

北京市生活用水年预报模型

顾月红^{1 2}, 葛朝霞^{1 2}, 薛 梅^{1 2}, 宋颖玲^{1 2}

(1. 河海大学水文水资源与水利工程科学国家重点实验室, 江苏 南京 210098 ;

2. 河海大学水文水资源学院, 江苏 南京 210098)

摘要 :采用偏最小二乘回归方法建立了用水量与影响因子之间的回归方程 ,利用 AR 模型对各影响因子进行外延预测 ,将预测的因子值分别代入 3 个回归模型 ,通过比较分析挑选出了预报精度最高的回归模型 ,并利用此回归模型对 2003 ~ 2010 年的北京市年生活用水量进行了预测 .结果表明 :北京市年生活用水量和前 3 a 的各项影响因子所组成的回归方程 ,能够很好地预报出预留年份的年生活用水量 ,2003 ~ 2010 年北京市年生活用水量呈明显增长趋势 .

关键词 :北京市 ;年生活用水量 ;影响因子 ;回归模型 ;AR 模型

中图分类号 :TU991.31 **文献标识码** :A **文章编号** :1000-198X(2008)01-0019-04

随着经济的发展、人民生活水平的逐步提高以及城市规模的不断扩大 ,城市的用水量呈现不断增长的趋势 ,城市水资源供需矛盾不断加剧 ,供水量紧缺的现象也越来越严重 .准确预测城市用水量 ,对城市供水规划和水务管理以及城市的可持续发展具有重要的作用 .

常用的水量预测方法有各类回归模型、时间序列分析模型、灰色模型、神经网络模型以及各类方法的组合模型等^[1-6] .其中 ,以分析预测对象和其影响因子之间回归关系的模型应用最为广泛 .本文采用偏最小二乘回归与时间序列分析相结合的方法 ,对 2003 ~ 2010 年北京城市年生活用水量进行了预测 .

1 预测方法

1.1 偏最小二乘回归

偏最小二乘回归^[7]是一种因变量对多自变量的回归建模方法 ,可以解决一般回归问题中自变量之间的多重相关性和样本数量太少的问题 ,它是多元线性回归分析、典型相关分析和主成分分析的集成和发展 .其主要思路是 :首先 ,从自变量集合 $X = \{X_1, X_2, \dots, X_m\}$ (m 为自变量的个数)中按交叉有效性原则提取主成分 t_h ($h = 1, 2, \dots, m$) ,各成分相互独立 ,然后 ,建立这些主成分与单因变量 y 之间的线性回归方程 .

在许多情况下 ,偏最小二乘回归并不需要选用 $t_1, t_2, \dots, t_m$ 的全部进行回归建模 ,而是采用截尾方式选择前 h 个成分 ,仅用这 h 个成分 $t_1, t_2, \dots, t_h$ 就可以得到一个预测性能较好的模型 .这种判断方式称为交叉有效性原则^[8] .

偏最小二乘回归所提取的主成分不仅能较好地概括原始自变量系统的信息 ,对因变量系统也具有很强的解释能力 ,而且主成分之间互不相关 ,可以有效地解决自变量之间多重相关情况下的回归建模问题 .

1.2 时间序列分析

时间序列分析的基本思想是 :先假定某一变量要素随时间变化时 ,任一时刻的变化和前期要素变化有关 ,再根据这种假定建立可以适当地描述这一要素变化规律的模型 ,最后利用所建立的模型对变量要素未来时刻的预测值做出估计^[9-10] .

对于某一非平稳时间序列 x_t ($t = 1, 2, \dots, n$, n 为样本数) ,首先需要判断此序列是否存在趋势、突变和周期等特性 ,若存在 ,则分别剔除趋势项、跳跃项和周期项等 .对于通过剔除得到的新序列 ,可以按照平稳序列的 AR(p)模型进行预测 .模型阶数 p 可以按照 AIC 准则确定 .

2 北京市年生活用水量的分析与预测

2.1 预报因子的选择

文献[11]将生活用水量定义为公共服务用水量、居民用水量和消防等特殊用水量之和,其中:公共服务用水是指为城市社会公共生活服务的用水,包括行政事业单位、部队营区和公共设施服务、社会服务业、批发零售贸易业、旅馆餐饮业和社会服务等单位的用水;居民用水是指城市范围内所有居民家庭的日常生活用水,包括城市居民、农民家庭和公共供水站用水;消防及其他特殊用水是指城市灭火以及除居民家庭、公共服务、生产运营用水范围以外的各种特殊用水,包括消防用水、深井回灌用水和其他用水。由此可见,影响生活用水量的因素有多种,所选自变量应该尽可能包含能有效影响生活用水系统的变量。通过物理考察,本文选择非农业城市人口、第三产业产值、人均居住面积、绿地因子(当年公共绿地面积与当年降水量之比^[6])以及城镇居民家庭可支配收入(即居民家庭可以用来自由支配的收入,它是家庭总收入扣除交纳的所得税、个人交纳的社会保障费以及调查户的记账补贴后的收入)为自变量。北京市年生活用水量与5项因子的数据来源于1985~2003年的《北京统计年鉴》。

分析北京市年生活用水量与5项因子间的相关系数可知,年生活用水量 y 和各项因子 X 以及各项因子之间多重相关显著。由于北京市年生活用水量及其影响因子的样本长度有限,采用偏最小回归方法建立模型。

2.2 模型的建立

年生活用水量不仅受到当年各项因子的影响,前期因子对后期用水量的影响也具有一定的滞后效应。考虑到实际预测当年的年生活用水量 y 时,当年各项的因子值还未知,所以将预测对象 y 和影响因子 X 序列分别错开1~3a,于是得到3个预报方案。

采用1986~2000年的北京市年生活用水量资料建模,利用2001年和2002年的北京市年生活用水量资料验证预报模型的精度。

首先,建立因子分别提前1a、2a和3a的偏最小二乘回归模型,并分别对2001年和2002年的北京市年生活用水量进行预测,同时通过预测结果的比较,挑选出预报精度最高的回归模型;其次,用AR模型预测未来年份各项因子的值,并将预测出的因子值代入已选定的回归模型,利用该回归模型预报出2003~2010年的北京市年生活用水量。

2.2.1 回归模型的选择

将 y 和 X 序列分别错开1~3a,利用偏最小二乘回归模型进行计算,并认为预报对象相对误差小于20%的为合格。拟合通过率定义为:合格点数与总点数(样本长度)之比。计算结果表明,3个方案均通过 $\alpha=0.05$ 的回归效果显著性检验,拟合通过率都为100%。但对于方案3,也就是在时间错开3a的情况下,2001年和2002年北京市年生活用水量预报的相对误差最小,并且复相关系数最大。据此,选择方案3预测北京市2003~2010年的年生活用水量。

方案3偏最小二乘回归方程为

$$\hat{F}_0 = -0.46t_1 + 0.33t_2 - 0.49t_3 \quad (1)$$

式中 t_1 、 t_2 、 t_3 是根据交叉有效性原则筛选出来的3个主成分,可分别表示为

$$\begin{aligned} t_1 &= -0.47X_1^* - 0.47X_2^* - 0.5X_3^* - 0.3X_4^* - 0.47X_5^* \\ t_2 &= 0.09X_1^* - 0.01X_2^* + 0.58X_3^* - 0.81X_4^* - 0.02X_5^* \\ t_3 &= 0.26X_1^* + 0.31X_2^* - 0.93X_3^* - 0.09X_4^* - 0.41X_5^* \end{aligned}$$

代入式(1)得

$$\hat{y}^* = 0.116X_1^* + 0.062X_2^* + 0.872X_3^* - 0.086X_4^* + 0.009X_5^* \quad (2)$$

式中 $\hat{y}^*$ ——预报对象的标准化的数据; X_i^* ($i=1, \dots, 5$)——预报因子的标准化的数据。

由方程(2)可以看出, X_3^* 对应的系数达到0.872,相应的因子 X_3^* 的影响最大,并且呈正相关。这说明人均居住面积越大,生活水平越高,北京市年生活用水量就越大。

将数据还原,可以得到原变量的偏最小二乘回归方程

$$\hat{y} = -22\,969.69 + 18.2X_1 + 2.36X_2 + 9\,887.16X_3 - 466.9X_4 + 0.04X_5 \tag{3}$$
式中： $\hat{y}$ ——北京市年生活用水量； X_i （ $i = 1, \dots, 5$ ）——北京市非农业城市人口、第三产业产值、人均居住面积、绿地因子和城镇居民家庭可支配收入数据。

2.2.2 各项因子的预测

由于选定的回归方程中的 y 和 X 错开了 3 a,也就是说,若要预测 2003 ~ 2010 年的北京市年生活用水量 y ,要将 2000 ~ 2007 年的各项因子 X 代入方程才可得到预测值.而 2000 ~ 2002 年的各项因子的值是已知的,必须预测出 2003 ~ 2007 年各项因子的值.为此,利用 AR(p)模型对各项因子的值进行预测.

预测前,先对各因子序列进行了如下检验:通过 Kendall 秩次检验^[12]判断因子序列是否存在趋势项;利用简单分波法^[13]分析因子序列周期项是否存在;利用有序聚类法^[14]检验因子序列是否含有跳跃项.结果表明,显著性水平 α 取 0.05 时,各因子序列均含有趋势项,但不存在跳跃项和周期项.利用最小二乘法估计出各因子序列趋势项的直线方程(表 1),并从原序列中剔除趋势项,然后利用 AR(p)模型对剔除了趋势项的新序列进行预测.模型阶数 p (表 1)按 AIC 准则确定.将 AR(p)模型的预测结果分别加上各对应的趋势项后,得到各项因子未来年的预测值(表 2).

表 1 5 项因子的趋势项直线方程系数和 AR 模型的阶数

Table 1 Linear equation for trend items of 5 factors and orders of AR model			
因子	趋势项直线方程系数		按 AIC 准则确定的模型阶数 p
	b_0	b_1	
X_1	563.05	23.14	1
X_2	-310.97	114.92	1
X_3	6.02	0.33	2
X_4	1.52	1.05	2
X_5	-1442.22	762.31	1

表 2 5 项因子最终预测结果

Table 2 Final predicted results of 5 different factors					
年 份	X_1	X_2	X_3	X_4	X_5
2003 年	1002.03	1996.22	12.13	22.11	12967.69
2004 年	1014.31	2032.71	12.32	22.39	13542.05
2005 年	1031.86	2094.98	12.55	23.22	14167.73
2006 年	1052.13	2174.56	12.83	23.95	14830.71
2007 年	1073.80	2265.76	13.16	24.88	15520.80

由表 2 可见,预测得到的各项因子都呈现逐年增长的趋势,它们的平均增长率分别为 1.7%、3.2%、1.1%、3.0%、4.6%.

2.2.3 年生活用水量预测结果

将 2000 ~ 2007 年的 5 项因子值代入已确定的回归方程(3)后,就可预测出 2003 ~ 2010 年的北京市年生活用水量,结果如表 3 所示.

表 3 2003 ~ 2010 年北京市年生活用水量预测结果

Table 3 Predicted result of annual domestic water consumption for Beijing City from 2003 to 2010								
年 份	2003 年	2004 年	2005 年	2006 年	2007 年	2008 年	2009 年	2010 年
年生活用水量预测结果/万 m^3	99207.6	103772.6	108480.8	110156.5	112239.2	114619.5	117633.0	121101.5

2.3 结果分析

从以上计算结果可知,选入方程的 5 项因子(北京市非农业城市人口、第三产业产值、人均居住面积、绿地因子以及城镇居民家庭可支配收入)2003 ~ 2007 年呈现明显的增长趋势,并且 2003 ~ 2010 年北京市年生活用水量也相应地呈现增长趋势(平均增长率达到了 2.9%).这说明,随着社会的发展、城市的逐渐扩张和人口的不断增加,人民的生活水平也逐渐提高,整个城市对水量的需求也会越来越高.

3 结 论

- a. 本文利用偏最小二乘回归方法从自变量集合 X 和因变量 y 中提取互不相关的主成分 t_h ,并对这些主成分进行多元线性回归分析,解决了原自变量多重相关和样本数量太少的问题.
- b. 由于各项因子呈现明显的增长趋势,在利用 AR(p)模型对各项因子进行预测前先剔除其趋势项,就可以得到更加合理的预测结果.
- c. 选择的 5 项因子都与北京市年生活用水量有很大关系(尤其是人均居住面积,其影响最大),将它们作为因子能很好地反映城市年生活用水量的变化趋势.
- d. 本文预测结果较为合理和可靠,可为北京城市的供水规划等提供参考依据.

参考文献:

- [1] 张智,陈金锥,魏忠庆,等.城市年用水量的分类预测探析[J].水资源与水工程学报,2006,17(4):10-13.
- [2] 李栋,王洪礼,杜忠晓,等.城市生活用水量的支持向量回归预测[J].天津大学学报:社会科学版,2006,8(1):64-67.
- [3] 李琳,左其亭.城市用水量预测方法及应用比较研究[J].水资源与水工程学报,2005,16(3):6-10.
- [4] 田一海,汪泳,迟海燕.偏最小二乘与灰色模型组合预测城市生活需水量[J].天津大学学报,2004,37(4):322-325.
- [5] 曹连海,王安明,陈南祥,等.偏最小二乘回归神经网络在城市生活用水量中的应用[J].华北水利水电学院学报,2005,26(2):9-11.
- [6] 张雅君,刘全胜,冯萃敏.多元线性回归分析在北京城市生活需水量预测中的应用[J].给水排水,2003,29(4):26-29.
- [7] 王惠文.偏最小二乘回归方法及其应用[M].北京:国防工业出版社,1999:1-20.
- [8] 王文圣,丁晶,赵玉龙,等.基于偏最小二乘回归的年用电量预测研究[J].中国电机工程学报,2003,23(10):17-20.
- [9] 黄嘉佑.气象统计分析与预报方法[M].北京:气象出版社,2004:195-199.
- [10] 李纪人,刘德平.水文时间序列模型及预报方法[M].南京:河海大学出版社,1991.
- [11] 北京市统计局.北京统计年鉴[M].北京:中国统计出版社,2003:137.
- [12] 李春晖,杨志峰.黄河流域分区天然径流量趋势性与持续性特征[J].水文,2005,25(1):13-17.
- [13] 张闻胜,王银堂.时间序列分解模型在水文要素中长期预报中的应用[J].水文,2001,21(1):21-24.
- [14] 丁晶,邓育仁.随机水文学[M].成都:成都科技大学出版社,1988:64-121.

Prediction model for annual domestic water consumption in Beijing City

GU Yue-hong^{1,2}, GE Zhao-xia^{1,2}, XUE Mei^{1,2}, SONG Ying-ling^{1,2}

(1. State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering,
Hohai University, Nanjing 210098, China;

2. College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China)

Abstract :The regression equation for the relationship between water consumption and influencing factors was developed by use of the partial least square regression method. The AR model was used for prediction of each influencing factor by extrapolation, and the predicted values were introduced to 3 regression models. Then, the regression model with the highest precision of prediction was selected by comparative analysis, and it was applied to prediction of annual domestic water consumption of Beijing City from 2003 to 2010. The result shows that the regression equation based on annual water consumptions and the influencing factors in the past three years can effectively predict the domestic water consumption of Beijing City, which demonstrates an obviously increase trend from 2003 to 2010.

Key words :Beijing City; annual domestic water consumption; influencing factor; regression model; AR model