手足口病传播的数学模型建立与分析

包 钰，施 昀，黄海洋，崔吟歌

（北京师范大学数学科学学院，北京 100875）

摘要：首先，根据传染病的传播特点，建立关于手足口病人率的常微分方程模型，并对出生率和超过6岁儿童的比率关系分别进行讨论。搜集2008年5月4日至5月11日传染病流行较严重的省份——安徽省比较准确的数据，用Matlab软件进行数值计算，以图形模拟方法求得*S-I-R*模型中的有关参数。得到手足口病人率比较符合实际数据的变化图，能正确地预测它的发展趋势。并导出手足口病人数关于时间的函数$I(t)$，利用相轨线理论对结果加以分析并作出预测。然后，建立时间序列的ARIMA模型。根据现有模型，利用时间序列软件EViews对数据进行拟合，得出模型的阶数和参数，对拟合结果进行相关性检验，验证模型的合理性。将拟合数据与原数据进行对比，证明拟合正确。并对未来几天可能的感染人数进行预测。最后根据预报得知，手足口疾病（hand-foot-and-mouth disease，HFMD）大概1个多月后，病情会有大大好转，而现实证明了这一点。

关键词：流行病与卫生统计学；手足口病；常微分方程；曲线拟合；时间序列；ARIMA模型

中图分类号：R181 **文献标识码**：A **文章编号**：1674-2850(2010)06-1157-7

Establishment and analysis of mathematical model of hand-foot-and-mouth disease

BAO Yu, SHI Yun, HUANG Haiyang, CUI Yinge

(*School of Mathematical Sciences, Beijing Normal University, Beijing* 100875, *China*)

Abstract: Firstly, according to the characteristics of infectious disease, this paper builds ordinary differential equation model of the patient rate of the hand-foot-and-mouth disease and discusses the relation between birth rate and the proportion of children over six years old. The data are from Anhui province where the disease is serious from May fifth to May eleventh in 2008. The research uses the math software Matlab to do numerical calculation and takes advantage of graph fitting to gain the parameter of *S-I-R* model and to get the changing graph which conforms to the fact and could forecast the trend. The work derives the function $I(t)$ of the number of hand-foot-and-mouth disease patients and makes a prediction. Secondly, ARIMA model of time series is established, and software Eviews is used to fit the data and calculate the degree and the parameter of the ARIMA model. The study tests the correlation to verify rationality of the model. Comparing the fitting data and the original data, the result proves that the fitting is valid and makes the forecast of the disease. Finally, based on the forecast, the condition of hand-foot-and-mouth disease would be better greatly after one month. The facts warrant the conclusion.

Key words: epidemiology and hygienic statistics; hand-foot-and-mouth disease; ordinary differential equation; curve fitting; time series; ARIMA model

0 引言

随着人类文明的不断进步和医学的不断发展，许多传染病都已得到有效的控制。手足口病虽然是一种很早就有的传染病，但是自从2008年以来，它的传播有愈演愈烈的趋势。因此，对手足口建立数学模型，有助于卫生部门对手足口病的现状作出合理判断，并能比较准确地预测这种病的趋势。手足

作者简介：包钰（1988—），女，本科生，主要研究方向：统计学

通信联系人：黄海洋，教授，主要研究方向：常微分方程. E-mail：hhywsg@bnu.edu.cn

口病是由肠道病毒引起的传染病，多发生于 5 岁以下儿童，并且一旦感染过一次手足口病之后就会获得终身免疫。

1 *S-I-R* 模型

1.1 模型的建立

1.1.1 模型假设

1）除感病特征之外，人群的个体间没有差异，感病者与易感病者的个体在人群中混合均匀；2）人群的数量足够大，只考虑传染过程的平均效应；3）易感者感病的机会与其他接触感病者的机会成正比；4）疾病的传染率为常数；5）感病痊愈者将进入移出者；6）超过 5 岁的人群不再感病；7）疾病流行过程中，人群的总数保持不变。

1.1.2 参数说明

$S(t)$ 为易感人群在 t 时刻占总人数的比例；$I(t)$ 为感病人群在 t 时刻占总人数的比例；$R(t)$ 为移出人群在 t 时刻占总人数的比例；k 为疾病的传染率；h 为感染者的日治愈率，即感染者每天被治愈的人数占感染者总人数的比例；I 为初始时刻（$t=0$），感染人群占总人口的比例；S 为初始时刻（$t=0$）健康人群占总人口的比例；R 为初始时刻病愈免疫者的比例，不妨设 $R_0=0$；a 为 6 岁以下儿童占的百分比；b 为人口的出生率；d 为单位时间内超过 6 岁的人口比率；N 为人群总数。

1.1.3 问题分析

由于手足口感病人群只在 5 岁以下，因而受到儿童人群总数的影响，与儿童的出生率和越过年龄界限的人数比率有关。并且手足口病有免疫效果，因而采用 *S-I-R* 的模型模拟。并且把传染病流行阶段的出生儿童和超过 6 岁界限的儿童人数考虑进去，通过对出生率和超过 6 岁界限儿童比率的比较，讨论不同情况下，手足口病的最终情况。

1.1.4 模型的阐述

由于假设总人口数在这段时间内不变，得到[1]

$$S(t)+R(t)+I(t)=a=\text{const},$$

其中，t 时刻总共有 $NS(t)$ 个易感者；每天一个患病者接触 k 个人，有 $k\cdot N\cdot a\cdot I(t)$ 个患病者，于是每天患病的易感者为 $kNaI(t)S(t)$，再加上这段时间出生的儿童人数 Nb，除去这段时间内过 6 岁生日的儿童人数 Nd，得到微分方程

$$Na\frac{\mathrm{d}S}{\mathrm{d}t}=-kNaIS+Nb-Nd,$$

化简得出

$$\frac{\mathrm{d}S}{\mathrm{d}t}=-kIS+\frac{b}{a}-\frac{d}{a}.$$

同理可以分别得到关于患病者和治愈者的微分方程：

$$\begin{cases}\dfrac{\mathrm{d}I}{\mathrm{d}t}=kIS-hI,\\[2ex] \dfrac{\mathrm{d}R}{\mathrm{d}t}=hI.\end{cases}$$

设初始时刻易感者的比例为 $S(0)=S_0(>0)$，感病者的比例为 $I(0)=I_0(>0)$，并且初始移出者 $R(0)=R_0=0$，就可以得到 *S-I-R* 模型的方程[2]：

$$\begin{cases}\dfrac{\mathrm{d}S}{\mathrm{d}t}=-kIS+\dfrac{b}{a}-\dfrac{d}{a},\ S(0)=S_0,\\ \dfrac{\mathrm{d}I}{\mathrm{d}t}=kIS-hI,\ I(0)=I_0,\\ \dfrac{\mathrm{d}R}{\mathrm{d}t}=hI,\ R(0)=R_0=0,\\ S(t)+R(t)+I(t)=a=\text{const}.\end{cases}\tag{1}$$

由于调查的时间段比较短，可以认为，在手足口病流行的这 2 个月中，儿童的人数没有太大变化。因而采用模型 $b=d$ 时的情况。由 $S(t)+R(t)+I(t)=a=\text{const}$ 可知，上述 3 个方程式相容，因而化简为

$$\begin{cases}\dfrac{\mathrm{d}S}{\mathrm{d}t}=-kIS,\ S(0)=S_0,\\ \dfrac{\mathrm{d}I}{\mathrm{d}t}=kIS-hI,\ I(0)=I_0.\end{cases}$$

用 Maple 可以得到方程的解为

$$I=\rho\ln S-S+\text{const},$$

带入初值 $I(S_0)=I_0$，得

$$I(S)=(S_0+I_0)-S+\rho\ln\frac{S}{S_0}.$$

1.2 模型的分析

1.2.1 参数的确定

1）治愈率 h 的确定

通过调查手足口传染病流行严重的安徽省，可以得到 2008 年 5 月 4 日至 5 月 11 日病情发展的详细情况，如表 1 所示[3]。

表 1 2008 年 5 月 4 日至 5 月 11 日病情发展情况统计表

Tab. 1 Statistics of disease progress from May, 4th to May, 11th in 2008

日期	总病例数/例	治愈人数/人	日病人数/人	日治愈数/人	日治愈率	死亡数/人
5-4	5 840	2 642	3 198			22
5-5	6 545	3 166	3 379	524	0.163 8	22
5-6	7 283	3 780	3 503	614	0.181 7	22
5-7	7 988	4 382	3 606	602	0.171 8	24
5-8	8 729	4 927	3 802	545	0.151 1	24
5-9	9 304	5 536	3 768	609	0.160 2	24
5-10	9 824	6 175	3 649	639	0.169 6	24
5-11	10 265	6 885	3 380	710	0.196 4	24

对日治愈率求平均值可以得到实际的日治愈率平均值为 $h=0.170\ 4$.

2）日接触率 k 的确定

很明显，从建立的模型无法得到 S，I，S_0，I_0 的解析解。为解决这个问题，用 Matlab 软件中的龙格-库塔方法求出它们的数值解。先通过实际统计数据算出每一天的 S，I，S_0，I_0，作出它们与时间的函数图像，如图 1 所示。

将这组图像与实际数据所得图像相比较、调试。发现当 $k\approx1.7$ 时，理论图形与实际图形有最佳的吻合。为简化运算，将 6 岁以下儿童看成一个不变的群体，这样就得到 $I(t)+S(t)+R(t)=1$. 根据确

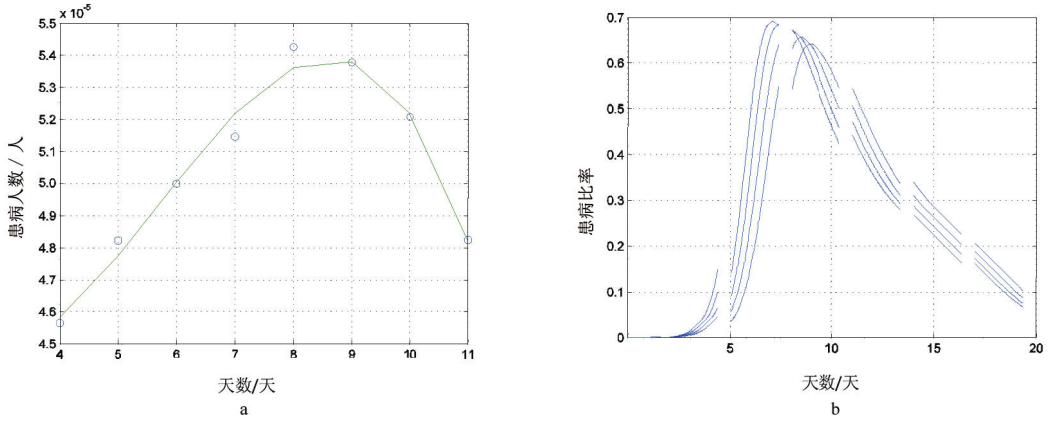

图 1　不同的参数取值下的患病人数图形

Fig. 1　Number of patients based on different parameters

a—固定某一参数之下的患病人数图形；b—不同的参数取值下的患病人数图形

a-Number of patients based on fixed parameters; b-Number of patients based on different parameters

定的参数值，就可以完全确定模型的方程为[2]

$$\begin{cases}\dfrac{\mathrm{d}S}{\mathrm{d}t}=-1.7IS,\\ \dfrac{\mathrm{d}I}{\mathrm{d}t}=1.7IS-0.1704I,\\ I(0)+S(0)=1.\end{cases}\tag{2}$$

1.2.2　结果分析和模拟检验

由于初始的治愈者数 $R(0)=0$，因而 $S_0+S_1=0$，代入参数值得 $\rho=\dfrac{h}{k}=0.1$，S-I 的微分方程就成为

$$I=1-S-0.1\ln S,$$

由 I-S 图形可以验证（如图 2 所示）：

1）当 $t\to\infty$时，$I\to\infty$. 即不管这个疾病的初始状态是什么，疾病最后都会消失。

2）由图 2 中可以看出，当 $I=0$ 时，方程变为 $1-S-0.1\ln S=0$，可以解出 $S_\infty<0.1$.

3）当 $S_0=\rho=0.1$ 时，$I(t)$ 达到最大值，

$$I_{\max}=S_0+I_0-\rho(1+\ln\frac{S_0}{\rho})=$$

$$1-0.1\times(1+\ln\frac{1}{0.1})=0.6697.$$

当 $S_0<\rho=0.1$ 时，$I(t)$ 单调增加；当 $S_0>\rho=0.1$ 时，$I(t)$单调递减到 0，这个时期病人比例 $I(t)$ 不会增加，传染病不会蔓延，逐渐好转。

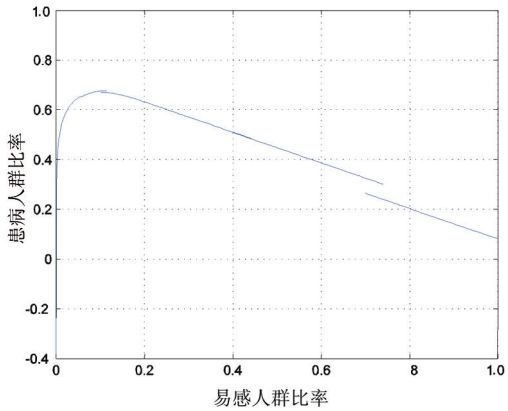

图 2　I-S 相轨线

Fig. 2　I-S phase trajectory

1.3　模型的评价

模型首先根据以往传染病模型，采用微分方程建立模型，分设变量。再通过统计数据与数据拟合求得各自的参

数值，利用数值计算得到结果并加以分析，得出传染病的传染规律，最后根据此分析提出对传染病的预测。

模型采用数值计算、图形观察与理论分析相结合的方法，先有感性认识，再用相轨线作理论分析，最后进行数值验证和估算，可以看作计算机技术与建模的配合。

模型采用微分方程本身就有一定的缺陷，其计算结果的准确性、可靠性将受到限制，再加之数值解的不确定性，对长时间的预测，模型有其局限性。因时间限制，模型没能更多地考虑交叉分类进行。

若能建立以随机偏微分方程组为基础的数学模型，将大大提高计算的准确性与可靠性，使得预测更加准确，但这样做将遇到模型求解、数据准确收集和数值求解的不精确性等诸多问题。

2 ARIMA 模型

2.1 模型的选择原因及概述

2.1.1 选择原因

时间序列分析（time seriesanalysis）是用随机过程理论和数理统计学的方法，研究随机数据序列所遵从的统计规律，以用于解决实际问题。由于在大多数的问题中，随机数据依照时间先后顺序排列，故称为时间序列。它包括一般统计分析，统计模型的建立与推断，以及关于随机序列的最优预测、控制和滤波等。在生物医学中，将时间序列分析方法应用于各种疾病发病趋势预测的例子越来越多，如肿瘤、肺结核、肝炎、SARS、动物鼠疫、肾综合征出血热（hemorrhagic fever with renal syndrome，HFRS）等。

根据时间序列数据有序列长短和发病频率不同的特点，人类、动物和植物的烈性传染性疾病具有发病周期短、发病率快速增加的特点。目前，时间序列方法在序列较长的事件上应用较多，对序列较短、发病率快速增加事件上的应用有待于进一步研究；另一方面，一维时间序列的预测方法应用相对较多。基于上述两个问题，研究采用常见的时间序列模型 ARIMA 模型进行拟合。

2.1.2 模型概述

称以下模型为 p 阶自回归与 q 阶滑动平均混合模型，以后简记为 ARMA(p，q) 模型[4]，即

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \cdots - \beta_q \varepsilon_{t-q},$$

其中，$\{\varepsilon_t\}$ 为白噪声序列，而且 $E(X_s \varepsilon_t)=0$，对一切 $s<t$ 成立，记

$$\alpha(u) = 1 - \alpha_1 u - \cdots - \alpha_p u^p,$$
$$\beta(u) = 1 - \beta_1 u - \cdots - \beta_q u^q,$$

$\alpha(u)$ 与 $\beta(u)$ 没有公共根。

ARMA(p，q) 模型适用于不存在任何趋势和周期性的平稳时间序列模型，而实际生活中常遇到的时间序列往往是不平稳的，这时通常对原始数据进行预处理，尽量使得经过预处理后的数据平稳。预处理包括确定性成份的分离、一阶或高阶差分、Box-Cox 变换以及零均值化等，使其达到平稳，进而建立求和自回归滑动平均模型（integrated autoregressive moving average model），简记为 ARIMA（p，d，q），其中，p，d，q 分别为自回归阶数、差分阶数和滑动平均阶数。

2.2 模型的拟合方法

1）序列平稳化[5]：ARIMA 的应用需要时间序列符合平稳性的要求。

2）模型参数估计和模型诊断：对提出的模型进行参数估计和诊断，如模型不恰当，则回到第二阶段，重新选定模型。

3）预测应用：4 月 27 日至 5 月 12 日的数据用于建立模型，4 月 29 日至 5 月 12 日的数据用于验证模型的预测效果。

2.3 模型的数据

表 2 数据来自腾讯手足口每日病情通报[6]，因安徽阜阳的数据比较多，所以取其数据进行拟合。

表 2 2008 年 4 月 27 日至 5 月 12 日病情发展情况统计表

Tab. 2 Statistics of disease progress from April, 27th to May, 12th in 2008

日期	4-27	4-28	4-29	4-30	5-1	5-2	5-3	5-4
感染人数/人	789	1 199	1 520	1 884	2 477	2 946	3 736	4 098
日期	5-5	5-6	5-7	5-8	5-9	5-10	5-11	5-12
感染人数/人	4 496	4 929	5 244	5 513	5 819	6 049	6 188	6 346

2.4 模型的阐述

一个平稳的随机过程应符合以下要求：均数不随时间变化；方差不随时间变化；自相关系数只与时间间隔有关，而与所处的时间无关。

2.4.1 模型的平稳化

研究的目的是期望建立的模型能够用来对手足口发病情况进行短期预测，结合本研究序列的特点，在各种短期预测方法中，尝试用 ARIMA 模型、指数平滑法进行建模和预测，并考察预测效果。建立 ARIMA 模型的基础是时间序列必须是平稳性的，因此首先考察序列的平稳性。

进行一次差分后，序列仍然有明显的下降趋势，所以序列不是平稳的[7]。进行二次差分后，得到如图 3 所示的结果。

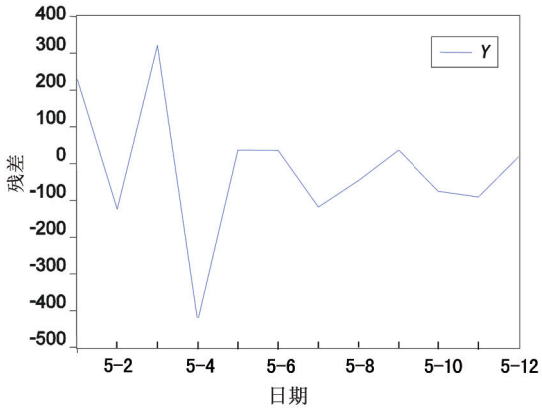

图 3 二次差分结果

Fig. 3 Result of second difference

可以看出进行二次差分后，序列没有明显的趋势，近似为平稳序列。

2.4.2 模型参数估计和模型诊断

对模型 ARIMA(2，2，0) 进行参数估计，其中，a_1，a_2 为参数[8]。

$$(1-a_1B-a_2B^2)(1-B)^2y=\varepsilon_t.$$

在这里，采用 AR (2) 模型，利用条件最小二乘估计计算结果，如表 3 所示，P 值均小于 0.05，参数可以直接代入模型。

表 3 模型参数的最小二乘估计

Tab. 3 Least squares estimation of parameter

方差	系数	标准差	t-统计量	概率
C	−28.304 700	50.373 930	−0.561 848	0.587 9
a_1	0.612 696	0.330 617	1.853 187	0.096 9
a_2	0.122 883	0.330 745	0.371 535	0.718 8

经过模型的识别、估计和诊断过程，得出了建立一个两次差分以后的 AR(2) 模型是合适的，即 ARIMA(2，2，0) 模型的表达式为

$$(1-0.612\ 7B-0.122\ 9B^2)(1-B)^2Y=\varepsilon_t.$$

2.4.3 预测应用

将数据带入拟合的式子中，可以得到预测效果图，如图 4 所示。预测结果如表 4 所示。

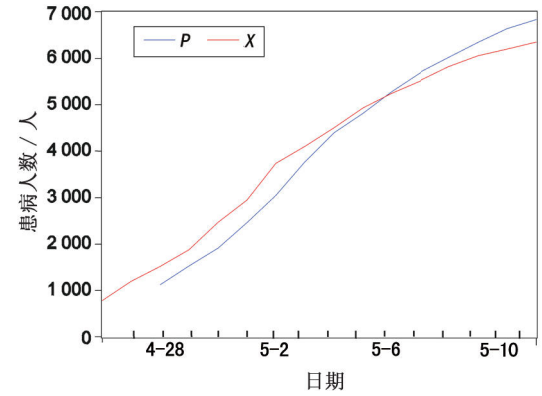

图 4 拟合数据与原数据对比

Tab. 4 Comparison between estimation and original data

由图 4 可以看出：预测值和原始病情指数吻合得较好。但病情增长速度有所差异，预测曲线中病情得到控制时增长比实际情况较快，可能是数据过少。

表 4 预测数据

Tab. 4 Forecasting data

日期	4-27	4-28	4-29	4-30	5-1	5-2	5-3	5-4
感染人数/人	789	1199	1 520	1 884	2 477	2 946	3 736	4 098
预测人数/人	—	—	1 131	1 534	1 918	2 464	3 050	3 769
日期	5-5	5-6	5-7	5-8	5-9	5-10	5-11	5-12
感染人数/人	4 496	4 929	5 244	5 513	5 819	6 049	6 188	6 346
预测人数/人	4 388	4 814	5 279	5 689	6 012	6 335	6 630	6 830

2.5 模型优缺点评价

ARIMA 模型建模过程严谨复杂，预测精度较高，但一般要求序列长度要达到 30. 由于手足口病爆发的周期较短、治愈率高、康复快，本建模过程只查找到 16 个时间点的数据，给后半部分的拟合带来了一些困难。从以上模型的拟合效果来说，在初期的拟合还是比较理想的，说明该模型具有广泛的应用价值，对其他病情发展的预测有较好的借鉴作用。

3 结论

从理论角度来说，*S-I-R* 模型的解受感染人数、治愈人数等变量的控制，对突发事件能做出较快的反应，对病情的走势有较好的把握，因而应该更准确一些。而 ARIMA 模型是根据病情随时间的发展，做出理想的估计，对突发事件反应较慢。从结果来说，*S-I-R* 模型的解较难求出，因而只能对走势进行大致的估计。而 ARIMA 模型能给出精确的数据，对医疗机构设置病人床位和药品供应量的大小都有很强的指导意义。

[参考文献] (References)

[1] 姜启源，谢金星，叶俊. 数学模型（3 版)[M]. 北京：高等教育出版社，2003.
JIANG Q Y，XIE J X，YE J. Mathematical model (3th Edition)[M]. Beijing：Higher Education Press，2003.

[2] 刘来福，曾文艺. 数学模型与数学建模（2 版)[M]. 北京：北京师范大学出版社，2002.
LIU L F，ZENG W Y. Mathematical model and mathematical modeling (2nd Edition)[M]. Beijing：Beijing Normal University Press，2002.

[3] 安徽省卫生厅. 疾病报告统计[OL]. [2008-05]. http：//www. ahwst. gov. cn/
Health Department of Anhui Province. Disease-reporting statistics[OL]. [2008-05]. http：//www. ahwst. gov. cn/ (in Chinese)

[4] 易丹辉. 数据分析与 EViews 应用[M]. 北京：中国统计出版社，2002. (in Chinese)
YI D H. Data analysis and application of EViews[M]. Beijing：China Statistics Press，2002.

[5] 邓丹，王润华，周燕荣. 时间序列分析及其在卫生事业中的应用[J]. 数理医药学杂志，2002，15 (5)：455～457.
DENG D，WANG R H，ZHOU Y R. The time series analysis and the application in the health fields[J]. Journal of Mathematical Medicine，2002，15(2)：455～457. (in Chinese)

[6] 腾讯新闻网站. 手足口疫情专题[OL]. [2008-05]. http：//news. qq. com/zt/2008/anhuifuyang/
Tencent News. Hand，foot and mouth epidemic feature[OL]. [2008-05]. http：// news. qq. com/zt/2008/anhuifuyang/ (in Chinese)

[7] 丁宁盍，唐家琦，王洁贞. ARIMA 模型在发病率预测中的应用[J]. 中国医院统计，2003，10 (1)：23～26.
DING N K，TANG J Q，WANG Z J. Application of ARIMA model on predictive incidence[J]. Chinses Journal of Hospical Statistics，2003，10(1)：23～26. (in Chinese)

[8] 温亮，徐德忠，林明和，等. 应用时间序列模型预测疟区疟疾发病率[J]. 第四军医大学学报，2004，25 (6)：507～510.
WEN L，XU D Z，LIN M H，et al. Prediction of malaria incidence in malaria epidemic area with time series models[J]. Journal of the Fourth Military Medical University，2004，25(6)：507～510. (in Chinese)