

桂林地区气温与降水量的时间序列预测模型*

The Prediction Model of Time Series for the Air Temperature and Precipitation in Guilin Region

黎玉芳¹,李志鸿²LI Yu-fang¹,LI Zhi-hong²

(1. 广西师范大学数学科学学院,广西桂林 541004;2. 广西陆川县中学,广西陆川 537700)

(1. College of Mathematics,Guangxi Normal University,Guilin,Guangxi,541004,China;

2. Guangxi Luchuan Middle School,Luchuan,Guangxi,537700,China)

摘要:根据桂林地区 2001~2010 年各月平均气温和 1987~2010 年每年 6 月份降水量的实测数据,利用时间序列的季节指数法和线性外推法分别建立气温预测模型和降水量预测模型,再利用模型对桂林市 2011 年各月的气温和 2011 年 9、11、12 月的降水进行预测,得到比较理想的预测效果。

关键词:气温 降水量 时间序列 季节指数法 线性外推法

中图分类号:O212 文献标识码:A 文章编号:1005-9164(2013)02-0107-04

Abstract:Based on the monthly mean air temperature and monthly total precipitation in Guilin region, the air temperature prediction model and precipitation forecast model are established by using seasonal index method and linear extrapolation method of stationary time series. The results verify that the models have satisfactory prediction performance.

Key words:air temperature,precipitation,time series,seasonal index,linear extrapolation

气温和降水的预测对于人们的日常生活、防洪防旱、农业发展以及水资源的合理分配与利用有着重大意义,是当前的一个重点领域研究。广西桂林是一个旅游城市,也是一个雨量充沛的城市,对该城市的气温和降水量的预测可以帮助人们更好地安排自己的出游计划,还可以为防洪防灾提供可靠的理论依据。由于气温和降雨都是自然现象,随机因素的影响较大,数据观测也比较困难,目前国内关于这方面研究的报道很少。石昱馨等^[1]利用 ARIMA 乘积季节模型对内蒙古海拉尔地区 2007 年 9~12 月的月平均气温和降水量进行预测;迟道才等^[2]应用自回归模型 ARMA(p,q) 对辽宁朝阳地区干旱天气进行预测。本文根据桂林地区月平均气温和月降水量的实测数据,分别利用季节指数法和平稳时间序列的线性外推法建立该地区气温和降水量的预测模型,得到了较理想的预测结果。

收稿日期:2012-07-30

修回日期:2012-12-24

作者简介:黎玉芳(1977-),女,讲师,主要从事概率统计研究。

* 国家自然科学基金项目(11061007,11201088),国家社会科学基金项目(10CTJ004),广西教育厅科研项目(201106LX051)资助。

广西科学 2013 年 5 月 第 20 卷第 2 期

1 时间序列分析方法

1.1 季节指数法

季节指数法是以时间序列含有季节性周期变动的特征,计算描述该变动的季节变动指数的方法^[3]。假设有一个时间序列 $\{y_t\}$,其中 t 表示该序列的下标。该序列的数据结构以 k 期为一周期,共有 n 个周期。如果设各个时期的季节指数为 $\alpha_i, i = 1, 2, \dots, k$,那么季节模型为

$$\hat{y}_{ji} = \bar{y} \cdot \alpha_i, i = 1, 2, \dots, k \quad (1)$$

其中 \hat{y}_{ji} 表示第 j 年第 i 个时期的预测值, \bar{y} 表示总的平均数。计算季节指数分为 3 步:

1) 计算一个周期内的各期平均数。设 $r_i, i = 1, 2, \dots, k$ 为第 i 期的平均数

$$r_i = \frac{\sum_{j=1}^n y_{ji}}{n}, i = 1, 2, \dots, k; \quad (2)$$

2) 计算总平均数

$$\bar{y} = \frac{1}{nk} \sum_{j=1}^n \sum_{i=1}^k y_{ji}; \quad (3)$$

3) 用各期平均数除以总平均数就可以得到各时

期的季节指数 $\alpha_i, i = 1, 2, \dots, k$,

$$\alpha_i = \frac{r_i}{y}, i = 1, 2, \dots, k. \quad (4)$$

1.2 平稳时间序列的线性外推法^[4]

设 $\{x_t\}$ 为一平稳时间序列, 并设一组实际观测值为 y_1, y_2, \dots, y_N , 则在具体应用时, 其均值 $E(x_t)$ 和相关函数 $B(k)$ 有下列的近似公式:

$$E(x_t) = c \approx \frac{1}{N} \sum_{i=1}^N y_i = \bar{y},$$

$$B(k) = E[x_{t+k} - c][x_t - c] \approx \frac{1}{N-k} \sum_{i=1}^{N-k} (y_{t+k} -$$

$$\bar{y})(y_i - \bar{y}) = \frac{1}{N-k} \sum_{i=1}^{N-k} y'_{t+k} y'_i,$$

其中 $y'_i = y_i - \bar{y}$ 称为 y_i 的距平。

采用平稳时间序列的线性外推法进行建模可以由下面几步完成:

1) 由 $\{x_t\}$ 的一组观测值 y_1, y_2, \dots, y_N , 先按状态分类, 然后计算样本平均数和各个数据的距平值,

$$\text{即 } \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i, y'_i = y_i - \bar{y}. \quad (5)$$

2) 由相关函数的性质及计算误差精度确定预报方程阶数 m 。一般取小于资料阶数的四分之一, 也可以取几种不同阶数分别进行预报值比较, 选择与历史资料吻合最好的作为应取阶数。

3) 在最小二乘法原则的基础上, 确定线性最优预报方程

$$\hat{y}'_{N+\tau} = \alpha_1 y'_N + \alpha_2 y'_{N-1} + \dots + \alpha_m y'_{N-m+1}, \quad (6)$$

其中 τ 取正整数(外推步长)。

$$4) \text{ 由 } B(k) = \frac{1}{N-k} \sum_{i=1}^{N-k} y'_{t+k} y'_i, \text{ 计算出 } B(0),$$

$B(1), \dots, B(m)$ 。

5) 解线性方程组

表 1 2001~2010 年的各月平均气温

Table 1 Monthly mean air temperatures between 2001 and 2010

月份 Month	平均气温 Mean air temperatures(0.1°)										r_i	α_i
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010		
1	87	112	91	80	62	82	77	56	85	97	82.9	0.424
2	105	123	117	127	63	89	149	75	143	119	111	0.567
3	151	151	134	133	125	140	148	161	143	144	143	0.731
4	181	196	197	210	205	204	185	195	197	170	194	0.992
5	236	231	232	233	253	238	252	236	234	232	237.7	1.215
6	261	271	261	268	266	266	267	264	270	244	263.8	1.348
7	283	276	301	267	292	283	289	283	282	295	285.1	1.457
8	276	265	290	280	279	284	294	281	294	293	283.6	1.450
9	265	245	259	264	266	254	249	266	277	264	260.9	1.334
10	222	202	206	216	216	236	221	230	236	208	219.3	1.121
11	161	162	159	170	178	169	170	158	134	166	162.7	0.832
12	84	93	104	110	94	109	114	111	103	115	103.7	0.530

$$\begin{cases} B(0)\alpha_1 + B(-1)\alpha_2 + \dots + B(1-m)\alpha_m = B(\tau), \\ B(1)\alpha_1 + B(0)\alpha_2 + \dots + B(2-m)\alpha_m = B(\tau+1), \\ \dots\dots \\ B(m-1)\alpha_1 + B(m-2)\alpha_2 + \dots + B(0)\alpha_m = \\ B(\tau+m-1), \end{cases} \quad (7)$$

确定系数 $\alpha_1, \alpha_2, \dots, \alpha_m$ 。

6) 确定最优步长 τ 后, 得出最优估计 $\hat{y}'_{N+\tau}$, 于是最终的预测值为

$$\hat{y}_{N+\tau} = \hat{y}'_{N+\tau} + \bar{y}. \quad (8)$$

2 基于时间序列的预测模型建立

2.1 气温预测模型的建立

桂林地区 2001~2010 年各月的平均气温数据(表 1)来源于中国气象科学数据共享服务网(<http://www.cma.gov.cn/2011qxw/2011qsjgx/index.htm>)。现将平均气温数据绘制成时序图 1。从图 1 可以看出, 桂林地区的气温具有明显的季节性, 每月的平均气温随着时间的推移都有着明显的规律变动。

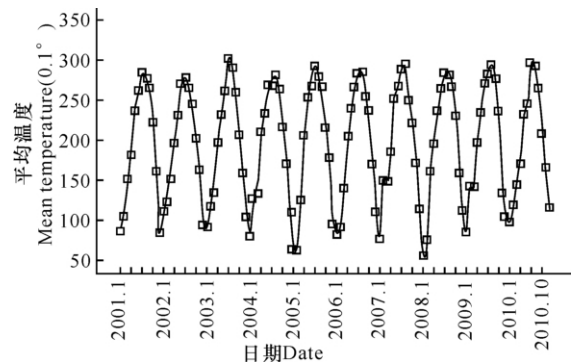


图 1 桂林市 2001~2010 年各月平均气温时序

Fig.1 Plot of time series of monthly mean air temperatures between 2001 and 2010 in Guilin City

以 12 个月作为一个周期,2001~2010 年即是 10 个周期,再利用(2)~(4)式,得到各期平均数 $r_i, i = 1, 2, \dots, 12$ 和各期季节指数 $\alpha_i (i = 1, 2, \dots, 12)$ 如表 1 所示,计算总平均数 $\bar{y} = 195.642$ 。从各期季节指数可以看出 7 月份和 8 月份的季节指数最大,所以说 7 月和 8 月是桂林一年中最热的月份;1 月份的季节指数最小,说明 1 月份是桂林一年中最冷的月份。

2.2 降水预测模型的建立

桂林地区 1987~2010 年 6 月份的降水量实测数据(表 2)来源于中国气象科学数据共享服务网(<http://www.cma.gov.cn/2011qxw/2011qsjgx/index.htm>)。首先利用 SPSS 软件对降水量数据进行平稳性分析,得到自相关函数如图 2。

表 2 桂林 1987~2010 年各年 6 月降水量

Table 2 Total precipitation in June between 1987 and 2010 in Guilin

日期 Date	降水量(0.1mm) Precipitation	日期 Date	降水量(0.1mm) Precipitation
1987.6	3027	1999.6	2370
1988.6	3700	2000.6	4324
1989.6	2897	2001.6	2287
1990.6	2421	2002.6	7580
1991.6	3288	2003.6	3037
1992.6	5475	2004.6	3017
1993.6	5163	2005.6	6448
1994.6	6018	2006.6	4174
1995.6	3024	2007.6	4536
1996.6	2887	2008.6	5417
1997.6	2128	2009.6	3352
1998.6	7667	2010.6	5307

从图 2 可以看出,该序列自相关系数 (ACF) 一

表 3 y'_i 的数据

Table 3 y'_i 's data

序号	y'_i	$y'_i \cdot y'_i$	y'_{i+1}	$y'_{i+1} \cdot y'_i$	y'_{i+2}	$y'_{i+2} \cdot y'_i$	y'_{i+3}	$y'_{i+3} \cdot y'_i$	y'_{i+4}	$y'_{i+4} \cdot y'_i$	y'_{i+5}	$y'_{i+5} \cdot y'_i$
1	-1120.667	1255893.778	-447.667	501685.111	-1250.667	1401580.444	-1726.667	1935017.778	-859.667	963399.778	1327.333	-1487498.222
2	-447.667	200405.444	-1250.667	559881.778	-1726.667	772971.111	-859.667	384844.111	1327.333	-594202.889	1015.333	-454530.889
3	-1250.667	1564167.111	-1726.667	2159484.444	-859.667	1075156.444	1327.333	-1660051.556	1015.333	-1269843.556	1870.333	-2339163.556
4	-1726.667	2981377.778	-859.667	1484357.778	1327.333	-2291862.222	1015.333	-1753142.222	1870.333	-3229442.222	-1123.667	1940197.778
5	-859.667	739026.778	1327.333	-1141064.222	1015.333	-872848.222	1870.333	-1607863.222	-1123.667	965978.778	-1260.667	1083753.111
6	1327.333	1761813.778	1015.333	1347685.778	1870.333	2482555.778	-1123.667	-1491480.222	-1260.667	-1673324.889	-2019.667	-2680770.889
7	1015.333	1030901.778	1870.333	1899011.778	-1123.667	-1140896.222	-1260.667	-1279996.889	-2019.667	-2050634.889	3519.333	3573296.444
8	1870.333	3498146.778	-1123.667	-2101631.222	-1260.667	-2357866.889	-2019.667	-3777449.889	3519.333	6582326.444	-1777.667	-3324829.222
9	-1123.667	1262626.778	-1260.667	1416569.111	-2019.667	2269432.111	3519.333	-3954557.556	-1777.667	1997504.778	176.333	-198139.889
10	-1260.667	1589280.444	-2019.667	2546126.444	3519.333	-4436706.222	-1777.667	2241045.111	176.333	-222297.556	-1860.667	2345680.444
11	-2019.667	4079053.444	3519.333	-7107880.222	-1777.667	3590294.111	176.333	-356134.556	-1860.667	3757926.444	3432.333	-69321.222
12	3519.333	12385707.111	-1777.667	-6256201.556	176.333	620575.778	-1860.667	-6548306.222	3432.333	12079525.111	-1110.667	-3908806.222
13	-1777.667	3160098.778	176.333	-313461.889	-1860.667	3307645.111	3432.333	-610644.556	-1110.667	1974395.111	-1130.667	2009948.444
14	176.333	31093.444	-1860.667	-328097.556	3432.333	605234.778	-1110.667	-195847.556	-1130.667	-199374.222	2300.333	405625.444
15	-1860.667	3462080.444	3432.333	-6386428.222	-1110.667	2066580.444	-1130.667	2103793.778	2300.333	-4280163.556	26.333	-48997.556
16	3432.333	11780912.111	-1110.667	-3812178.222	-1130.667	-3880824.889	2300.333	7895510.778	26.333	90384.778	388.333	1332889.444
17	-1110.667	1233580.444	-1130.667	1255793.778	2300.333	-2554903.556	26.333	-29247.556	388.333	-431308.889	1269.333	-1409806.222
18	-1130.667	1278407.111	2300.333	-2600910.222	26.333	-29774.222	388.333	-439075.556	1269.333	-1435192.889	-795.667	899633.778
19	2300.333	5291533.444	26.333	60575.444	388.333	893296.111	1269.333	2919889.778	-795.667	-1830298.556	1159.333	2666853.111
20	26.333	693.444	388.333	10226.111	1269.333	33425.778	-795.667	-20952.556	1159.333	30529.111		
21	388.333	150802.778	1269.333	492924.444	-795.667	-308983.889	1159.333	450207.778				
22	1269.333	1611207.111	-795.667	-1009966.222	1159.333	1471580.444						
23	-795.667	633085.444	1159.333	-922442.889								
24	1159.333	1344053.778										
Σ		62325949.333		-18245940.444		2715662.111		-11285341.000		11225896.222		-6526833.889

直都比较小,始终控制在 2 倍标准差范围内并且逐渐趋于 0,而自相关系数在某一固定水平线附近摆动且按周期逐渐衰减,说明该时间序列具有平稳性。

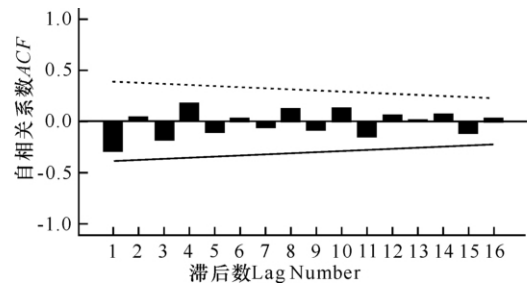


图 2 降雨量自相关图

Fig. 2 Autocorrelogram of precipitation

■: 系数;: 上限; ——: 下限.

■: Coefficient;: Upper coefficient limit; ——: Lower coefficient limit.

由(5)式计算得 1987~2010 年 6 月的平均降雨

量: $\bar{y} = \frac{1}{24} \sum_{i=1}^{24} y_i = 4147.667$,再对原序列做平稳化

处理,这里以距平值 $y'_i = y_i - \bar{y}, i = 1, 2, 3, \dots, 24$ 进行平稳化处理。取 $\tau = 1$,因为 $N = 24$,所以阶数 $m = 5$ 。由式(6)得

$$\hat{y}'_{2011} = \alpha_1 y'_{2010} + \alpha_2 y'_{2009} + \alpha_3 y'_{2008} + \alpha_4 y'_{2007} + \alpha_5 y'_{2006}。$$

由式(5)及递推关系得到距平值 y'_i 的一系列数据,其结果如表 4 所示。

由表 3 的距平值 y'_i 计算 $B(0), B(1), B(2), B(3), B(4), B(5)$ 如下:

$$B(0) = \frac{1}{24} \sum_{i=1}^{24} (y'_i)^2 = 2596914.556, B(1) = \frac{1}{23} \sum_{i=1}^{23} y'_{i+1} y'_i = -793301.758, B(2) = \frac{1}{22} \sum_{i=1}^{22} y'_{i+2} y'_i = 123439.187, B(3) = \frac{1}{21} \sum_{i=1}^{21} y'_{i+3} y'_i = -537397.190, B(4) = \frac{1}{20} \sum_{i=1}^{20} y'_{i+4} y'_i = 561294.811, B(5) = \frac{1}{19} \sum_{i=1}^{19} y'_{i+5} y'_i = -343517.573.$$

由(7)式得到线性方程组

$$\begin{cases} B(0)\alpha_1 + B(1)\alpha_2 + B(2)\alpha_3 + B(3)\alpha_4 + B(4)\alpha_5 = B(1), \\ B(1)\alpha_1 + B(0)\alpha_2 + B(1)\alpha_3 + B(2)\alpha_4 + B(3)\alpha_5 = B(2), \\ B(2)\alpha_1 + B(1)\alpha_2 + B(0)\alpha_3 + B(1)\alpha_4 + B(2)\alpha_5 = B(3), \\ B(3)\alpha_1 + B(2)\alpha_2 + B(1)\alpha_3 + B(0)\alpha_4 + B(1)\alpha_5 = B(4), \\ B(4)\alpha_1 + B(3)\alpha_2 + B(2)\alpha_3 + B(1)\alpha_4 + B(0)\alpha_5 = B(5). \end{cases}$$

然后将 $B(0), B(1), B(2), B(3), B(4), B(5)$ 的值代入到上述方程组中,运用 Matlab 数学软件计算出 $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$ 的值分别为

$$\alpha_1 = -0.305, \alpha_2 = -0.123, \alpha_3 = -0.203, \alpha_4 = 0.079, \alpha_5 = -0.058.$$

将 $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$ 的值代入预测方程(6)中,得

$$\hat{y}'_{2011} = -0.305y'_{2010} - 0.123y'_{2009} - 0.203y'_{2008} + 0.079y'_{2007} - 0.058y'_{2006} = -0.305 \times 1159.333 - 0.123 \times (-795.667) - 0.203 \times 1269.333 + 0.079 \times 388.333 - 0.058 \times 26.333 = -483.689.$$

故由式(8)得 2011 年 6 月降水量的预测值 $\hat{y}_{2011} = \hat{y}'_{2011} + \bar{y} = -483.689 + 4147.667 = 3663.978$.

3 预测结果

利用(1)式,预测出桂林 2011 年各月的气温,得到预测结果如表 2。再根据桂林 1987~2010 年 9 月、11 月和 12 月的降水量实测数据,对 2011 年 9 月、11 月和 12 月的降水量进行预测,得到的结果如表 5 所示。

从表 4 可以看出,利用季节指数法建立的气温预测模型的预测效果从整体来说都比较好的,除了 1 月、3 月和 11 月,其余月份的预测绝对误差都在 5% 以内,模型预测的精度高。

表 4 2011 年气温预测分析结果表

Table 4 Prediction results of air temperatures in 2011

月份 Month	预测值(0.1°) Predictive value	实测值(0.1°) True value	绝对误差(%) Absolute error
1	83.111	72	15.4
2	111.282	115	3.2
3	143.363	128	12.0
4	194.493	200	2.8
5	238.304	227	5.0
6	264.470	276	4.2
7	285.825	297	3.8
8	284.321	291	2.3
9	261.563	255	2.6
10	219.857	207	6.2
11	163.114	182	10.4
12	103.964	100	4.0

表 5 2011 年降水预测误差

Table 5 Prediction errors for precipitation in 2011

月份 Month	预测值(0.1mm) Predictive value	实测值(0.1mm) True value	绝对误差(%) Absolute error
6	3663.978	3978	7.89
9	658.869	632	4.3
11	345.646	368	6.0
12	118.08	131	9.8

从表 5 可以看出,2011 年降水预测值与真实值的误差都较小。这说明平稳时间序列的线性外推法建立的预测模型具有较高的精度,可以作为降水预测的一个有效方法。

参考文献:

- [1] 石昱馨,刘喜波.乘积季节模型在气温及降水时间序列分析中的应用[J].数学·力学·物理学高新技术交叉研究进展,2010(13):792-797.
- [2] 迟道才,张宁宁,袁吉,等.时间序列分析在辽宁朝阳地区干旱灾变中的应用[J].沈阳农业大学学报,2006,37(4):627-630.
- [3] 王燕.应用时间序列分析[M].北京:中国人民大学出版社,2008.
- [4] 唐延芳.灌溉用水量预测方法的研究——以铁甲灌区为例[D].沈阳:沈阳农业大学,2007.

(责任编辑:尹 闯)