

DOI :10.3876/j.issn.1000-1980.2011.01.001

非线性逐步回归在宜昌站年径流量预报中的应用

胡己坤^{1,2},曹丽青^{1,2},葛朝霞^{1,2},郑苏娟³,黄美花^{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;3.河海大学理学院,江苏 南京 210098)

摘要:为了分析非线性关系的引入对中长期水文预报精度的影响,以对数变换后得到的非线性逐步回归模型为基础,利用长江宜昌站年径流量进行中长期水文预报,并与传统的逐步回归模型预报值进行比较.结果表明,宜昌站年径流量与该站上一年汛期 7 月、8 月、10 月流量存在明显非线性关系,并对太平洋副高活动以及厄尔尼诺效应反应敏感.非线性逐步回归模型拟合和试预报准确度均高于线性模型.引入非线性关系能在一定程度上提高模型的预报精度,具有实用价值.

关键词:非线性;逐步回归;中长期水文预报;宜昌站

中图分类号:P338+.2 文献标志码:A 文章编号:1000-1980(2011)01-0001-04

中长期水文预报是根据前期水文要素,用成因分析与数理统计的方法,对未来较长时间的水文要素进行科学的预测.逐步回归方法作为统计方法的重要组成部分,具有物理意义清晰、因子主次明确、独立性好等优点^[1],但该方法对资料的依赖性较高.很多学者从不同角度对逐步回归的改良进行了有意义的探索^[2-8].本文在逐步回归相关分析的步骤中添加对数变换,考虑因子和预报对象之间存在的幂函数关系以及对数关系,建立非线性预报模型,以期提高预报精度.

1 对数变换应用原理

设 y 对 x 存在幂函数关系:

$$y = ax^b \tag{1}$$

对式(1)取对数,有:

$$\ln y = \ln a + b \ln x \tag{2}$$

作变换,令 $z = \ln y, b_0 = \ln a, m = \ln x$,式(2)化为

$$z = b_0 + bm \tag{3}$$

这就是一个一元线性回归问题^[1].

利用最小二乘法得到系数 b 的大小 b_d .现用 y_t 和 $x_t (t = 1, 2, \dots, n)$ 表示预报对象和预报因子,令 $e =$

$$\sum_{t=1}^n (y_t - ax_t^b)^2, \text{ 则 } \hat{e} = e_{\min} \Leftrightarrow b = b_d \text{ 令 } r = \frac{\sum_{t=1}^n (x_t^b - \bar{x}^b)(y_t - \bar{y})}{\sqrt{\sum_{t=1}^n (x_t^b - \bar{x}^b)^2 \sum_{t=1}^n (y_t - \bar{y})^2}}, \text{ 则 } \hat{r} = r_{\max} \Leftrightarrow b = b_d.$$

若 y 与 x 存在对数关系,处理方法同幂函数关系类似,这里不再赘述.

为从指数、对数和线性 3 种关系当中挑选出最适合的线型代入逐步回归计算,分别用 f_1, f_2, f_3 依次表示上述 3 种函数型, r_1, r_2, r_3 表示其与预报对象 y_t 之间的单相关系数,选择 $y = f_i$,使得 $1 - |r_i|$ 最小 ($i = 1, 2, 3$),代入逐步回归模型进行计算.

2 模型建立及分析

本文选用的资料为 1951—2005 年长江宜昌站逐月平均径流量资料、1951—2000 年 74 项气象环流因子的

月资料以及 1951—2005 年厄尔尼诺 4 个关键海区月海温平均值资料。

2.1 预报模型建立

采用线性逐步回归模型(以下简称模型 1)和增加对数变换的逐步回归模型(以下简称模型 2)建立宜昌站年径流量的预报模型。模型建立步骤如下:

步骤一 利用 1951—2002 年资料系列,选择置信度 $\alpha = 0.01$,模型 1 计算宜昌站年径流量与各预报因子(气象环流因子、海温)提前 1~5 年的相关系数,通过检验的因子进入步骤二。模型 2 首先选择最优线型后(即在线性、指数型和对数型三者中挑选)计算相关系数,通过信度检验的因子进入步骤二。

步骤二 将两方案初选的因子代入逐步回归程序,建立年径流量预报方程,并对方程进行 F 检验(取置信度 $\alpha = 0.01$)。

步骤三 利用 2003 年、2004 年和 2005 年资料做试预报检验。

定义相对误差为拟合值、实测值之差与实测值的比值,以相对误差不超过 20% 为预报合格。

下面应用模型 1 和模型 2 分别建立宜昌站年径流量预报方程。模型 1 采用 $\alpha = 0.05$ 的置信度建立预报方程,模型 2 取置信度 $\alpha = 0.01$ 。

模型 1 得到的宜昌站年径流量预报方程为

$$y = 12496.57 - 416.07x_1 + 70.33x_2 - 0.13x_3 + 93.03x_4 + 238.37x_5 + 10.11x_6 + 224.67x_7 \quad (4)$$

式中: x_1 ——北非副高强度指数提前 2 年的 7 月测量值; x_2 ——南海副高脊线提前 2 年的 7 月测量值; x_3 ——编号台风提前 3 年的 5 月测量值; x_4 ——北非副高脊线提前 4 年的 3 月测量值; x_5 ——东太平洋副高面积指数($175^\circ\text{W} \sim 115^\circ\text{W}$)提前 5 年的 5 月测量值; x_6 ——印度副高脊线提前 5 年的 1 月测量值; x_7 ——宜昌站提前 5 年的 4 月径流量。

模型 2 得到的宜昌站年径流量预报方程为

$$y = -32399.19 + 2034.10\ln x_1 + 114.20x_2 + 2059.42\ln x_3 + 0.01x_4^{1.23} + 18.34x_5 + 19.25x_6 \quad (5)$$

式中: x_1 ——宜昌站上一年的 7 月径流量; x_2 ——南海副高北界提前 3 年的 9 月测量值; x_3 ——宜昌站上一年的 8 月径流量; x_4 ——宜昌站上一年的 10 月径流量; x_5 ——东太平洋副高强度指数提前 5 年的 9 月测量值; x_6 ——南方涛动指数上一年的 5 月测量值。

2.2 计算结果分析

分析模型 1 得到的预报方程(式(4)),1 号和 4 号因子与北非副高有关,且相关性一正一负,反映出其对宜昌站上游不同地区有不同的影响。2 号、5 号和 6 号因子均与太平洋副高活动有关,呈正相关关系。7 号因子为宜昌站前期流量,也为正相关关系。模型 1 拟合方程复相关系数为 0.8884,通过了置信度为 0.01 的 F 检验。模型 1 的拟合成果见图 1。

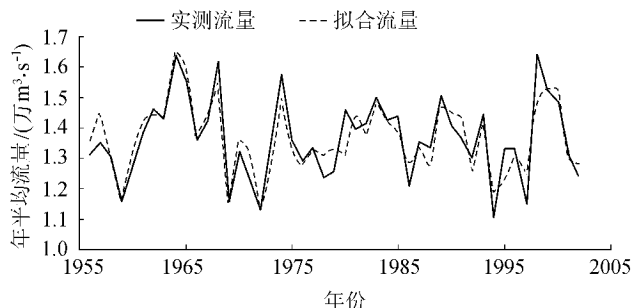


图 1 模型 1 拟合成果

Fig. 1 Actual and fitting curves of mean annual discharge of Model 1

分析模型 2 得到的预报方程(式(5)),所选中的 6 个因子中,2 号和 5 号因子与副高活动有关,同样为正相关关系。1 号、3 号和 4 号因子为上一年汛期 7 月、8 月、10 月径流量,且均为正相关关系,6 号因子为南方涛动指数,且与宜昌站年径流量呈正相关关系,即厄尔尼诺的发生使长江上游发生旱灾几率增大,而拉尼娜事件则使长江上游发生涝灾的几率增大。模型 2 拟合方程复相关系数为 0.9556,通过了置信度为 0.01 的 F 检验。模型 2 的拟合成果见图 2。

一些学者^[9-12]从不同角度出发,得出的结论与模型 2 是一致的。

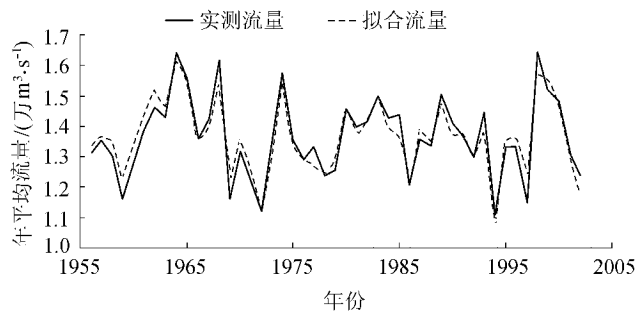


图 2 模型 2 拟合成果

Fig. 2 Actual and fitting curves of mean annual discharge of Model 2

综合比较来看,模型 2 得到的拟合序列复相关系数较大,精度较高,拟合检验效果为优。
表 1 给出了模型 1 和模型 2 对 2003—2005 年宜昌站年径流量试预报的结果及其误差,可以看出模型 2 的预报效果明显好于模型 1。

表 1 宜昌站年径流量试预报结果

Table 1 Predicted results of mean annual discharge at Yichang station

年份	实测年平均流量/ ($\text{m}^3 \cdot \text{s}^{-1}$)	模型 1		模型 2	
		拟合年平均流量/($\text{m}^3 \cdot \text{s}^{-1}$)	相对误差/%	拟合年平均流量/($\text{m}^3 \cdot \text{s}^{-1}$)	相对误差/%
2003	12 933	15 875	22.75	12 672	- 2.02
2004	13 081	14 694	12.33	12 412	- 5.12
2005	14 397	14 723	2.26	14 381	- 0.11

3 结 论

- a. 以对数变换为基础建立的逐步回归模型,能够考虑预报对象和预报因子之间的非线性关系,更适合于高度非线性化的水文气象过程。
- b. 以径流量作为预报因子,实际上就是以产生该流量的大气环流背景集合作为预报因子,通过逐步回归的筛选,有利于将那些物理上相对独立但在降水效果上配合作用很高的环流因子剔除出方程,而那些可以产生特殊天气过程的因子则得以突出,如模型 2 的南方涛动指数.通过模型 1 和模型 2 的比较可以看出,前期径流量与预报对象之间的非线性关系要强于线性关系。
- c. 以对数变换为基础的多因子逐步回归方法对长江宜昌站年径流量的拟合和试预报合格率都达到了 100%,效果明显优于线性模型,具有较高的实用价值。

参考文献：

[1] 范忠秀. 中长期水文预报[M]. 南京: 河海大学出版社,1999 :75-76.

[2] 葛朝霞,薛梅,宋颖玲. 多因子逐步回归周期分析在中长期水文预报中的应用[J]. 河海大学学报:自然科学版,2009,37(3) :255-257. (GE Zhao-xia,XUE Mei,SONG Ying-ling. Application of multi-factor stepwise regression cycle analysis in medium and long-term hydrological forecast[J]. Journal of Hohai University :Natural Sciences,2009,37(3) :255-257. (in Chinese))

[3] 段红,訾丽. 应用回归聚类预报物理模型预报三峡水库月平均来水量的实验与探讨[J]. 水资源研究,2007,28(3) :30-32. (DUAN Hong,ZI Li. Application of regression physical clustering forecast model in average monthly inflow of Three Gorges Reservoir[J]. Water Resource Research,2007,28(3) :30-32. (in Chinese))

[4] 张瑞好. 逐步回归曲线拟合模型及其在气象要素预报中的应用[J]. 成都气象学院学报,1991,16(1) :43-45. (ZHANG Rui-hao. A model for stepwise regression curve fitting and its application in meteorological essential factors forecasting [J]. Journal of Chengdu Institute of Meteorology,1991,16(1) :43-45. (in Chinese))

[5] 何丽娟,白玉良,赫明天. 双重逐步回归分析在中长期水文预报中的应用[J]. 东北水利水电,2006(2) :34. (HE Yu-juan,BAI Yu-liang,HE Ming-tian. Application of dual progressive regression analysis in hydrological medium-long range forecast[J]. Water Resource & Hydropower of Northeast China,2006(2) :34. (in Chinese))

[6] 王德隽,夏萍. 多元样条逐步回归模型在夏收中期预报中的应用[J]. 气象科学,2000,20(2) :224-228. (Application of multiple

- spline progressive regression model to medium-range prediction of summer harvest[J]. *Scientia Meteorologica Sinica*, 2000, 20(2): 224-228. (in Chinese))
- [7] 崔景山, 叶柳. 基于逐步回归均生函数模型的哈尔滨站年径流量预测分析[J]. *黑龙江水利科技*, 2009, 37(4): 7-9. (CUI Jing-shan, YE Liu. Function based on regression models of mean annual runoff of Harbin Station Forecast[J]. *Heilongjiang Science and Technology of Water Conservancy*, 2009, 37(4): 7-9. (in Chinese))
- [8] 孙树青. 长江三峡中长期水文预报研究[D]. 长沙: 长沙理工大学, 2007.
- [9] 假拉, 周顺武, 丁锋. 北非高压的气候特征及其对中国夏季降水的影响[J]. *南京气象学院学报*, 2002, 25(6): 816-822. (JIA La, HOU Shun-wu, DING Feng. The climatic character of North Africa subtropical high and its effect on summer precipitation in China[J]. *Journal of Nanjing Institute of Meteorology*, 2002, 25(6): 816-822. (in Chinese))
- [10] 董婕, 刘丽敏. 赤道东太平洋海温与中国温度、降水的关系[J]. *气象*, 2000, 26(2): 25-28. (DONG Jie, LIU Li-min. The relationship of temperature and precipitation in China and the ocean temperature in the equatorial East Pacific[J]. *Meteorological Monthly*, 2000, 26(2): 25-28. (in Chinese))
- [11] 张存杰, 高学杰, 赵红岩. 全球气候变暖对西北地区秋季降水的影响[J]. *冰川冻土*, 2003, 25(2): 157-164. (ZHANG Cun-jie, GAO Xue-jie, ZHAO Hong-yan. Impact of global warming on autumn precipitation in Northwest China[J]. *Journal of Glaciology and Geocryology*, 2003, 25(2): 157-164. (in Chinese))
- [12] 张强, 姜彤, 吴宜进. ENSO 事件对长江上游 1470—2003 年旱涝灾害影响分析[J]. *冰川冻土*, 2004, 26(6): 691-696. (ZHANG Qiang, JIANG Tong, WU Yi-jin. Impact of ENSO events on flood/drought disasters of Upper Yangtze River during 1470—2003[J]. *Journal of Glaciology and Geocryology*, 2004, 26(6): 691-696. (in Chinese))

Application of nonlinear stepwise regression in prediction of annual discharge at Yichang station

HU Ji-kun^{1,2}, CAO Li-qing^{1,2}, GE Zhao-xia^{1,2}, ZHENG Su-juan³, HUANG Mei-hua^{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* ;

3. *College of Sciences, Hohai University, Nanjing 210098, China*)

Abstract : In order to analyze the influences of introducing nonlinear relationship on the precision of middle-and long-term hydrological prediction, the mean annual discharge at Yichang station on Changjiang River was employed for the hydrological prediction based on the nonlinear stepwise regression model through log-transformation. The predicted values were compared with those by the traditional stepwise regression models. The results show that the annual discharge at Yichang station has a remarkable nonlinear relationship with its discharge in July, August and October in pervious year, and it is sensitive to the activity of Pacific Subtropical High and El Nino Effect. The precision of the fitting values of the nonlinear stepwise regression model and the predicted values is higher than that of the linear models. Introducing the nonlinear relationship may improve the model precision to a certain extent and is applicable.

Key words : nonlinearity ; stepwise regression ; middle-and long-term hydrological prediction ; Yichang station