB)^{-1}B^TC \tag{8}$$

将式(8)代入方程(1)求解即得关于时间响应的灰色差分模型

$$\hat{x}^{(1)}(k+1) = (x^{(0)}(1) + \hat{a}/\hat{b} - 2k)e^{\hat{b}k} - \hat{a}/\hat{b} \tag{9}$$

该模型相应的计算值 $\hat{x}^{(1)}(k)$ 累减生成还原, 即得原始数据相应的拟合值。其计算公式为

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k), \quad k = 0, 1, \cdots, n-1 \tag{10}$$

3 预测模型的建立

设 $x(t)$ 是某地区关于血吸虫病蔓延周期记载的原始数集, σ 为 $x(t)$ 中的某个异常阀值(零点值)。其中所有大于 σ 的点组成的子集称为 $x(t)$ 的疫变子集。在疫变子集中, 对异常点出现的时刻 t 作预测即称为疫变周期预测。

设原始数列 $x(t)$ 中的元素与其相应的时刻 t 构成一个二维平面, 记 $a_t(t, x(t))$ 为该平面上的点, P 为该平面上某点的水平投影算子, 即

$$P(a_t) = P(t, x(t)) = t$$

所有这些投影算子构成的集$\{P(a_1), P(a_2), \cdots, P(a_n)\}$称为$x(t)$的序列。疫变子集所对应的序列即称为$x(t)$的疫变序列。

取阈值σ, 设$x(t)$构成的疫变子集为$x_\sigma(\hat{t})$, $x_\sigma(\hat{t}) \in \{x(t)\}$, $\hat{t} \in \{P(a_t)\}$, 于是$x_\sigma(\hat{t})$的序列为

$$\{P(\hat{1}, x_\sigma(\hat{1})), P(\hat{2}, x_\sigma(\hat{2})), \cdots, P(\hat{m}, x_\sigma(\hat{m}))\} = \{\hat{1}, \hat{2}, \cdots, \hat{m}\}$$

其中$\hat{m} \leq n$。可以看出, 由阈值σ建立了一个$\{x(t)\}$到$\{x_\sigma(\hat{t})\}$的映射:

$$\sigma: \{x(t)\} \rightarrow \{x_\sigma(\hat{t}\}, \quad t \rightarrow \hat{t}$$

对新的二元组$(\hat{t}, x_\sigma(\hat{t}))$, 记$x_\sigma^{(0)}(\hat{t}) = P(\hat{t}, x_\sigma(\hat{t}))$, 于是得疫变序列

$$x_\sigma^{(0)}(k) = \{x_\sigma^{(0)}(\hat{1}), x_\sigma^{(0)}(\hat{2}), \cdots, x_\sigma^{(0)}(\hat{m})\} = \{\hat{1}, \hat{2}, \cdots, \hat{m}\}$$

对疫变序列$x_\sigma^{(0)}(k)$实施一次累加得到 1- AGO 生成序列$x_\sigma^{(1)}(K)$, 根据式(9)的推导过程即可推得疫变周期预测模型

$$\hat{x}_\sigma^{(1)}(k+1) = (x_\sigma^{(0)}(\hat{1}) + \hat{a}/\hat{b} - 2k)e^{\hat{b}k} - \hat{a}/\hat{b} \tag{11}$$

其中$k = 0, 1, \cdots, \hat{m} - 1$, $\hat{a}$, $\hat{b}$满足式(7)。模型(11)的计算值$x_\sigma^{(1)}(k)$累减生成还原, 即得到疫变序列值相应的拟合值$\hat{x}_\sigma^{(0)}(k)$。其计算公式同式(10)。

4　预测实例

以疫情严重的湖北荆州市国营太湖农场为例, 对该农场血吸虫病蔓延周期进行预测。血吸虫病的蔓延因素是比较复杂的, 其综合考察的主要因素有: 血吸虫卵繁殖速率, 人体感染的比例、农畜(牛、羊等)感染的比例, 丁螺生长密度, 被感染的人、畜排泄废物的处理等。太湖农场血防所记载的 1950 年至 1985 年综合数据如表 1。

表 1　太湖农场 1950～1985 年综合递增率(%)表

年份	递增率	年份	递增率	年份	递增率	年份	递增率	年份	递增率	年份	递增率
1950	4.9722	1956	7.6253	1962	8.7903	1968	7.0316	1974	3.3011	1980	3.1327
1951	7.9276	1957	8.2072	1963	3.7825	1969	4.1721	1975	6.5467	1981	4.0121
1952	4.8613	1958	3.6794	1964	7.3927	1970	3.6081	1976	4.2141	1982	5.7609
1953	8.4127	1959	9.1375	1965	4.3931	1971	3.7204	1977	3.1072	1983	3.1901
1954	4.7697	1960	4.9891	1966	3.6972	1972	6.3942	1978	3.0911	1984	2.9712
1955	4.9781	1961	4.7932	1967	4.0713	1973	3.1799	1979	6.0982	1985	2.7211

根据表 1, 设原始数列$x(t) = \{x(1950), x(1951), \cdots, x(1985)\}$。由血防部门经验总结, 一般在该地区调查的综合递增率高于 5% 以上, 即认为该地区血吸虫病蔓延较为严重。因此, 选择阈值$\sigma = 0.05$。由表 1 知疫变子集$x_\sigma(\hat{t}) = \{7.9276, 8.4127, 7.6253, 8.2072, 9.1375, 8.7903, 7.3927, 7.0316, 6.3942, 6.5467, 6.0982, 5.7609\}$, 其相应的疫变序列$x_\sigma^{(0)}(k) = \{2, 4, 7, 8, 10, 13, 15, 19, 23, 26, 30, 33\}$, 则疫变序列的 1-AGO 生成序列$x_\sigma^{(1)}(k) = \{2, 6, 13, 21, 31, 44, 59, 78, 101, 127, 157, 190\}$。由算式(7)计算得

$$C_1 = \frac{1}{20}[155 + 184] = 16.95$$

$$C_2 = \frac{1}{20}[13867 + 16615] = 1524.1$$

$$X_2 = \frac{1}{10} \times 64087 = 6408.7, \quad Y_1 = \frac{1}{10} \times 637 = 63.7$$

则

$$B = \begin{pmatrix} 1 & 63.7 \\ 63.7 & 6408.7 \end{pmatrix} \quad C = \begin{pmatrix} 16.95 \\ 1524.1 \end{pmatrix}$$

由式(8)即得

$$\hat{A}=\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix}=\begin{pmatrix}2.7259 & -0.0271\\ -0.0271 & 0.0004\end{pmatrix}\begin{pmatrix}16.95\\ 1524.1\end{pmatrix}=\begin{pmatrix}4.9009\\ 0.1503\end{pmatrix}$$

将以上数据代入模型(11)即得疫变周期预测的计算公式

$$\hat{x}_\sigma^{(1)}(k+1)=(34.6075-2k)e^{0.1503k}-32.6075$$

又由文献[1]知, 灰色模型拟合程度的残差 $\epsilon(k)=x_\sigma^{(0)}(k)-\hat{x}_\sigma^{(0)}(k)$, 相对误差 $\delta(k)=\left|(x^{(0)}(k)-\hat{x}^{(0)}(k))/x_\sigma^{(0)}(k)\right|$, 因此, 计算结果及精度如表 2。

由表 2 中的计算数据我们可以看到, 当相对误差 $\delta(k)<0.02$ 时, 其拟合效果为佳。表中有三行满足该条件, 即 $k=5,6,7$ 时, 拟合效果较好。当 $k=5$ 时, 拟合值为 12.7982, 则预测值为

$$1962+12.7982=1974.7982\approx 1975$$

该值说明, 太湖农场将于 1975 年出现血吸虫病蔓延高峰, 从表 1 可以看到, 1975 年综合递增率值为 6.5467%, 超过 0.05, 属疫变异常值。同理, 当 $k=6$ 时, 预测 1979 年将再出现蔓延高峰, 结果与表 1 中反映出的事实相符。当 $k=7$ 时, 预测太湖农场将于 1987 年再次出现蔓延高峰, 预测结果正确, 1987 年该农场血防所调查数据的综合递增率为 5.2764%。

表 2　血吸虫病蔓延高峰期预测拟合计算值及精度表

k 值	年份	$x_\sigma^{(0)}(k)$	$\hat{x}_\sigma^{(1)}(k+1)$	$\hat{x}_\sigma^{(0)}(k)$	$\epsilon(k)$	$\delta(k)$%
0	1951	2	2	3.2889	-1.2889	64.4450
1	1953	4	5.2889	3.4421	0.5579	13.9475
2	1956	7	8.7310	6.4022	0.5978	8.5400
3	1957	8	15.1332	8.3271	-0.3271	4.0889
4	1959	10	23.4603	11.7008	-1.7018	17.0180
5	1962	13	35.1611	12.7982	0.2018	1.5523
6	1964	15	47.9593	15.2907	-0.2907	1.9380
7	1968	19	63.2500	18.7211	0.2789	1.4679
8	1972	23	81.9711	21.8284	1.1716	5.0939
9	1975	26	103.7995	24.0902	1.9098	7.3454
10	1979	30	127.8897	27.3972	2.6028	8.6760
11	1982	33	155.2869	31.8725	1.1275	3.4167

5　结束语

文中根据邓聚龙先生提出的灰色理论, 利用双向差分法建立的灰色预测模型, 能较为准确地预测血吸虫病蔓延的高峰期。若适当地调整阀值 σ, 多次进行计算比较, 将会得到更佳的预测效果。但文中对模型的拟合精度及可靠性的分析没有涉及, 因有不少文献对精度的传统分析提出异议[4], 所以笔者将另行文探讨。文中提出的模型, 具有广泛性, 只须稍加改变, 即可用于其他方面的预测。

参 考 文 献

1　邓聚龙. 灰色预测与决策. 武汉: 华中工学院出版社, 1986
2　陈兆国. 时间序列及其谱分析. 北京: 科学出版社, 1988
3　杨秋明. 非线性灰色微分方程 $dx/dt+ax=b$ 的拟合. 应用数学, 1990, 3(3): 60～70
4　朱宝璋. 关于灰色系统基本方法的研究和评论. 系统工程理论与实践, 1994, 14(4): 52～60