

文章编号: 1001-6791 (2000) 04-0451-06

一种新的时间序列综合分析和应用^{*}

李 玲¹, 施 勇²

(1. 南京经济学院, 江苏 南京 210003; 2. 水利部南京水文水资源研究所 江苏 南京 210024)

摘要: 在传统时间序列分解计算方法的基础上, 采用一种新的灰色预测方法模式对趋势项进行预测, 建立以传统分析方法和灰色系统理论相结合的长期预测数学模型, 对 1985 年蚌埠站、峰山站、明光站等月流量进行预测, 结果令人满意。

关 键 词: 时间序列; 综合分析; 趋势; 周期; 随机; 灰色理论

中图分类号: p 338 **文献标识码:** A

时间序列预测的传统分解法包括趋势预测、季节预测 (周期预测) 和随机预测, 过去趋势预测主要采用曲线配合方法, 然后进行时间外推。该方法适合于原始时间序列规律性较好的情况, 若时间序列中包含了随机因素的影响, 曲线配合法所得出的预测结果可能会失真。灰色预测方法 (GM (1, 1)) 模型不是直接使用原始数据进行预测, 而是对由原始数据产生的生成数作预测, 一方面, 将离散变量连续化, 用微分方程代替差分方程, 使得数学处理变得比较容易, 另一方面, 即使原始数据的变化规律不明显, 一旦形成生成数后, 由于累加作用, 抵消了不少随机因数的影响, 规律性容易表现出来。但由于 GM (1, 1) 模型中, 背景值采用一次累加系列的前后时刻变量的平均值, 其适用的范围局限于变量变化较平稳的时间序列, 而对变化较剧烈的时间序列进行模拟和预测时产生较大的误差。针对变化较剧烈的时间序列存在趋势项预测滞后的问题, 对 GM (1, 1) 模型中的背景值的表达式进行修正, 不仅能适应变化平缓的时间序列, 而且能够适用于时间变化剧烈的序列

1 建模思路

从时间系列变量自身演变规律这一角度出发, 采用时间序列预测的传统分解法来建立时间序列中长期预报模型。

时间序列 $X(t)$, 一般由趋势项 $A(t)$ 、周期项 $P(t)$ 、突变项 $B(t)$ 和随机项 $R(t)$ 组成, 表达式为

$$X(t) = A(t) + P(t) + R(t)$$

时间序列可表示为三个组成部分之和。趋势项 $A(t)$ 反映变量的季节性趋势或多年变化趋势; 周期项 $P(t)$ 反映变量的周期性变化。趋势项、周期项这两项反映了时间序列变量变化中

^{*} 收稿日期: 2000-02-08; 修订日期: 2000-06-12

作者简介: 李玲 (1963-), 女, 江苏南京人, 南京经济学院讲师, 主要从事随机理论研究与应用。

的确定性成分, 在把这二项分离出去, 余下的就是随机项了。随机项可用平稳的时间序列来分析, 随机项 $R(t)$ 可分为两项: 平稳时间序列项 $S(t)$ 和纯随机项 $N(t)$, 即

$$R(t) = S(t) + N(t)$$

纯随机项 $N(t)$ 作为白噪声来处理。

在将时间序列变量分解成四个组成部分之后, 就可按各组成项的变化规律对未来时刻进行外推, 再将各项合成作出预报。

2 趋势项分析

采用灰色系统理论来提取时间序列中的趋势项。从灰色系统理论的观点看, 随机变量可视作为在一定范围内变化的灰色量, 而随机过程可视作为在一定范围内变化的灰色过程。灰色系统理论认为: 对于一个预测系统, 必然是具有已知的信息, 又有未知信息或不确知信息, 且处于被动变化之中。尽管系统的行为现象是朦胧的, 数据是杂乱的, 但它毕竟是有序的, 在这杂乱无章的数据背后, 必然潜在着某种规律。

由于时间序列变量自身变化的复杂性, 受诸多因素制约, 具有很大的不确定性, 所以时间序列变量系统就是一个处于动态变化之中的灰色系统, 可用 $GM(1, 1)$ 建模。

$GM(1, 1)$ 模型是单序列的一阶线性动态模型, 设某原始数据系列为

$$x^{(0)} = \{x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)}\} \quad (1)$$

由它产生的 1 次累加生成 1-AGO 为

$$x^{(1)} = \{x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}\} \quad (2)$$

$$\text{式中} \quad x_t^{(1)} = \sum_{i=1}^t x_i^{(0)} = x_{t-1}^{(1)} + x_t^{(0)} \quad (t = 1, 2, \dots, n) \quad (3)$$

由于生成序列接近指数曲线, 故可认为是光滑的离散系数, 可用微分方程描述。一阶带系数的微分方程表达式为

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \quad (4)$$

其解 (即预测方程) 可写成

$$x^{(1)}(t+1) = (x^{(0)}(1) - \frac{b}{a})e^{-at} + \frac{b}{a} \quad (5)$$

式中参数 a , b 的估计, 应用最小二乘法估参, 写成矩阵形式

$$Y = XB \quad (6)$$

其中

$$X = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix} \quad Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \quad B = \begin{bmatrix} a \\ b \end{bmatrix}$$

式中背景值 $z^{(1)}(t+1)$: $z^{(1)}(t+1) = \frac{1}{2} (x^{(1)}(t) + x^{(1)}(t+1)) \quad t=1, 2, \dots, n-1$

$$B = (X'X)^{-1}X'Y \quad (7)$$

则

在计算出 $\hat{x}(t)$

其还原数据为 $\hat{x}^{(0)}(t) = x^{(1)}(t) - x^{(1)}(t-1) \quad (8)$

讨论背景值 $z^{(1)}(t+1)$ 的选取: $z^{(1)}(t+1)$ 是 $[t, t+1]$ 时段内 $dx^{(1)}/dt$ 的背景值, 当时间间隔较小, 系列变量变化平稳时, 背景值由生成系列的前后时间变量的平均值构成是适合的, 因为生成系列的趋势线与时间轴围成的面积可采用梯形面积近似处理, 模型预测的误差较小。当系列变量的变化剧烈时, 其趋势线与时间轴围成的面积小于由 $x^{(1)}(t), x^{(1)}(t+1), t, t+1$ 构成的梯形面积, 若仍采用这种背景值并进行预测, 往往产生较大的滞后误差, 为了克服这缺点, 需重新构造背景值的表达式。

构造新背景值的设计思想是使在 $[t, t+1]$ 时段内背景值所确定的面积与生成系列的趋势线与时间轴围成的面积近似相等。首先对生成系列进行函数拟合, 当然在拟合时尽量使得在 x_k 的拟合值与实测值重合, 而后在生成系列变化剧烈时段 $[t, t+1]$ 内分成 n 等分, 则时间 $t_i, i=1, 2, \dots, n$ 对应 $x^{(1)}_i$ 的值由拟合函数确定, 将 n 个小梯形面积相加就可以得到 $[t, t+1]$ 的背景值。这种背景值的构成仍保留原背景值构成的简洁, 无需多次迭代, 又能使得背景值梯形面积与实际面积接近, 提高模型拟合和预测精度。

3 周期项分析

时间序列分离趋势之后, 可将剩余的序列 $X(t) - A(t)$ 进行周期分析。本文采用简单分波法, 是将时间序列 (即分离趋势之后的序列) 看成是由不同周期的规则波动迭加而成。因而在分离周期时, 是逐步分解出一些比较明显的周期波, 然后叠加起来作为该时间序列的周期项, 把这个周期项进行外推就可用于预测。

方差分析法简述如下

设有时间序列 $x_1, x_2, \dots, x_n$

(1) 在分析周期之前, 事先并不知道这一序列存在多少个周期, 所以要根据序列长度, 列出可能存在的周期, 逐个试验, 即
式中 n 为序列长度, k 为周期个数。

$$k = \left\{ \left\lfloor \frac{n}{2} \right\rfloor \right\} = \begin{cases} \frac{n}{2} & n \text{ 为偶数} \\ \frac{n-1}{2} & n \text{ 为奇数} \end{cases} \quad (9)$$

(2) 将时间序列按每一试验周期排列, 计算离均差平方和, 包括排列组内离均差平方和及组间离均差平方和。

组内离均差平方和

$$Q_2^2 = \sum_{j=1}^k \sum_{i=1}^m (y_{ij} - \bar{x}_j)^2, \quad \bar{x}_j = \frac{1}{m} \sum_{i=1}^m y_{ij} \quad (10)$$

式中 m 为组内项数; y_{ij} 为时间序列 x_i 按试验周期分组排列后的序列值, 其自由度 $f_2 = n - k$ 组间离均差平方和

$$Q_3^2 = \sum_{j=1}^k m(\bar{x}_j - \bar{x}), \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n \bar{x}_i \quad (11)$$

其自由度 $f_3 = k - 1$

(3) 计算试验周期的方差比。

$$F = \frac{Q_3^2/f_3}{Q_2^2/f_2} \quad (12)$$

(4) 方差的 F 检验。选定某一信度 $\alpha = 0.05$, 查 F 分布表得 F_α , 由 F 是否大于 F_α , 即可分析出方差是否存在显著性差异, 如是, 则该试验周期存在, 反之, 则不存在。

(5) 从原序列中将第一周期滤去, 剩余序列再重复上述步骤, 即可继续提取第二、第三周期, 直至无显著周期为止。

(6) 将提取出的各周期波迭加即可求出周期项。

4 随机项分析

一般认为, 时间序列在分离出趋势项、周期项之后, 剩余的序列具有平稳随机过程的性质, 因此可以用平稳随机过程的方法来分析。前面所述, 随机项还可分离成平稳时间序列项 $S(t)$ 和纯随机项 $N(t)$ (白噪声), 因此还可以分离求解。

对平稳时间序列可建立自回归模型 $AR(P)$:

$$X(t) = \beta_{p,1}X(t-1) + \beta_{p,2}X(t-2) + \dots + \beta_{p,p}X(t-p) \quad (13)$$

式中 $X(t)$ 为时刻 t 的变量序列值, 一般以离均差表示; $\beta_{p,j}$ ($j = 1, 2, \dots, P$) 为自回归系数, P 为模型阶数。

自回归系数依据实测变量序列数据, 利用最小二乘法, 建立

$$\begin{cases} \beta_{k,1} = Y_1 \\ \beta_{k,k} = \frac{Y_k - \sum_{j=1}^{k-1} \beta_{k-1,k} Y_{k-1}}{1 - \sum_{j=1}^{k-1} \beta_{k-1,j} Y_j} & (k = 2, 3, \dots) \\ \beta_{k,j} = \beta_{k-1,j} - \beta_{k,k} \beta_{k-1,k-j} & (j = 1, 2, \dots, k-1) \end{cases} \quad (14)$$

Yule-Walker 方程组, 采用递推公式求解。

式中 Y_k 为 $X(t)$ 的 k 阶样本自相关系数。

$$Y_k = \frac{\sum_{t=1}^{n-k} X_t X_{t+k}}{\sum_{t=1}^n X_t^2} \quad (15)$$

模型阶数 P , 可通过 AIC 准则来确定。

$$AIC = \text{Min} \left\{ n \ln \frac{\sum (x_i - \bar{x})^2}{n - P - 1} + 2P \right\} \quad (16)$$

式中 X_i 为时间序列; $\bar{X}$ 为均值; n 为时间序列长度; P 为模型阶数。

对纯随机项, 本研究中对 $N(t)$ 采用同平稳项 $S(t)$ 一样, 用 $AR(P)$ 模型来求解。

在对时间序列的四个组成项分别建立分析模型的基础上, 合成各项的预测值, 即可作为时间序列的预报。

5 预报模型的应用

为了检验模型的模拟功能和 GM (1, 1) 背景值构成的改进对模拟精度的影响, 选用 1985 年蚌埠站、峰山站、明光站、戴村坝站和临清站月流量值对背景值构成的改进前后分别预测, GM (1, 1) 背景值构成模型预测结果如表 1, 改进后模型预测结果如表 2。

表 1 GM (1, 1) 模式改进前, 1985 年月流量预报成果^[1] $\text{m}^3 \cdot \text{s}^{-1}$

Table 1. The GM (1, 1) model, month discharge forecasted in 1985

月份	蚌埠站			峰山站			明光站			戴村坝站			临清站		
	差值	实测	预测	差值	实测	预测	差值	实测	预测	差值	实测	预测	差值	实测	预测
1 月	686	536	-150	24	9.7	-14.3	5.6	5.7	0.1	5.3	8.6	3.3	17.5	11.2	-6.3
2 月	455	442	-13	29.4	23.4	-6	5.7	7.7	2	5.9	7.7	1.8	6.2	5.6	-0.6
3 月	619	535	-84	37.6	27.2	-10.4	7.1	7.2	0.1	3.3	3.5	0.2	0.9	3.3	2.4
4 月	421	547	126	9.6	7.9	-1.7	6.5	10.3	3.8	0.2	3.9	3.7	1.7	9.6	7.9
5 月	1928	1239	-689	16.2	21.2	5	51.6	30.8	-20.8	12.1	23.6	11.5	12	11.4	-0.6
6 月	1058	788	-270	8.4	6.8	-1.6	9.3	6.3	-3	2.1	2	-0.1	12.3	13.1	0.8
7 月	1510	1364	-146	18.9	22.8	3.9	81.7	24.6	-57.1	117.1	104	-13.1	15.3	7.6	-7.7
8 月	1018	1111	93	4.7	6.6	1.9	8	4.2	-3.8	107.7	79.7	-28	23.4	31.7	8.3
9 月	1094	1306	212	108.1	48.7	-59.4	18.4	21.4	3	116.7	50.7	-66	59.1	76.6	17.5
10 月	1655	1833	178	153.4	26.2	-127.2	66.8	52.3	-14.5	70.7	62.1	-8.6	60.1	42.5	-17.6
11 月	1679	1282	-397	148.2	67.8	-80.4	14.6	10.7	-3.9	37.4	22	-15.4	60.9	32.2	-28.7
12 月	473	427	-46	38.7	24.1	-14.6	4.6	7.3	2.7	24.4	17	-7.4	23.8	14.4	-9.4

从表 1 可看出, 若采用 GM (1, 1) 改进前模式, 对于枯水季节 1 月和 2 月的流量预测值的精度较好, 当雨季到来时, 流量增加幅度较大, 如 4 月到 5 月蚌埠站流量增加 1 500 m^3/s , 5 月流量的预测值明显偏小。表 2 采用 GM (1, 1) 改进后模式预测, 拟合和预测结果较表 1 结果预测精度有较大提高, 尤其对于流量增加幅度较大的月份, 克服了预测滞后的问题。两种模式的预测过程见图 1、图 2。

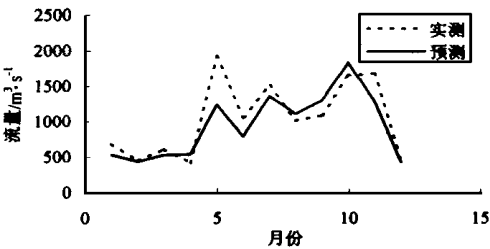


图 1 GM (1,1) 模式改进前, 蚌埠站月流量预测
Fig. 1. The GM (1, 1) model, month discharge forecasted in Bangpu station

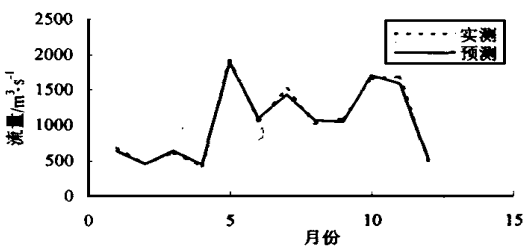


图 2 GM (1, 1) 模式改进后, 蚌埠站月流量预测
Fig. 1. The new GM (1, 1) model, month discharge forecasted in Bangpu station

表 2 GM (1, 1) 模式改进后, 1985 年月流量预报成果 $\text{m}^3 \cdot \text{s}^{-1}$

Table 2. The new GM(1, 1) model. month discharge forecasted in 1985

月份	蚌埠站			峰山站			明光站			戴村坝站			临清站		
	差值	实测	预测	差值	实测	预测	差值	实测	预测	差值	实测	预测	差值	实测	预测
1 月	686	636	-50	24	21.7	-2.3	5.6	6.2	0.6	5.3	6.6	1.3	17.5	18.1	0.6
2 月	455	462	7	29.4	27.4	-2	5.7	6.1	0.4	5.9	5.7	-0.2	6.2	6.9	0.7
3 月	619	635	16	37.6	40.2	2.6	7.1	6.8	-0.3	3.3	4.1	0.8	0.9	0.7	-0.2
4 月	421	447	26	9.6	8.9	-0.7	6.5	7.3	0.8	0.2	0.9	0.7	1.7	2.6	0.9
5 月	1928	1889	-39	16.2	20.1	3.9	51.6	54.8	3.2	12.1	13.1	1	12	13.2	1.2
6 月	1058	1088	30	8.4	7.8	-0.6	9.3	11.3	2	2.1	1.9	-0.2	12.3	10.1	-2.2
7 月	1510	1434	-76	18.9	20.3	1.4	81.7	78.6	-3.1	117.1	124	6.9	15.3	16.1	0.8
8 月	1018	1061	43	4.7	5.2	0.5	8	9.2	1.2	107.7	119.7	12	23.4	25.7	2.3
9 月	1094	1046	-48	108.1	98.9	-9.2	18.4	20.1	1.7	116.7	120.7	4	59.1	66.3	7.2
10 月	1655	1704	49	153.4	134.6	-18.8	66.8	67.4	0.6	70.7	69.1	-1.6	60.1	57.5	-2.6
11 月	1679	1582	-97	148.2	156.7	8.5	14.6	15.7	1.1	37.4	34.6	-2.8	60.9	56.2	-4.7
12 月	473	507	34	38.7	43.2	4.5	4.6	3.9	-0.7	24.4	27	2.6	23.8	24.1	0.3

6 结 语

在原 GM (1, 1) 模式的基础上, 改进背景值的构成, 建立传统分析方法和灰色系统理论相结合的长期预测数学模型, 在分析预测水文时间序列过程中, 发现模型预测能力与精度较 GM (1, 1) 模式改进前分析预测模型有较大的提高。具体表现为改进后的 GM (1, 1) 模式, 增强了模型对受诸多因素制约, 自身变化复杂和变量剧烈变化的处理能力。因此该模型拓宽了应用范围, 不仅能模拟变量变化缓慢的时间系列, 而且能预测剧烈变化的时间系列, 确保预测结果的合理性。

参考文献:

[1] 刘国纬. 跨流域调水运行管理[M]. 北京: 中国水利水电出版社, 1995. 93- 103.
[2] 潭冠军. GM (1, 1) 模型的背景值构造方法和应用[J]. 系统工程理论与实践, 2000, 20(4). 99- 103
[3] 邓聚龙. 灰色预测与决策[M]. 武汉: 华中理工大学出版社, 1986.

A New Time Serial Synthetical Analysis Method

LI Ling¹, SHI Yong²

(1. Nanjing Economy College, Nanjing, 21003, China;

2. Nanjing Institute of Hydrology and water Resources, Nanjing 210024, China)

Abstract: Based on the traditional method of time serial analysis, a new model of GM (1, 1) is adopted and a mathematics model is built by combining the analysis method with the grays theory. When applied the model to the monthly discharge forecasting of Bengful, Fengshan and other stations in 1985, the results are satisfactory.

Key words: time series; synthetical analysis; tendency; period; randomicity; gray theory