城市用水量预测中的多变量灰色预测模型

楼 玉[1],张 清[2],刘国华[1],范庆来[1]

(1.浙江大学建筑工程学院,浙江 杭州 310027;2.耀江房地产开发有限公司,浙江 杭州 310006)

摘要:当观测资料的数据量少,而又存在多个相互影响或关联的变量时,常用的灰色预测模型 GM(1,1)不能全面考虑多个变量。采用自适应 MGM(1,n)模型——多变量灰色预测模型,较好地解决了这一问题。MGM(1,n)模型是 GM(1,1)模型在 n 元多变量情况下的推广,通过联立求解 n 个 n 元微分方程,使模型中的参数能够反映实际工程或社会系统中多个变量间相互影响、相互制约的关系。以杭州市用水量统计资料为研究对象,运用灰色理论建立 MGM(1,n)模型,获得了较好的预测效果。

关键词:灰色系统;MGM(1,n)模型;用水量

中图分类号:TU991.31 文献标识码:A 文章编号:1004-6933(2005)01-0011-03

Application of grey multi-variable forecasting model in predicting urban water consumption

LOU Yu[1], ZHANG Qing[2], LIU Guo-hua[1], FAN Qing-lai[1]

(1. *College of Civil Engineering and Architecture of ZheJiang University*, *Hangzhou* 310027, *China*; 2. *Yaojiang Real Estate Co. Ltd*, *Hangzhou* 310006, *China*)

Abstract: The general grey model GM (1,1) cannot take multi variables into account in forecast problems with several variables relating with each other and few observation data available. In this paper, a self-adapting MGM (1,n) model, which is a grey multi-variable forecasting model, is introduced to solve this problem. The MGM (1,n) model is an extension of GM (1,1) in the case of n variables. Through the application of the grey MGM (1,n) model to statistical data of water consumption in Hangzhou City, satisfying results are obtained. The results show that the proposed method is feasible and effective for application.

Key words: grey system; MGM (1,n) model; water consumption

常用用水量预测方法可分为两类:一类是解释性预测方法,即找出被预测量的各影响因素,建立回归分析模型;另一类为时间序列分析方法,它只依赖于被预测量的历史观测数据及数据模式,通过序列分析,找出其顺序变化规律[1]。因为城市用水量受经济、人口、生活水平等多种因素的影响,具有一定的灰色特征,特别是没有较长系列的可靠数据可以利用时,根据近期少量数据和它们的发展态势,通过利用灰色系统模型进行预测,常不失为一种有效途径[2]。

目前普遍采用的 GM(1,1)模型,仅利用单一的时间序列数据,无法反映多个变量间的相互影响、协同发展与制约情况,而且 GM(1,n)模型又主要描述变量间的相互关系,不用于预测[3]。为此本文采用能够考虑多个相关变量的自适应灰色预测模型 MGM(1,n),它是 GM(1,1)模型在 n 元多变量情况下的推广,但不是 GM(1,1)模型的简单组合,也不同于 GM(1,n)模型只建立单个 n 元一阶微分方程,而是建立 n 个 n 元微分方程,通过联立求解,使 MGM(1,n)模型中的参数能够反映变量间的相互影响。文中列出了对杭州市用水量预测的应用示例,对比数据表明预测效果良好。

作者简介:楼玉(1980—),女,浙江诸暨人,硕士研究生,研究方向为给排水及系统工程.E-mail:dairy_zj@163.com

1 MGM(1,n)灰色模型的基本方程

MGM(1,n)模型采用生成数列建模。设问题有 n 个 m 维时间序列数据,每个序列代表系统的一个因素变量的动态行为,即有

$$x_i^{(0)} = \{x_i^{(0)}(1), x_i^{(0)}(2), \cdots, x_i^{(0)}(m)\} \quad (1)$$

对 $x_i^{(0)}$ 作一次累加生成,得

$$x_i^{(1)} = \{x_i^{(1)}(1), x_i^{(1)}(2), \cdots, x_i^{(1)}(m)\}$$
$$i = 1,2,\cdots,n$$
$$x_i^{(1)}(k) = \sum_{j=1}^{k} x_i^{(0)}(j) \quad (2)$$

则 MGM(1,n)模型的一阶常微分方程组的形式为

$$\left.\begin{aligned} \frac{dx_1^{(1)}}{dt} &= a_{11}x_1^{(1)} + a_{12}x_2^{(1)} + \cdots + a_{1n}x_n^{(1)} + b_1 \\ \frac{dx_2^{(1)}}{dt} &= a_{21}x_1^{(1)} + a_{22}x_2^{(1)} + \cdots + a_{2n}x_n^{(1)} + b_2 \\ &\square \\ \frac{dx_n^{(1)}}{dt} &= a_{n1}x_1^{(1)} + a_{n2}x_2^{(1)} + \cdots + a_{nn}x_n^{(1)} + b_n \end{aligned}\right\} \quad (3)$$

式中 :$x_1^{(1)}, x_2^{(1)}, \cdots, x_n^{(1)}$ 为一次累加生成数列 ;a_{ij}, b_i 为系统参数。

如记

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \square & & \square \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

$$\boldsymbol{B} = \begin{bmatrix} b_1 \\ b_2 \\ \square \\ b_n \end{bmatrix} \qquad \boldsymbol{X}^{(1)} = \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ \square \\ x_n^{(1)} \end{bmatrix}$$

则式(3)可记为

$$\frac{d\boldsymbol{X}^{(1)}}{dt} = \boldsymbol{A}\boldsymbol{X}^{(1)} + \boldsymbol{B} \quad (4)$$

式(4)中的参数向量 $\boldsymbol{a}_i$ 可用最小二乘法估计得,即:

$$\boldsymbol{a}_i = [\hat{a}_{i1} \quad \hat{a}_{i2} \quad \cdots \quad \hat{a}_{in} \quad \hat{b}_i]^{\mathrm{T}} = (\boldsymbol{L}^{\mathrm{T}}\boldsymbol{L})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{Y}_i$$
$$i = 1,2,\cdots,n \quad (5)$$

其中:

$$L = \begin{bmatrix} \frac{1}{2}\left(x_1^{(1)}(2) + \frac{1}{2}x_1^{(1)}(1)\right) & \cdots & \frac{1}{2}\left(x_n^{(1)}(2) + \frac{1}{2}x_n^{(1)}(1)\right) & 1 \\ \square & & \square & \\ \frac{1}{2}\left(x_1^{(1)}(m) + \frac{1}{2}x_1^{(1)}(m-1)\right) & \cdots & \frac{1}{2}\left(x_n^{(1)}(m) + \frac{1}{2}x_n^{(1)}(m-1)\right) & 1 \end{bmatrix}$$

$$\boldsymbol{Y}_i = [x_i^{(0)}(2) \quad x_i^{(0)}(3) \quad \cdots \quad x_i^{(0)}(m)]^{\mathrm{T}}$$

灰色 MGM(1,n)模型的时间响应序列为

$$\hat{X}^{(1)}(k) = e^{\hat{A}(k-1)}X^{(1)}(1) + \hat{A}^{-1}(e^{\hat{A}(k-1)} - I)\hat{B}$$
$$k = 1,2,\cdots,m \quad (6)$$

式(6)中矩阵指数函数 $e^{\hat{A}t}$ 可按下式计算:

$$e^{\hat{A}t} = I + \hat{A}t + \frac{\hat{A}^2}{2!}t^2 + \cdots = I + \sum_{k=1}^{\infty}\frac{A^k}{k!}t^k \quad (7)$$

通过累减生成,还原为相应变量的原数列值:

$$\hat{X}^{(0)}(k) = \hat{X}^{(1)}(k) - \hat{X}^{(1)}(k-1)$$
$$k = 2,3,\cdots,m \quad (8)$$
$$\hat{X}^{(0)} = \hat{X}^{(0)}(1)$$

用灰色系统模型进行短期预测较为成功。随着预测期增长,未来的扰动或随机因素会对系统产生重大的影响。为了预测稍长时刻的系统情况,可采用等维灰数递补模型 :将获得的预测数据,如 $\hat{x}^{(0)}(m+1)$,充实到原始数列中,并去掉最老的数据,如 $\hat{x}^{(0)}(1)$,形成等维的新数列用以建模,并预测下一步:

$$\hat{x}^{(0)} = \left\{ x^{(0)}(k) \,\middle|\, k = 2,3,\cdots,m \text{ 和 } \hat{x}^{(0)}(m+1) \right\} \quad (9)$$

这类动态模型具有适应环境的特点,亦称为自适应模型。

2 计算实例

为了验证上述模型与方法的实用程度,下面对杭州市用水量历年统计资料[4]进行分析与计算。杭州市的统计数据不多且前期统计有间断,故仅利用后期连续的等间距(10 年)数据 9 组,见表 1。

表 1 杭州市用水量统计

年份	年供水量/亿 t	GDP 发展指数/%	人口/万人	生活价格指数/%
1992	26.20	5.30	136.30	3.19
1993	27.73	6.80	138.33	3.55
1994	28.75	8.40	141.27	4.24
1995	30.43	10.05	143.52	5.03
1996	31.72	11.50	166.73	5.56
1997	31.56	13.14	169.29	5.93
1998	32.04	14.41	171.89	6.04
1999	31.70	15.65	175.27	6.06
2000	32.61	17.40	179.18	6.12

注 :GDP 发展指数和生活价格指数都以 1978 年为基准.

2.1 建模计算

对原始数据进行初值化处理后,用斜率关联度法计算得 GDP 发展指数、人口、生活价格指数的关联系数分别为 0.78,0.71,0.83,可见所取的影响因素与城市用水量之间的关系比较密切,故本文在预测模型中取城市用水量、GDP 发展指数、人口、生活价格指数 4 个变量(分别用 x_1, x_2, x_3, x_4 表示)建立 MGM(1,4)模型。

取等时间间距的 1 ~ 6 组数据,按式(3)建立 MGM(1,4)模型,取表 1 中数据求得相应的 $\boldsymbol{L}$ 与 $\boldsymbol{Y}$,

表 2　逐点残差(1992 ~ 1997 年)

年份	用水量/亿 t			GDP 发展指数/%			人口/万人			价格指数/%		
	实测值	拟合值	绝对误差	实测值	拟合值	绝对误差	实测值	拟合值	绝对误差	实测值	拟合值	绝对误差
1992	26.20	26.20	0	5.30	5.30	0	136.30	136.30	0	3.19	3.19	0
1993	27.73	27.66	0.070	6.80	6.80	0	138.33	137.53	0.80	3.55	3.54	0.01
1994	28.75	28.94	0.188	8.40	8.42	0.02	141.27	139.53	1.75	4.24	4.30	0.06
1995	30.45	30.47	0.044	10.05	10.00	0.05	143.52	150.12	6.60	5.03	5.02	0.01
1996	31.72	31.53	0.182	11.50	11.58	0.08	166.73	162.16	4.57	5.56	5.59	0.03
1997	31.56	31.87	0.310	13.14	13.09	0.05	169.29	170.93	1.64	5.93	5.94	0.01
平均相对误差/%			0.430			0.29			1.69			0.42

注 相对误差以各变量实测值为分母。

进一步按式(6),(7)计算时间响应序列,得 $\widehat{X}_1^{(1)}(k)$,$\widehat{X}_2^{(1)}(k)$,$\widehat{X}_3^{(1)}(k)$,$\widehat{X}_4^{(1)}(k)$,$k=1,2,\cdots,9$,再按照式(8)还原计算即可得拟合-预测值。

2.2　模型精度检验

为了检验所建模型是否合适,对所建自适应 MGM(1,4)模型拟合值和实测值的误差 $e_i(k)=x_i^0(k)-\hat{x}_i^0(k)$进行逐点检验,结果见表 2。

由表 2 可见,残差值并不大,平均相对误差在 2%以下,特别是预测的主要变量(用水量)的平均相对误差小于 1%。

2.3　预测结果比较

利用上述所建 MGM(1,4)模型,预测得第 7 组数据,将预测值加入原数据组并去掉最老的第 1 组数据,建立 MGM(1,4)模型,继续预测出第 8 组数据,加上刚才预测得的第 7 组数据,并同时去掉第 2 组数据之后建模。如此反复,数据‘新陈代谢”,一直预测出序号 9 时刻的沉降,结果如表 3 所示。为了对比,表 3 中还列出了自适应 GM(1,1)模型的预测结果。

表 3　两种模型预测用水量

用水量/亿 t	自适应 GM(1,1)模型			自适应 MGM(1,4)模型		
	预测值/亿 t	绝对误差/亿 t	相对误差/%	预测值/亿 t	绝对误差/亿 t	相对误差/%
32.62(1998 年)	32.04	0.59	1.84	31.63	0.41	1.28
33.59(1999 年)	31.70	1.89	5.97	31.26	0.44	1.40
34.58(2000 年)	32.61	1.97	6.05	31.26	1.35	4.15
平均相对误差/%			4.62			2.27

注 相对误差以用水量实际值为分母.

由表 3 可见,自适应 MGM(1,4)模型预测精度比自适应 GM(1,1)模型的预测精度高,这是因为 MGM(1,4)模型比 GM(1,1)更全面地考虑了相关变量的结果。两个模型的拟合和预测值情况见图 1,同样可以看出 MGM(1,4)模型的拟合效果也要比 GM(1,1)模型的拟合效果好。

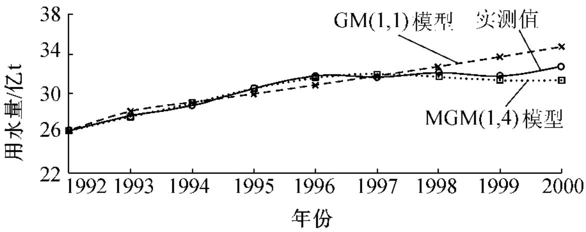

图 1　两种模型拟合曲线和用水量预测

3　结　语

本文讨论了多变量灰色预测模型 MGM(1,n)及其在城市用水量预测中的应用。城市用水量是一个复杂的非线性系统,受多种因素影响,而 MGM(1,n)模型既能适用于小样本量的情况,又较好地考虑了多个相关因素,因而预测精度较高,在工程实践中具有良好的应用价值。

同时应该指出,在 MGM(1,n)建模过程中,如果所选变量不当,矩阵及其逆矩阵会出现病态,导致参数失真,所以应该避免采用线性相关系数过高的一个变量同时进入微分方程组。如何恰当选择变量建立 MGM(1,n)模型和改进模型参数估计,还需在实际应用中进一步研究和改善。

参考文献:

[1] 李斌,许仕荣,柏光明,等.灰色—神经网络组合模型预测城市用水量[J].中国给水排水,2002,18(2):66 ~ 68.
[2] 汪树玉.系统分析[M].杭州:浙江大学出版社,2002.152 ~ 159.
[3] 翟军,盛建明,冯英浚,等.MGM(1,n)灰色模型及应用[J].系统工程理论与实践,1997(5):109 ~ 113.
[4] 俞亭超.自适应遗传—人工神经元网络模型及在城市用水量预测中的应用研究[D].杭州:浙江大学,2001.
[5] 刘思峰,郭天榜,党耀国,等.灰色系统理论及其应用[M].北京:科学出版社,1999.

(收稿日期:2003-10-08　编辑:傅伟群)