

DOI: 10.3969/cjz.j.issn.1002-2694.2013.01.021

## 温特斯模型在重庆市戊肝监测中的应用

杨艳红<sup>1</sup>, 曾庆<sup>1</sup>, 赵寒<sup>2</sup>, 易娟<sup>2</sup>, 李勤<sup>2</sup>, 肖达勇<sup>2</sup>, 夏宇<sup>2</sup>, 杨荣刚<sup>3</sup>, 方明金<sup>3</sup>

**摘要:**目的 建立戊型肝炎(戊肝)分月发病情况预测的时间序列模型,探讨时间序列预测模型在戊肝预测预警方面的应用。方法 利用 SAS 9.13 软件的时间序列预测系统(time series forecasting system)进行建模与分析。采用 2005 年 1 月至 2010 年 12 月的数据建模,并利用 2011 年 1 月至 2011 年 6 月的数据验证模型。结果 重庆市 2005—2010 年戊肝分月发病情况适合用温特斯模型(Winters method-additive)来拟合。结论 温特斯模型能够较好的应用于重庆市戊肝的预测,具有实际应用价值。

**关键词:** 戊型肝炎; 预测; 时间序列; 温特斯法

中图分类号: R373.2

文献标识码: C

文章编号: 1002-2694(2013)01-0096-05

### Application for Winters model on epidemic forecasting of hepatitis E in Chongqing Municipality

YANG Yan-hong<sup>1</sup>, ZENG Qing<sup>1</sup>, ZHAO Han<sup>2</sup>, YI Juan<sup>2</sup>, LI Qin<sup>2</sup>,  
XIAO Da-yong<sup>2</sup>, XIA Yu<sup>2</sup>, YANG Rong-gang<sup>3</sup>, FANG Ming-jin<sup>3</sup>

(1. Teaching and Research Section, Chongqing Medical University, Chongqing 400016, China;

2. Chongqing Center for Disease Control and Prevention, Chongqing 400042, China;

3. Chongqing Municipal Health Bureau, Chongqing 401147, China)

**ABSTRACT:** We aimed to construct time series model for forecasting the monthly cases of hepatitis E in Chongqing Municipality. The monthly cases of hepatitis E from 2005 to 2010 in Chongqing Municipality were analyzed by time series forecasting system in SAS 9.13. The model was established with the data from January in 2005 to December in 2010, and the data in 2011 was used for model validation. As a result, the Winters method-additive model was the automatically fitting model. In brief, the monthly cases of hepatitis E in Chongqing can be fitted by Winters model with an excellent application, and the model is practical for forecasting and warning hepatitis E epidemic situation in Chongqing.

**KEY WORDS:** hepatitis E; prediction; time series; Winters method

Supported by the Science and Technology Foundation of Chongqing Municipal Health Bureau (No. 2011-2-583)

Corresponding author: Zeng Qing, Email: zengqing1@gmail.com

全球已有多个国家对戊型肝炎流行的报道,在印度、墨西哥、东南亚、中亚、中东和非洲部分国家,戊肝是引发病毒性肝炎最重要的病原体。在美国、英国和法国等发达国家中,散发性病例也时有出现。我国 2004—2008 年卫生部传染病疫情报告显示<sup>[1]</sup>:近 5 年平均年报告戊肝 18 237 例,平均发病率为 1.38/10 万,与 2003 年相比,报告发病数平均每年增加约 88.89%,发病率平均每年增加约 93.07%,

报告发病数和发病率均呈现连续、快速增长态势。

本研究应用重庆市戊肝网络直报资料,通过建立时间序列预测模型来预测重庆市戊肝的流行情况,探讨使用预测模型进行预测预警的可行性,同时为戊肝的防控提供依据。

#### 1 资料与方法

1.1 资料来源 戊肝数据(表 1)来源于重庆市疾病预防控制中心《疾病监测信息报告管理系统》传染病个案病例,为重庆市 2005—2010 年按发病日期统计的戊肝数据,删除重卡,剔除已删除卡。数据真实可靠,具有权威性。

重庆市卫生局科技计划项目(项目编号 2011-2-583)

通讯作者: 曾庆, Email: zengqing1@gmail.com

作者单位: 1. 重庆医科大学卫生统计教研室, 重庆 400016;

2. 重庆市疾病预防控制中心, 重庆 400042;

3. 重庆市卫生局, 重庆 401147

表 1 重庆市 2005—2010 年戊肝按月发病例数  
Tab. 1 Monthly cases of HEV in Chongqing (2005—2010)

年份 Year	1	2	3	4	5	6	7	8	9	10	11	12
2005	5	5	11	15	7	11	6	5	3	5	3	4
2006	12	13	20	18	13	12	8	12	4	7	7	8
2007	12	15	38	26	27	23	16	15	11	7	10	12
2008	14	12	14	26	20	7	9	15	12	3	6	7
2009	12	24	27	22	6	7	11	9	6	8	12	14
2010	22	27	61	38	45	33	38	19	13	23	29	43

1.2 方法 利用 2005 年 1 月至 2010 年 12 月的戊肝月发病数建立时间序列预测模型,同时用 2011 年 1 月至 2011 年 6 月的数据进行模型验证。用 SAS 9.13 软件的 SAS/ETS 时间序列预测系统(time series forecasting system)进行自动分析,检验水准  $\alpha = 0.05$ 。

1.3 温特斯(Winters)模型 温特斯模型类似指数平滑模型,综合考虑了长期趋势、季节变动和不规则变动。按照长期趋势和季节比率  $s_t$  的关系,模型分为乘法模型和加法模型。其基本形式如下:

$$x_t = (a + bt)s_t + \epsilon_t \text{ 乘法模型} \quad (1)$$

$$x_t = a + bt + s_t + \epsilon_t \text{ 加法模型} \quad (2)$$

上两式中,  $x_t$  表示时间序列  $t$  时刻的实际值,  $a_t$  表示时间序列  $t$  时刻的常数平滑水平,  $b_t$  表示时间序列  $t$  时刻的线性平滑趋势,  $s_t$  表示第  $t$  个季节比率的指数平滑值,  $\epsilon_t$  为随机波动。

该模型中的  $(a + bt)$  部分表示长期趋势,可以拟合二项式(非线性)、线性式和常数式在内的 3 种长期趋势模型。因为二项式模型不稳定,在实际应用中常用线性式模型。

温特斯模型(1)、(2)的拟合以下述 3 个平滑方程为基础,3 个平滑方程如下:

$$a_t = \omega_1 x_t / s_{t-1} + (1 - \omega_1)(a_{t-1} + b_{t-1}) \quad (3)$$

$$b_t = \omega_2 (a_t - a_{t-1}) + (1 - \omega_2) b_{t-1} \quad (4)$$

$$s_t = \omega_3 (x_t / a_t) + (1 - \omega_3) s_{t-1} \quad (5)$$

式中  $\omega_1, \omega_2, \omega_3$  表示平滑权数,表示了前后两个时间点值的关系。由这 3 个平滑方程可知,模型(1)、模型(2)的参数  $a, b, s_t$  是随着时间序列不断更新的。

## 2 结果

2.1 时间序列特征分析 2005 年 1 月至 2010 年 12 月重庆市戊肝分月发病数呈明显波动趋势,每年 3—4 月发病数明显增多,达高峰,见图 1,该时间序

列是非平稳和非随机序列,序列存在长期递增的趋势模式和以年为周期的季节模式。由戊肝原始时间序列自相关系数图和偏自相关系数图(图 2)也可得到相同结果,自相关图呈现出明显的正弦波动规律,这是具有周期变化规律的非平稳序列的典型特征。

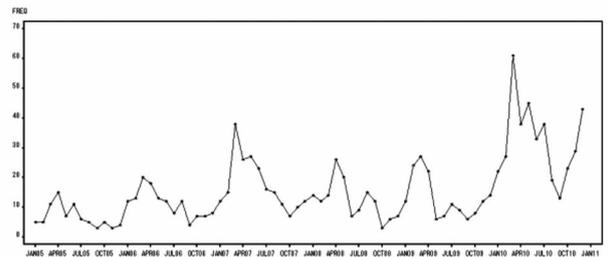


图 1 重庆市 2005—2010 年戊肝按月发病数

Fig. 1 Monthly cases of HEV in Chongqing (2005—2010)

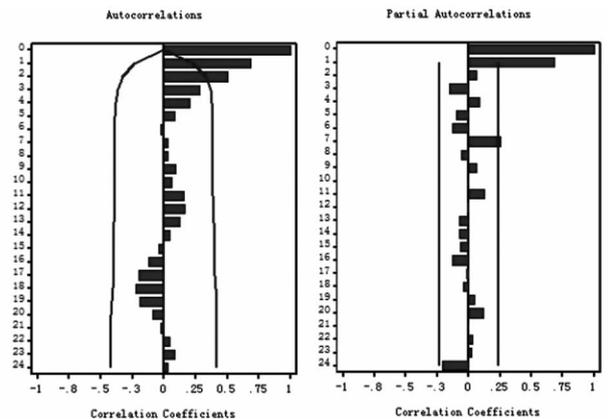


图 2 原始序列自相关系数图和偏自相关系数

Fig. 2 Correlation coefficients of autocorrelation and partial autocorrelation of the raw time series data

2.2 建立模型 利用 SAS/ETS 统计分析软件时间序列预测系统(time series forecasting system)进行自动筛选建模。在 SAS/ETS 统计分析软件中可以使用 Mean Square Error, Root Mean Square Error, Mean Absolute Error, Mean Absolute Percent Error 和 R-Square 等作为模型的选择准则。

本文选用均方根误差最小的模型选择准则即

Root Mean Square Error 选择准则,根据重庆市戊肝的时间序列特点,SAS/ETS 系统自动筛选出对数转换温特斯加法(Log Winters Method-Additive)预测模型,均方根误差为 5.74503,决定系数  $R^2$  为 0.74147。自动拟合的白噪声和单位根检验结果如图 3 所示,由图可知,白噪声检验是显著的,但是单位根检验却不显著。结果提示,需要对模型进行调整。本文采用了把平滑权重(smoothing weights)的边界(bounds)设为无限制(unrestricted)。调整

后重新拟合,此时模型的均方根误差为 5.50873,决定系数  $R^2=0.762$ 。经调整后的检验结果如图 4 所示,可知白噪声检验和单位根检验均显著了,说明拟合模型满足了模型的平稳性假定。模型拟合后自相关和偏相关系数图(图 5)也显示序列为平稳随机序列,这些都说明该预测模型建模成功。预测模型参数见表 2。模型的具体形式为:  $\log x_t = 3.70082 + 0.01984t + s_t + \epsilon_t$ 。

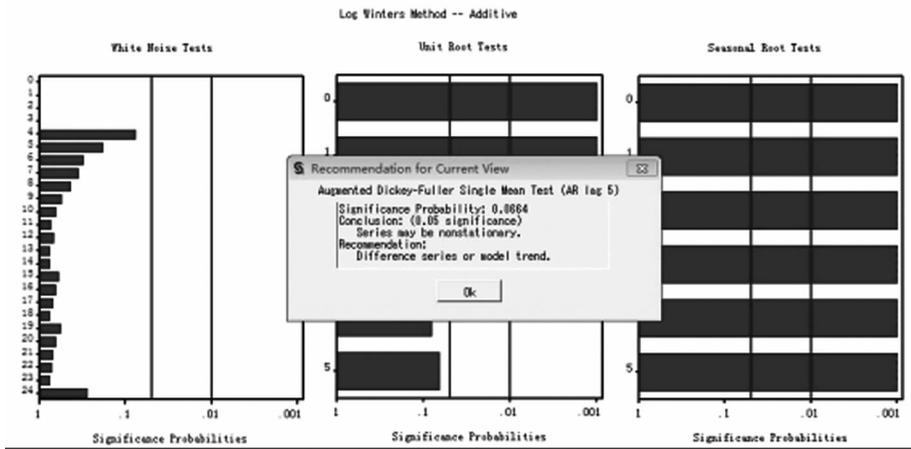


图 3 自动拟合结果白噪声检验图和单位根检验  
Fig. 3 White noise tests and unit root tests by automatically fitting

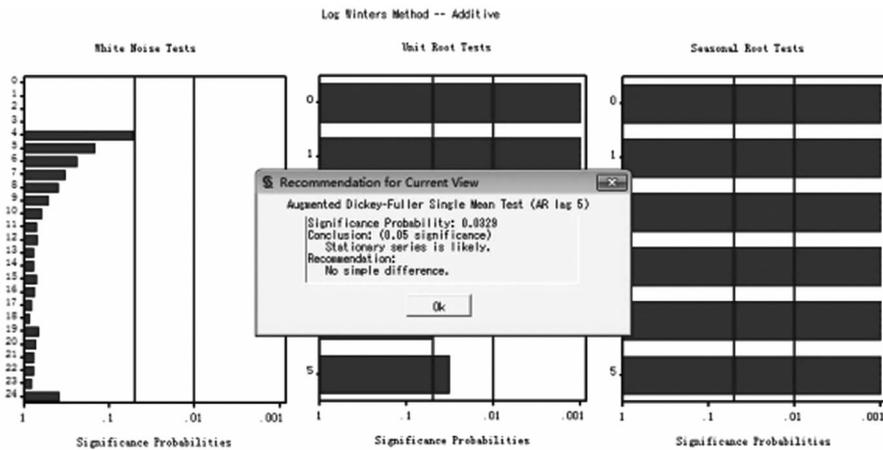


图 4 经调整后的白噪声检验图和单位根检验  
Fig. 4 White noise tests and unit root tests after modification

2.3 模型预测 利用确定的模型对 2011 年上半年戊肝发病数进行预测,结果如表 3 所示,实际值均在预测值 95% CI 内。预测值与实际值对比发现,总体平均绝对值相对误差为 13.25%(平均相对误差为 9.66%),提示用此模型对重庆市戊肝发病数的预测分析基本合理。

利用该模型对重庆市戊肝报告发病数预测值和实际值作图,结果显示预测数据与实际数据吻合程

度较好,预测值在 2011 年 3~4 月有一个高峰,这与实际情况也是相吻合的,说明该模型能较好拟合重庆市戊肝报告发病数的变化趋势,见图 6。

### 3 讨论

近年来,戊肝作为一种人兽共患病越来越引起国家重视。从重庆市戊肝分月发病数时序图中可以看出,戊肝发病逐年增加,但无暴发或流行,从 2006

表 2 重庆市 2005—2010 年戊型肝炎温特斯加法预测模型参数估计

Tab. 2 Parameter estimates of Winters model of HEV in Chongqing (2005—2010)

Model parameter	Estimate	Std error	t	P
Level smoothing weight( $w_1$ )	0.39124	0.0711	5.4998	<0.0001
Trend smoothing weight( $w_2$ )	0.0003519	0.0125	0.0281	0.9777
Seasonal smoothing weight( $w_3$ )	-0.14262	0.0701	-2.0342	0.0458
Smoothed level(a)	3.70082	—	—	—
Smoothed trend(b)	0.01984	—	—	—
Smoothed seasonal factor 1	-0.06001	—	—	—
Smoothed seasonal factor 2	0.12731	—	—	—
Smoothed seasonal factor 3	0.65601	—	—	—
Smoothed seasonal factor 4	0.63640	—	—	—
Smoothed seasonal factor 5	0.21601	—	—	—
Smoothed seasonal factor 6	0.07973	—	—	—
Smoothed seasonal factor 7	-0.05735	—	—	—
Smoothed seasonal factor 8	-0.08774	—	—	—
Smoothed seasonal factor 9	-0.58599	—	—	—
Smoothed seasonal factor 10	-0.55781	—	—	—
Smoothed seasonal factor 11	-0.36690	—	—	—
Smoothed seasonal factor 12	-0.15341	—	—	—

Note: The residual variance of the model=0.12641; AIC=251.71224;  $R^2=0.762$

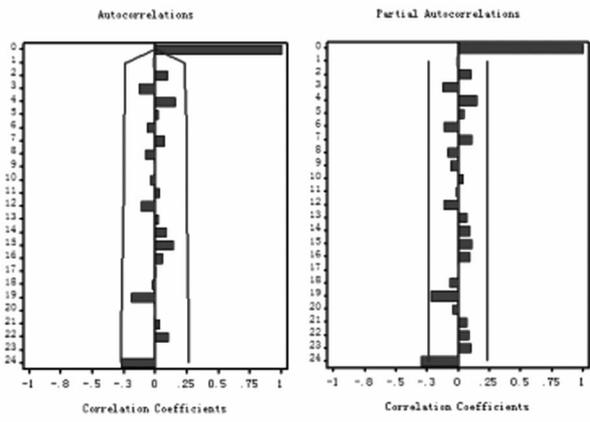


图 5 模型拟合后自相关系数图和偏自相关系数  
Fig. 5 Correlation coefficients of autocorrelation and partial autocorrelation model fitting

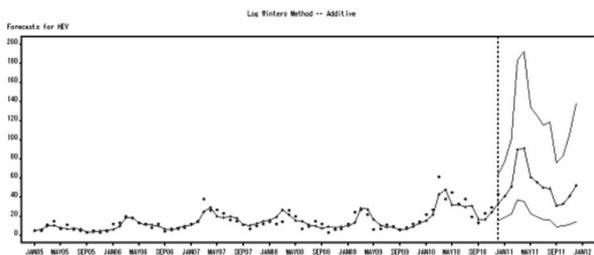


图 6 温特斯法预测模型拟合  
Fig. 6 The scatter plot of HEV cases and Winters model forecast values

年开始,增长环比依次为 67.50%、58.21%、-31.60%、8.97%、147.47%,提示 2010 年增长速

表 3 拟合模型对 2011 年上半年重庆市戊肝发病数预测效果验证

Tab. 3 Prediction of model of HEV from January to June in Chongqing, 2011

Month	Actual value	Forecast (95% confidence interval)	Relative error (%)
2011-01	47	41.42 (19.37—78.06)	11.87
2011-02	74	51.45 (22.63—101.10)	30.47
2011-03	109	89.92 (37.33—183.62)	17.50
2011-04	82	90.82 (35.68—192.20)	-10.76
2011-05	63	61.44 (22.89—134.44)	2.48
2011-06	59	55.22 (19.56—124.70)	6.41

度较快,同时考虑全国快速增长的态势<sup>[1]</sup>,提示应加强对戊肝流行态势的关注度,并做好预测预警的工作。

时间序列分析与预测是一种历史悠久而又富有生命力的预测方法,最早的时间序列可以追溯到 7000 年前的古埃及。时间序列分析在各行各业应

用广泛,可用于产品寿命的预测<sup>[2]</sup>、疾病发病率的预测<sup>[3-4]</sup>、人群死亡率的预测<sup>[5-6]</sup>、病虫灾害的预报<sup>[7]</sup>和地震预报<sup>[8]</sup>等。

目前 ARIMA 时间序列模型应用最为广泛,对难以辨别时间序列资料特征且需要综合考虑趋势和周期变化以及随机干扰的情况下,可借助 ARIMA 模型进行量化,处理带有趋势、季节平稳和含有随机项数据的资料,但有时需综合考虑线性、季节因素和不规则变化,这时可选用温特斯法。温特斯法是与指数平滑法相结合的季节时间序列预测法。目前该法应用在疾病监测<sup>[9]</sup>、航空运输<sup>[10]</sup>、警情预测<sup>[11]</sup>和医院管理<sup>[12]</sup>等领域。

温特斯模型由 3 部分构成<sup>[13]</sup>: 稳定成分,长期趋势成分和季节因素成分。该模型可综合解释趋势变动、季节变动、循环变动和随机变动。根据趋势成分与季节因素的关系,可将该模型分为乘法模型与加法模型两种形式,当长期趋势与季节效应不是简单的相加关系时就要考虑乘法模型。本文通过拟合温特斯加法模型说明,重庆市戊肝发病情况的长期趋势与季节效应间没有乘积关系,只是简单的相加关系,这说明月份对戊肝发病情况虽有影响,但是这个影响力还是在有限的范围内,而且从季节指数的大小、正负来看,2、3、4、5、6 月份为发病高发月份,而其他月份为低发月份,同时其值波动范围依然是有限的;模型的线性趋势  $b_1$  值是 0.01984,其值为正,说明总的来看,重庆市戊肝流行的长期趋势是增加的,短期内不会下降,但是该值又不是很大,说明戊肝流行的上升是一个缓慢的过程。

预测精度反应了预测模型的好坏,常用的预测精度指标有:均方误差,平均误差,误差标准差,均方根误差,平均百分误差,误差平方和平均绝对百分误差等。通常根据实际情况选择其中一个或几个指标综合考虑。本文采用均方根误差最小<sup>[9,13]</sup>作为选择最优模型的标准,效果较好。

自动拟合的模型与调整后模型的绝对值平均相对误差分别为 10.42%,13.25%,自动拟合的拟合优度相对较高,这是由于 SAS/ETS 系统考虑的是模型的拟合优度而不是拟合的有效性,所以在实际应用中要特别注意模型的残差检验和稳定性检验,当拟合模型不符合模型假定条件时就需要调整。本文正是如此。

目前大多数统计软件(如 SPSS, SAS, R 等)都可直接用于时间序列分析。本研究采用的是 SAS/ETS 的时间序列预测系统,它内置了 42 种常见的时间序列分析模型,包括对数转换和未经对数转换

的 ARIMA 模型、指数平滑模型、线性趋势拟合模型和温特斯模型等。同时该系统可利用时间序列数据集根据指定的建模指标自动选择拟合效果相对最好的时间序列模型,这对于初学者来说是一个很容易掌握的方法,同时也可以用它来探索时间序列的特征,为我们选取模型提供参考。

## 参考文献:

- [1] Chinese Center for Disease Control and Prevention. Legal infectious diseases report, 2005—2008 [DB/OL]. <http://www.chinacdc.net.cn/n272442/n272530/n272757/index.html> (in Chinese)  
中国疾病预防控制中心. 2005—2008 年度中国法定传染病疫情报告 [DB/OL]. <http://www.chinacdc.net.cn/n272442/n272530/n272757/index.html>
- [2] Di X, Zhang BH. Lifetime prediction model of wireless sensor networks[J]. *Comput Eng Des*, 2010, 31(22): 4769-4772. (in Chinese)  
底欣,张百海. 无线传感器网络寿命预测模型[J]. *计算机工程与设计*, 2010, 31(22): 4769-4772.
- [3] Yang PR, Tian H, Yan CY. Application of autoregressive integrated moving average model in forecasting of the monthly incidence rate in Baoji[J]. *J Pub Health Prev Med*, 2012, 23(1): 16-19. (in Chinese)  
杨培荣,田辉,严钊元. 自回归移动平均模型在宝鸡市细菌性痢疾月发病率预测中的应用[J]. *公共卫生与预防医学*, 2012, 23(1): 16-19.
- [4] Hu JL, Liang Q, Wu Y, et al. Applications of SARIMA Model on Forecasting in Incidence of the *Baoji* *Dysentery*[J]. *Chin J Health Stat*, 2012, 29(1): 34-36, 39. (in Chinese)  
胡建利,梁祁,吴莹,等. 季节时间序列模型在菌痢发病预测中的应用[J]. *中国卫生统计*, 2012, 29(1): 34-36, 39.
- [5] An HQ, Ma GF, Li XY, et al. Prediction and analysis of the neonatal mortality in China[J]. *Matern Child Health Care Chin*, 2012, 27(8): 1156-1158. (in Chinese)  
安洪庆,马桂峰,李向云,等. 我国新生儿死亡率预测分析[J]. *中国妇幼保健*, 2012, 27(8): 1156-1158.
- [6] Zhou ZY, Li JQ, Xu XY. Time series analysis of gastric cancer mortality trend in Changshu City during 1977-2008[J]. *Occup Health*, 2010, 26(13): 1505-1507. (in Chinese)  
周正元,李建清,徐晓燕,等. 1977-2008 年常熟市胃癌死亡率趋势的时间序列分析[J]. *职业与健康*, 2010, 26(13): 1505-1507.
- [7] Wu WH, Fu HP, Wu XD, et al. Forecasting the population dynamics of *Cricetulus barabensis* and *Meriones unguiculatus* by time series method[J]. *J Inner Mongolia Agric Univ (Nat Sci Edn)*, 2007, 4: 6-11. (in Chinese)  
武文华,付和平,武晓东,等. 应用时间序列分析预测黑线仓鼠和长爪沙鼠种群数量[J]. *内蒙古农业大学学报(自然科学版)*, 2007, 4: 6-11.
- [8] Zheng H, Li W, Qiu JF, et al. Earthquake predicting based on a matching algorithm of the time series similarity[J]. *J Beijing Elec Sci Tech Inst*, 2010, 2: 10-16. (in Chinese)