

117-122

洪水预报中特征值预报的若干数学方法比较

李致家

(河海大学水资源水文系, 南京 210024)

孔祥光

(沂沭泗水利管理局, 徐州 221009)

p 338

提要 讨论研究了水文特征值预报的数学方法: 统计回归模型、神经网络模型和模糊回归模型。三个计算实例表明如果系统的线性关系较好, 统计回归模型的结果最好; 如果系统的线性关系差, 神经网络模型的结果最好; 如果用于率定模型的资料太短, 任何一个模型都不可靠。

关键词 洪水预报, 神经网络方法, 回归分析, 模糊回归

数学法 特征预报

水文特征值预报在洪水预报中使用得非常频繁, 对于一个防洪系统要进行蓄洪决策, 洪峰流量或最高水位的精确和及时预报非常重要。水文上常用的特征值预报方法有回归分析和相关图, 对于非线性的系统用分段线性或曲线来表示。数学上的一些方法可以用于水文的特征值预报, 如回归分析、模糊数学和神经网络模型等。数学模型在水文上的应用丰富了洪水预报的方法, 但是也应该看到不少研究只是站在推广数学模型的角度而单纯的把一些方法用于水文。赵人俊和 Klimes 指出这种研究方法的危害性^[1,2]。为此, 应当结合水文现象的规律来研究这些数学方法在什么条件下适用? 什么条件下不适用? 本文以此来比较几个数学方法。

1 3种常用的数学预报方法

1.1 模糊回归模型

统计回归分析方法具有如下的特点: 假定参数服从正态分布, 用最小二乘法来估计参数, 其估计的参数是不偏的。

模糊回归分析方法是, 假定参数是模糊数, 把参数经过一个截集转换为上、下限数, 用线性规划的方法求出上、下限, 也可以假定参数的隶属函数是三角型, 求出模糊的中心, 用于预报, 下面简要的介绍模糊回归模型^[3]。

设应变量 $Y(\omega)$ 为连续变量 ($\omega = 1, 2, \dots, n$), n 是样本大小, 自变量 $X(\omega)$ 有 m 个, 即:

$$X(\omega) = (x_1(\omega), x_2(\omega), \dots, x_m(\omega))$$

这里自变量是 1 或 2 型模糊集, 描述 $Y(\omega)$ 和 $X(\omega)$ 间关系的模糊线性模型为:

$$Y(\omega) = \sum A_i x_i(\omega)$$

其中, $Y(\omega)$ 是对目标变量的模糊估计, A 是待定的模糊系数, 是普通模糊。以下假定模糊数都是正规和凸的, $Y(\omega)$ 的模糊度定义为:

$$S = \sum (\bar{a}_i - \underline{a}_i)$$

式中, \bar{a}_i 和 a_i 为系数 $A_i (i=1, 2, \dots, m)$ 的截集 A_i 的端点值, 即: $A_i = [a_i, \bar{a}_i]$.

经推导可得到当 X_i 取为 1 型模糊集时, 求解模糊系数 A_i 的 h^* 截集就变为解如下线性规划问题在约束条件

$$\begin{aligned} a_i x_i(1) &\leq y(1) \cdots \cdots a_i x_i(n) \leq y(n) \\ \bar{a}_i x_i(1) &\geq y(1) \cdots \cdots \bar{a}_i x_i(n) \geq y(n) \\ \bar{a}_1 &\geq a_1 \cdots \cdots \bar{a}_m \geq a_m \end{aligned}$$

下使目标函数(2)最小.

假设上述讨论的模糊数的隶属函数均呈三角型, 即:

$$u_{\Delta}(u) = \begin{cases} 1 - |u - a| / c & (a - c) \leq u \leq (a + c) \\ 0 & \text{其它} \end{cases} \quad (3)$$

这里 a 为模糊数的中心, c 反映模糊数的模糊程度, 可以取 $a - c = a_i$, $a + c = \bar{a}_i$.

1.2 神经网络的 B-P 模型

神经网络是由一个简单的元件及其层次组织的大规模并行联接构造的网络, 它致力于按照生物神经系统的同样方式处理真实世界的客观事物^[4]. 神经元是神经网络中的基本处理单元, 它的信息处理过程包括三个部分. 神经网络模型由网络的拓扑结构、神经元特性和学习规则这三个因素所决定. 神经网络按拓扑结构, 可以分为分层模型和相互联接网络模型. 为了比较起见, 本文采用常用的带隐含层 (Hidden layer) 网络的反向误差传播法 (Back-Propagation Algrith, 简称 B-P 算法). 神经网络模型最大的特点是较强的非线性处理能力.

B-P 模型是一种多层感知器结构, 除输入层和输出层以外, 还可含有多个中间蔽层, 隐蔽层第 j 个神经元的输入为:

$$\text{net}_j = \sum W_{ij} O_i \quad (4)$$

第 j 个神经元的输出为:

$$O_j = f(\text{net}_j) \quad (5)$$

其中 f 是激活函数:

$$f(\text{net}) = \frac{1}{1 + \exp[-(\text{net}_j + \theta_j) / \theta_0]} \quad (6)$$

输出层第 k 个神经元的输入为:

$$\text{net}_k = \sum W_{kj} O_j \quad (7)$$

对应的输出为:

$$O_k = f(\text{net}_k) \quad (8)$$

神经网络的自学习是一个反复迭代过程. 首先, 随机给定一组权重和阈值, 然后输入一个样本并计算其输出, 通过实际输出与实测之间的差值用一定的方法修改网络的权, 来达到减小这个差值的目的. 这个算法就是误差最小平方法 (LMS), 它使用梯度搜索技术来达到误差最小. 对每个样本其误差平方为:

$$E_p = \frac{1}{2} \sum (t_{pk} - O_{pk})^2 \quad (9)$$

对所有样本

$$E = \frac{1}{2} \sum \sum (t_{pk} - O_{pk})^2 \quad (10)$$

B—P 模型的自学习过程是:(1)随机给定网络的权重和阈值一组小的随机值;(2)将数据标准化(0—1 化),确定输入和输出;(3)前向传递,按(1)式—(5)式计算隐蔽层和输出层各神经元的输出;(4)按(6)式—(7)式计算期望值与实测值的偏差,看一下是否小于误差阈值,如果满足则结束,反之则继续;(5)从输出层开始,反向调整权重,计算公式如下:

$$\Delta W_{jn}(n+1) = \eta(\delta_j O_j) + \alpha \Delta W_{jn}(n) \quad (11)$$

$$\text{其中} \quad \delta_{pk} = (t_{pk} - O_{pk}) O_{pk} (1 - O_{pk}) \quad (12)$$

$$\delta_{pj} = O_{pj} (1 - O_{pj}) \sum \delta_{pk} W_{kj} \quad (13)$$

式中, η 和 α 是学习效率和动量系数,是介于[0, 1]之间的常数.

在 B—P 模型中,隐蔽层个数、学习效率和动量系数都是人为确定.为了避免陷入局部最小,可以从几组不同的隐蔽层个数、学习效率和动量系数开始计算.

2 水文特征值的模拟和预报

比较各种模型的优劣有两种办法:第一是理论分析,第二是通过用具有一定代表性的资料进行计算对比.从理论上来看,无论是统计回归分析,还是模糊回归模型,都是处理的线性关系,而神经网络模型即可以处理线性关系,也可以处理非线性关系,那么应当不管系统是线性或非线性,神经网络模型都应当好,实际如何呢?下面选取三组有代表性的资料通过计算来进行比较研究.

取沂河流域实测的水文资料.资料 1 是用沂河临沂水文站上游水文(位)站的洪峰流量和最高水位来预报临沂站的洪峰流量;资料 2 是用流域上的降雨来预报临沂站的洪峰流量;资料 3 是新沂河用沭阳站的流量和水位来预报水位站盐河南闸的水位.资料 1 和资料 2 是每次洪水摘录一套相关数据(峰值),故属于短系列.资料 3 是从每次洪水摘录相应的水位、流量值,属于长系列.下面分别介绍三套资料的计算结果.

2.1 资料 1

沂河临沂水文站以上有干流葛沟水文站、支流姜庄湖水位站,区间面积有 15%.用葛沟的洪峰流量、姜庄湖最高水位和区间次洪径流深来预报临沂的洪峰流量.选取 1960—1991 年的 32 次洪水进行研究.有两种方案:第一是选取其中 24 次做为率定期,8 次洪水检验模型;第二是全部资料用来率定模型.确定出最优的神经网络模型结构是一个隐蔽层,该隐蔽层有 6 个神经单元.在计算时神经网络模型的输入和输出及模糊回归模型的输入(自变量)要先进行[0, 1]化.可用下面的公式计算:

$$x_i = \frac{(x_i(w) - x_i(l))}{(x_i(s) - x_i(l))} \quad (14)$$

其中, $X_i(s) > \max[X_i(w)]$, 且使 $\max[X_i(w)]$ 近似等于 0.90, $X_i(l) < \min[X_i(w)]$, 且使 $\min[X_i(w)]$ 近似等于 0.10.

模糊回归模型可以预报出一个区间,本文取最可能发生的值,该值是当隶属函数 $u(y) = 1$ 时的预报值,3 个方法的计算结果见表 1.

表 1 资料 1 各种预报方法的结果统计

Tab. 1 The results of 3 forecasting methods of Case 1

方 法	率 定 期		检 验 期		全部资料用来率定	
	均方误差	合格率	均方误差	合格率	均方误差	合格率
回归分析	435	91.6%	407	87.5%	426	90.6%
模糊回归	676	58.3%	805	75.0%	698	59.3%
神经网络	813	72.7%	1050	62.5%	1084	65.6%

2.2 资料 2

临沂是沂河洪水预报的关键,该水文站以上流域面积为 5194km²(不包括水库).从历史资料来看,降雨强度对洪峰影响较大,这是由于有部分超渗产流.经过分析取 5 个因子:次洪产流量、起涨流量、2h、4h、和 6h 雨量预报临沂的洪峰流量,选取 1960 年至 1993 年 16 次洪水进行研究.考虑两种方案:①是取 10 次洪水做为率定期,6 次检验模型;②是全部资料用来率定模型.最优神经网络模型的结构是一个隐蔽层,该隐蔽层有 5 个神经元.三种方法的计算结果见表 2.

表 2 资料 2 各种预报的结果统计

Tab. 2 The results of 3 forecasting method of Case 2

方 法	率 定 期		检 验 期		全部资料用来率定	
	均方误差	合格率	均方误差	合格率	均方误差	合格率
回归分析	708	60%	1126	33.3%	915	75%
模糊回归	无最优解				1313	50%
神经网络	91	100%	1105	50%	210	100%

2.3 资料 3

用盐河南闸上两个时段的水位、沭阳站上两个时段的水位和流量来预报盐河南闸本时段的水位.取 1960 年至 1991 年 11 次洪水共计 338 个资料.238 个用来率定模型,100 个用来检验模型.神经网络模型通过人工优化得出最优结构和参数是:一个隐蔽层,该隐蔽层有 6 个神经元,学习系数 $\eta = 0.30$,动量系数 $\alpha = 0.60$.计算结果见表 3,表中例出了误差限为 0.10 和 0.20m 的合格率.

表 3 资料 3 各种预报方法的结果统计

Tab. 3 The results of 3 forecasting methods of Case 3

方法	误差限	率 定 期		率 定 期	
		均方误差	合格率	均方误差	合格率
回归分析	0.10	2.02	70%	1.46	60%
	0.20	2.02	88%	1.46	84%
模糊回归	0.10	3.12	52%	4.50	59%
	0.20	3.12	75%	4.50	69%
神经网络	0.10	1.89	60%	11.64	6%
	0.20	1.89	92%	11.64	13%

3 计算结果的比较和分析

由表 1 可以看出对于资料 1 的情况,回归分析的结果无论率定期还是检验期,都比神经网络模型和模糊回归模型的好,而且前者的各项指标均优于后者.由这个资料无法比较出神经网络模型回归的好坏.表 2 的结果表明对于资料 2,率定期神经网络模型模拟的精度远远高于其它两个模型.对于检验期,模糊回归模型无最优解,其它两个模型的预报精度都不高,合格低于 50%.表 3 的结果表明对于资料 3,率定期若合格率以 0.10m 为限,回归分析的最高,神经网络的次之;若以 0.20m 为限,回归分析与神经网络的差不多,模糊回归的要差一些.对于检验期回归分析的要远远高于其它两个模型.

从这 3 个例子可以得出几点:(1)资料 1 的情况回归模型较好;(2)资料 2 的神经网络模型较好;(3)资料 3 的情况回归模型较好;(4)资料较短时模型的率定和检验效果都不好.分析原因如下:

表 4 是 3 套资料的各个自变量与应变量的偏相关系数.资料 1、3 的各个偏相关系数都较高,均大于 0.700.资料 2 各个因子的偏相关系数除次洪径流深之外,其它因子的都较小,可以归纳为资料 1、3 水文系统的线性关系较好,资料 2 水文系统的非线性较强.

对于线性关系较好的系统,回归模型模拟和预报的结果较好;对于非线性较强的系统,神经网络模型较好.

率定模型包括确定模型的结构和估计模型的参数.概念性模型的结构可由物理意义确定,而数学模型却由输入输出资料确定.为了较可靠的率定模型要求率定期的资料:(1)足够的长;(2)包括各种水文情况.Jakemen^[5]分析了流域模型,认为考虑了蒸发和土壤含水量的季节性变化,2—3 个串并联的线性水库可以足够精度的模拟产汇流,而且现有的水文资料也只能较为可靠的率定 2—3 个线性水库的参数,K. Beven 指出模型过多的参数导致了模型的不可靠性^[6],这是由于实际的水文资料不可能可靠地率定出这么多的参数.

对于本文的特征值预报及水文中一般的特征值预报问题,要利用次洪的洪峰流量或水位或次洪产流量.对于大多数的流域只有几十场的洪水资料.对于神经网络模型,如果输入因子是 3—5 个,1 个输出和 1 个隐蔽层,该隐蔽层有 3—10 个神经单元,那么这个模型至少有 16 个参数(包括权重和阈值);如果有 2 个隐蔽层则参数按指数增加.目前还没有看到文献说明,需要多少资料才能可靠的率定神经网络模型的参数.但暂且不谈神经网络模型的结构,就其参数率定来讲,是使用的最小误差平方准则和梯度收缩搜索法.对率定资料的要求应该与统计方法无异,从这点来看至少要有 30—40 次的洪水资料.故对于资料 2 只有 16 次洪水,虽然神经网络模型模拟的精度很高,但无法可靠地率定是出模型的参数,所以检验期的精度很差.

还有一个值得考虑的问题,对于线性关系较好的长系列资料 3,即使在率定期神经网络模型和回归模型的效果差不多,检验期回归模型的效果远高于神经网络模型的.这是一个在理论上无法解释的问题.

表 4 3 套资料的偏相关系数

Tab.4 The correlation coefficients

相关系数	(x_1, y)	(x_2, y)	(x_3, y)	(x_4, y)	(x_5, y)
资料 1	0.922	0.833	0.700		
资料 2	0.802	0.235	0.603	0.507	0.543
资料 3	0.960	0.840	0.830		

4 结论与建议

就上面讨论的三种情况而言,可以看出:

(1)对于水文特值的预报,当线性关系较好时,回归统计模型的精度较高,其参数较少,可以比较可靠的率定。

(2)对于水文特征值的预报,当线性关系较差时或者是非线性系统,神经网络模型的模拟精度较高;该资料系列较短时,该方法参数较多,需要有较长系列的资料,难以可靠的率定。

(3)即使对于长系列的线性关系较好的水文系统,回归分析的结果也不会比神经网络模型的差。

应当指出的是上面的结论是根据作者使用的3种资料,神经网络用B—P模型研究得出的,不是一般的结论,但可作进一步研究的参考。

参 考 文 献

- 1 赵人俊.从实际出发研究水文学.见:(文集编辑小组)编.赵人俊水文预报文集.北京:水利电力出版社,1994.239—243
- 2 Klemes V. A hydrological perspective. *J Hydrol*, 1988, **100**, 3—28
- 3 曹鸿兴.模糊集方法及其在气象中的应用.北京:气象出版社,1986.285—310
- 4 包约翰.自适应模式识别与神经网络.北京:科学出版社,1992.109—127
- 5 Jakemen A J. How Much complexity is warranted in a Rainfall-Runoff Model? *Water Resour Res*, 1993, **29**:2637—2648
- 6 Anderson M G and T P Burt. Hydrological forecasting. New York: John Wiley and Sons Ltd, 1985. 312—340

COMPARISON ON THREE MATHEMATICAL MODELS FOR SPECIAL VALUES IN FLOOD FORECASTING

Li Zhijia¹ Kong Xiangguang²

(1: Department of Water Resources & Hydrology, Hohai University, Nanjing 210024;

2: Yichang Civil Engineering Bureau, Ministry of Water Resources, Xuzhou 221009)

Abstract

Three mathematical models, i. e. regressive analysis method, artificial neural method and fuzzy regressive method, are commonly used in the flood forecasting for special values. The practical calculation results of three cases provided in this paper show that either one is suitable for all cases. The regressive analysis method is favorable when the system has better linear correlations; otherwise the artificial neural net method is better if the system is not linearly correlated. None of the above-mentioned method is reliable when the data needed for the calibration are not enough.

Key Words Flood forecasting, artificial neural method, regressive analysis, fuzzy regressive analysis.