

文章编号: 1007-4929(2007)05-0038-03

# 灰色系统与多元逐步回归耦合模型 在年用水量预测中的应用

黄 胜

(西南科技大学环境与资源学院,四川 绵阳 621002)

**摘 要:**探讨了将多元回归分析模型和灰色系统 GM(1,1)耦合应用于北京市总用水量预测的方法和技术。对北京市 1980~2001 年的序列资料,利用灰色系统的关联度分析和逐步回归的思想,通过比较和选择,建立了多元线性逐步回归方程;将用 GM(1,1)模型拟合反推得到的预测值带入方程求出总用水量的未来值。结果表明:人均日生活用水量和工业用水量是影响北京市总用水量变化的主要因素。灰色系统与多元线性回归的耦合模型用于对城市总用水量的预测是行之有效的。

**关键词:**灰色系统;逐步回归分析;耦合模型;年用水量;预测

**中图分类号:**TV213.4 **文献标识码:**A

## Prediction of Annual Water-consumed Quantity Based on Coupling Model of Grey-system and Multiple Stepwise Regression

HUANG Sheng

(School of Environment and Resources, Southwest University of Science and Technology, Mianyang 621002, Sichuan Province, China)

**Abstract:** The coupling model of Grey-system and multiple stepwise regression was discussed and applied to the prediction of annual water-consumed quantity of Beijing. Firstly, using the method of correlation analysis and stepwise regression of grey system, the data sequences of Beijing from 1980 to 2001 were compared and selected to set up multiple stepwise regression formula. Then, the prediction value gained through GM(1,1) model was brought to the formula to get the future value of the total water quantity. The result showed that the water-consumed quantity per day, per person and industrial water-consumed quantity were the main influence factors for the change of total water quantity. It is feasible to apply the coupling model of grey-system and multiple stepwise regression for the prediction of total water quantity of the city.

**Key words:** Grey-system; stepwise regression; coupling model; annual water-consumed quantity; prediction

目前,对年用水量的预测采用的主要方法有:灰色系统预测模型、数理统计方法和神经网络等<sup>[1~4]</sup>。目前的方法在年用水量的预测中取得了一定的成果,但预测模型多为单一模型,难以反映年用水量与其主要影响因素之间的复杂关系。为此,本文提出一个灰色系统与多元逐步回归耦合模型来研究年用水量,并以北京市为例,建立北京市总用水量的耦合模型,预测了北京市未来总用水量的变化趋势,通过检验证明此耦合模型为一种简单、有效的年用水量预测模型。

### 1 研究区概况

北京市的水资源量为近 40 亿  $m^3$ 。按联合国规定的人均

水资源丰水线 3 000  $m^3$ /人和警戒线 1 700  $m^3$ /人的标准,北京市最佳人口规模为 133 万人,最大人口规模为 235 万人<sup>[5]</sup>。而 2004 年北京市人口总量已达 1 380 万人,再加上工业和农业发展的用水需求,使得北京已成为世界上严重缺水的特大城市之一。

根据北京市 1980~2001 年统计年鉴资料,从中选取 8 个因子:总人口数(万人)、固定资产(亿元)、国内生产总值 GDP(亿元)、社会总产值(亿元)、人均日生活用水量、日供水能力(万吨)、工业用水量(亿  $m^3$ )和总用水量(亿  $m^3$ )作为研究对象。其中,1980~1999 年的资料用于数学建模,2000~2001 年的资料用于模型检验。

收稿日期:2006-11-12

作者简介:黄 胜(1973-),男,讲师,博士,主要从事水文水资源、环境监测和环境评价工作。

## 2 数学建模

### 2.1 灰色系统关联系数及关联度的求解

#### 2.1.1 确定时间序列

作为比较的母序列,同时以几个因素作为比较序列。现选定北京市总用水量作为参考序列,而影响总用水量的7个因子作为比较序列,选取1980~2001年的总用水量和各个因子的数值见表1。

关于灰色系统的关联度分析<sup>[6]</sup>,首先要确定参考序列,即

表1 总用水量和各个影响因子的数值

年份	总人口数 /万人	GDP /亿元	日供水能 力/万 t	工业用水 量/亿 m <sup>3</sup>	人均日生活 用水量/亿 m <sup>3</sup>	固定资 产/亿元	社会总 产值/亿元	总用水 量/亿 m <sup>3</sup>
1980	904.3	137.9	163.0	13.50	141.0	33.17	275.8	42.08
1981	919.1	151.6	167.0	13.56	153.0	36.62	285.2	42.67
1982	928.6	165.7	161.0	13.49	151.0	38.67	304.2	41.23
1983	954.1	183.5	165.9	13.36	151.0	51.35	344.6	42.19
1984	966.6	214.4	164.5	13.62	142.2	66.29	397.1	41.47
1985	989.1	257.1	169.7	12.84	147.4	93.95	482.2	42.36
1986	1 032.0	284.8	175.6	12.91	155.4	106.23	525.9	40.98
1987	1 067.0	326.8	181.3	13.28	160.4	136.24	625.9	42.76
1988	1 081.0	410.2	188.8	12.75	157.9	162.96	818.2	41.54
1989	1 086.0	455.9	205.7	12.50	155.2	139.49	972.1	41.33
1990	1 104.0	500.8	215.1	12.34	165.4	135.6	1 053.7	41.12
1991	1 116.0	558.7	220.6	12.48	178.2	144.4	1 201.4	41.38
1992	1 125.0	709.1	225.6	12.39	184.3	201.0	1 481.0	41.76
1993	1 140.0	863.5	234.3	12.54	198.9	318.2	2 543.2	42.09
1994	1 164.0	1 084.0	242.0	12.61	212.5	507.9	3 048.7	42.56
1995	1 171.0	1 394.9	263.6	12.43	219.2	841.5	4 073.0	42.89
1996	1 184.0	1 615.7	266.4	12.64	237.3	876.9	5 133.2	43.21
1997	1 217.0	1 810.1	301.8	11.05	229.0	961.2	5 839.2	40.26
1998	1 223.0	2 011.3	330.3	10.84	238.2	1 115.3	6 598.5	40.47
1999	1 250.0	2 174.5	356.6	10.56	250.1	1 170.6	7 088.8	41.71
2000	1 278.0	2 478.8	367.1	10.52	240.2	1 297.4	8 026.5	40.43
2001	1 367.0	2 845.7	370.7	10.47	238.7	1 530.5	9 608.8	41.01

注:数据来自北京市统计年鉴(1990~2001年)。

为消除表1数据由于量纲的不同而对分析结果产生不良影响,需对表1数据作初值化变换,即对总用水量和各个影响因子用1980年的数据分别除每年的数据,从而得到一个新的数据表。

#### 2.1.2 计算关联系数及关联度

关联分析作为一种系统分析技术,是分析系统中各因素的关联程度并对系统动态过程发展态势的量化作比较分析的方法。其实质是对数据进行几何关系的比较,找出比较序列的发展趋势与参考序列的发展趋势的吻合情况,并因此判断比较序列与参考序列的关联程度,从而得出影响参考序列的主要因素。利用表2(经过初值变换数据)进行关联度分析,得到总用水量和各个影响因子之间的关联系数,从而求出总用水量与各个影响因子之间的关联度(见表2)。

表2 总用水量与各个影响因子的关联度

影响因子	总人口数	GDP	日供水能力	工业用水量	人均日生活用水量	固定资产	社会总产值
关联度	0.99	0.82	0.98	0.99	0.99	0.73	0.78

从表2可以看出,总用水量与总人口数、GDP、日供水能力、工业用水量、人均日生活用水量的关联度比较大,而与固定产值、社会产值的关联度较小,故选取总人口数、GDP、日供水能力、工业用水量、人均日生活用水量5个因子作为自变量,总用水量作为因变量并用于建立逐步多元线性回归模型。

### 2.2 多元回归分析

多元线性回归分析<sup>[7]</sup>的基本原理是:设预测对象 $Y$ 与多个解释变量 $X_i (i=1,2,\dots,p; p>1)$ 之间存在线性关系,则其数学模型为: $Y = a + b_1X_1 + b_2X_2 + \dots + b_pX_p$ 。利用 $(i=1,2,\dots,n)$ 组观测值 $(X_{1i}, X_{2i}, \dots, X_{pi}, Y_i)$ ,根据最小二乘法原理求出上式中的待定系数 $a, b_1, b_2, \dots, b_p$ 。

通过对可能引起总用水量变化的各种影响因子进行多变量分析,进而建立了多元线性回归模型系统分析,以便确定总用水量的变化规律。通过灰色系统关联度分析得知,所选取的7个影响因子中对总用水量变化的相关性最为紧密的是人口、国内生产总值GDP、日供水能力、人均日生活用水量和工业用水量5个因子。为了避免对总用水量没有显著影响的因子进

入多元回归模型,故采用逐步回归的办法,利用 MATLAB 里的 stepwise( )函数进行计算<sup>[8]</sup>,最后得出北京市总用水量与其主要影响因子(人均日生活用水量和工业用水量)的多元线性回归模型为:

$$Y = 22.40 + 0.023X_1 + 1.21X_2$$

### 2.3 建立灰色动态模型 GM(1,1)

在预测技术中,选择预测对象的量化指标是重要的。邓聚龙依据 1982 年提出的灰色系统理论,使用累加生成概念作为灰色量“白化”的工具,用以构造趋势模型。其中 GM(1,1)模型就是取得显著成效的灰色动态模型,在经济与环境预测中得到广泛地应用<sup>[9]</sup>。

灰色数据行预测以一次累加生成的数据行(称为生成数据行)建立单行的一阶线性微分方程,即 GM(1,1)模型。根据各行的发展系数,看各行所代表的行业在未来的发展趋势以及根据其发展趋势应采取的相应措施。其相应的微分方程的一般形式如下:

$$\frac{d(x)^{(1)}}{dt} + \alpha x^{(1)} = u$$

其时间响应函数为:

$$x^{(1)}(t) = \left[ x_{(0)}^{(1)} - \frac{u}{\alpha} \right] \times e^{-\alpha t} + \frac{u}{\alpha}$$

式中: $a, u$ 为预测参数; $e^{-\alpha t}$ 为系统发展指数; $t$ 为时间值; $x^{(1)}$ 为生成数据行; $x_{(0)}^{(1)}$ 为生成数据行的第一个元素。

对表 1 中的人均日生活用水量和工业用水量(1980~1999 年)的数据,按照 GM(1,1)模型的要求,分别建立相应的微分方程,并求出各自的发展灰数  $a$  和内生控制灰数  $u$ (见表 3)。

表 3 发展系数和内生控制灰数

项目	人均日生活用水量	工业用水量
$a$	-0.034 6	0.010 6
$u$	124.666 0	13.978 1

### 3 应用实例

将表 3 中人均日生活用水量和工业用水量的发展系数和内生控制灰数代入 GM(1,1)模型,可得到人均日生活用水量和 2000 年、2001 年、2010 年和 2015 年工业用水量的预测值,再代入逐步回归方程,便得到北京市 2000 年、2001 年、2010 年和 2015 年总用水量的预测值,分别见表 4 和表 5。

表 4 2000 年和 2001 年预测值与实际值的比较

项目	2000 年			2001 年		
	人均日生活用水量/万 t	工业用水量/亿 m <sup>3</sup>	总用水量/亿 m <sup>3</sup>	人均日生活用水量/万 t	工业用水量/亿 m <sup>3</sup>	总用水量/亿 m <sup>3</sup>
预测值	245.72	11.37	41.81	254.37	11.25	41.86
实际值	240.20	10.52	40.43	238.70	10.47	41.01
相对误差/%	2.30	8.10	3.40	6.60	7.45	2.10

表 5 2010 年和 2015 年的预测值

项目	2000 年			2001 年		
	人均日生活用水量/万 t	工业用水量/亿 m <sup>3</sup>	总用水量/亿 m <sup>3</sup>	人均日生活用水量/万 t	工业用水量/亿 m <sup>3</sup>	总用水量/亿 m <sup>3</sup>
预测值	347.3	10.2	42.8	412.9	9.7	43.6

由表 3 可以看出,北京市随着经济的发展,人们生活水平不断提高,人均日生活用水量(其发展灰数  $a = -0.034 6 < 0$ )不断增长;同时,随着科技的发展,工业结构的调整和工业技术的改进,工业用水量(其发展灰数  $a = -0.010 6 > 0$ )呈现出一种减少的趋势,但人均日生活用水量增长的幅度大于工业用水量减少的幅度。

由表 4 可知:将 2000 年和 2001 年的预测值与实际值比较,具有较高的预测精度,预测结果能反映北京市总用水量的变化趋势。说明在进行总用水量的预测中,用多元逐步回归与灰色系统模型 GM(1,1)耦合,预测总用水量,是一种有效、适用和快速的方法。

此外,到 2010 年和 2015 年,北京市的总用水量分别将达到 42.8 亿 m<sup>3</sup> 和 43.6 亿 m<sup>3</sup>,要想缓解北京市水资源短缺的矛盾,必须解决好以下三方面的问题:①增加中水回用,降低人均日生活用水量;②对工业结构进行调整,对落后企业进行技术改造,提高工业用水利用率,大力发展低耗水,高科技的工业产业,在现有基础上,进一步降低工业用水量;③实行开源和节流并重的政策,利用南水北调工程,提高可供水资源量。

### 4 结 语

本文构造的灰色系统与多元逐步回归的耦合模型概念清晰、结构简单、适用方便,反映了年用水量与其主要影响因素之间的复杂关系,对年用水量的预测具有潜在的优势;通过以北京市为实例应用研究表明,此耦合模型预测效果较好(以北京市为例相对误差小于 4%);此耦合模型用于年用水量的预测还仅是一种新尝试,取得了初步成果,还需要对模型作进一步的验证和改进。

参考文献:

- [1] 苏文利. 城市用水量的神经网络预测[J]. 数量经济技术经济研究, 2003, (7): 131-133.
- [2] 田一梅, 汪泳. 偏最小二乘与灰色模型组合预测城市生活用水量 J, 天津大学学报, 2004, 37(4): 322-325.
- [3] 陈意平, 张张爱. 城市用水灰色动态模拟与预测[J]. 山西水利科技, 1994, 24(1): 18-22.
- [4] 潘敏, 范庆来. 灰色模型在城市用水量预测中的应用[J]. 浙江水利科技, 2004, (1): 21-22.
- [5] 崔凤垣. 从水资源状况看北京市的人口控制[J]. 人口学刊, 2003, 137(1): 23-27.
- [6] 邓聚龙. 灰色预测与决策[M]. 武汉: 华中理工大学出版社, 1986.
- [7] 肖庭延. 实用预测技术及应用[M]. 武汉: 华中理工大学出版社, 1993.
- [8] 苏金明. MATLAB6.1 实用指南(下册)[M]. 北京: 电子工业出版社, 2002.
- [9] 徐洪福. 灰色预测模型在年用水量预测中的应用[J]. 哈尔滨建筑工业大学学报, 2001, 34(4): 61-65.