

文章编号:1001-4810(2009)04-0375-06

喀斯特流域降雨-径流人工神经网络 模型结构分析及模拟^①

陈才,陈喜,张志才,魏琳娜

(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)

摘要:喀斯特流域降雨-径流响应是一个非线性过程,分析确定地下河流量过程的主要影响因子对喀斯特流域水文过程模拟具有重要意义。本文利用普定后寨河流域实测降雨、径流系列资料,采用神经网络权重分析法确定该流域的人工神经网络模型结构为两个隐含层、三个输入变量,该人工神经网络模型结构可以保持降雨-径流模拟的稳定性。模型经交叉训练与验证,训练期效率系数(NSC)达0.9以上,验证期NSC达0.88以上。说明神经网络权重分析法能够较好地确立预报因子与预报对象的关系,为喀斯特流域降雨-径流模拟提供一种有效的分析手段。

关键词:喀斯特;神经网络;权重分析;降雨-径流模拟;模型结构;普定后寨河流域;贵州

中图分类号:P333

文献标识码:A

0 引言

我国西南喀斯特流域具有地表—地下双重结构,属特有的二元三维流域空间系统^[1],降雨径流关系复杂,降雨经岩溶裂隙、漏斗、落水洞等流入地下,地下水相当丰富。但在西南喀斯特山区,通常缺乏地下水位观测资料,因而地下水数值模型难以模拟喀斯特地下暗河径流对降雨的响应。目前,通常采用分层结构的水箱模型^[2]以及输入—输出系统模型进行降雨—径流模拟^[3]。人工神经网络(ANN)模型是一种具有高度非线性特点的黑盒子模型,它运用统计方法建立输入输出的非线性映射,只需要简单的降雨、径流资料便能模拟喀斯特降雨—地下径流响应,适用于复杂喀斯特系统地下径流预测^[3]。

基于输入—输出关系的系统分析模型用于水文预测,预报精度及可靠性取决于所选择的预报因子,因此,选择预报因子是合理建模与预测的首要步骤^[4]。系统分析模型通常采用线性相关分析方法选取与预报对象显著相关的因子,但是相关系数大小受因

子中个别数据影响,容易造成预报因子漏选、错选^[5]。另外,在大量应用中也发现人工神经网络模型存在“过度训练”问题,导致模型通过增加输入因子个数、隐含层个数及训练次数提高训练期模拟精度,而预测精度降低的现象。

针对上述问题,本文将人工神经网络非线性分析方法(权重分析法)应用于水文预报因子选择,计算了输入、输出的权重系数,分析输入因子个数、隐含层个数对径流模拟精度的影响,提出在降雨—径流模拟中神经网络模型结构确定方法,成功地用于贵州普定后寨河流域降雨—地下暗河径流模拟。

1 流域概况

普定后寨河流域位于贵州省普定县城以南,面积80.65 km²,流域地势东南高、西北低,海拔高度在1 300 m左右,属于贵州高原分水岭的岩溶地貌水文特征典型区,地面较平缓。区内气候温和湿润,多年平均降雨量为1 380 mm,5~10月雨季降雨量占全年

① 基金项目:国家重点基础研究(973)资助项目(2006CB403200)、国家自然科学基金重点项目(40930635)

第一作者简介:陈才(1985—),男,硕士研究生,主要从事水文模型、模拟研究。E-mail:chencai_85@163.com。

收稿日期:2009-06-29

80 %以上。后寨河流域由三条地表河(灯盏河、号营河、后寨河)和一条地下暗河(打油寨—冒水坑)组成(如图1)。地表干支流常年处于干枯状态,只在汛期起排洪作用,雨季灯盏河与号营河汇合流入青山水库,

再经后寨河流出同地下河汇合,大部分雨洪径流经打油寨—冒水坑地下暗河流出^[6]。该流域地下水丰富,冒水坑地下暗河最大流量达 3.7 m³/s。

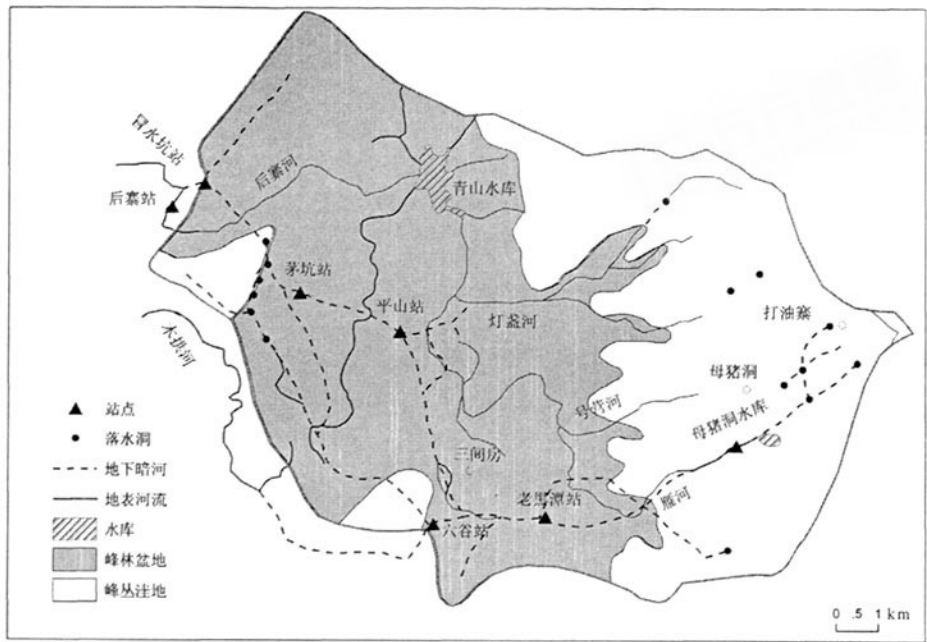


图1 后寨河流域水系分布图
Fig.1 Houzhai watershed system and stream network

2 基本原理与方法

2.1 BP 神经网络模型

本文采用广泛应用于降雨—径流模拟的误差反传播算法(Back Propagation,简称BP)^[7]。BP 网络不仅结构简单,而且具有很好的非线性映射能力。以 (n, m_1, m_2, q) 双隐层BP 网络结构(图2)为例, w_{ij} 表示输入层单元 i 与隐层单元 j 之间的连接权重,隐层神经元的净输入 S_j 为:

$$S_j = \sum_{i=1}^n w_{ij} x_i + \theta_j \quad j = 1, 2, \dots, m_1 \quad (1)$$

式中, θ_j 为阈值, m_1 表示隐含层神经元个数。净输入经过转移函数的作用后,得到神经元的输出 b_j 为:

$$b_j = f(S_j) = f\left(\sum_{i=1}^n w_{ij} x_i + \theta_j\right) \quad j = 1, 2, \dots, m_1 \quad (2)$$

转移函数 $f(u)$ 又称激活函数,能模拟生物神经元所具有的非线性转移特性。

$$f(u) = \frac{1}{1 + \exp(-u)} \quad (3)$$

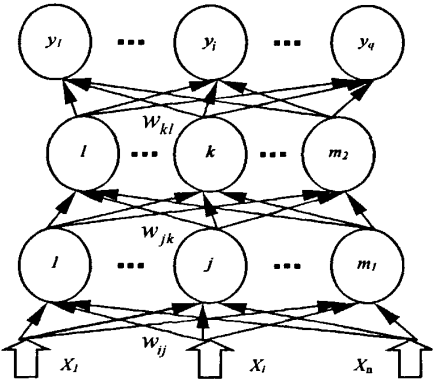


图2 BP 神经网络结构示意图
Fig.2 Sketch map of BP ANN structure

应用梯度法调节网络中的连接权重,使实际输出 (y_i) 与期望输出 $(\hat{y}_i)$ 的误差 E 小于等于一限定值 E_0 。

$$E = \frac{1}{2} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \leq E_0 \quad (4)$$

模型计算结果采用效率系数(NSC)、均方根误差(RMSE)、平均相对误差绝对值(MPE)三个指标作为检验标准,表达式为:

$$NSC = 1 - \frac{\sum_{i=1}^N (T_i - Y_i)^2}{\sum_{i=1}^N (T_i - \bar{T})^2} \quad (5)$$

$$RMSE = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (Y_i - T_i)^2} \quad (6)$$

$$MPE = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i - T_i|}{T_i} \quad (7)$$

式中, T_i 、 Y_i 分别为实测和计算值, $\bar{T}$ 为实测值均值, N 为观测系列长度。

2.2 神经网络权重分析法

流域某一时刻出口断面流量的影响因子有前期流量、上游站点前期水位、降雨量、蒸发量等^[8]。通常假定预报因子与预报对象为线性关系,采用相关分析,选取相关系数最大的因子作为模型输入^[9]。但由于预报因子与预报对象之间的关系大都呈非线性,采用线性方法所选因子不一定最优,因而需要采用非线性方法确定预报因子与预报对象之间的因果关系,确定最优预报因子。

ANN 模型通过神经元之间的连接权重实现输入-输出的非线性映射^[10],可用于选择最佳预报因子^[11]。以图2所示双隐层网络结构为例,分析第 i 个输入($i=1,2,\dots,n$)对第 l 个输出的影响权重,它与相连的 j,k 个神经元有关。第 i 个输入量对第一隐层上第 j 个神经元的影响表示为 w_{ij} ,来自第一隐层,施加于第二隐层第 k 个神经元上的作用信号为:

$$R_k = \sum_{j=1}^{m_1} |w_{ij}| |w_{jk}| \quad k = 1, 2, \dots, m_2 \quad (8)$$

来自第二隐层,施加于输出量 l 的作用信号为:

$$T_{il} = S_l = \sum_{k=1}^{m_2} R_k |w_{kl}| \quad (9)$$

将第 i 个输入变量对第 l 个输出变量的影响程度归一化得到权重系数 T_{il} :

$$T_{il} = T_{il} / \sum_{i=1}^n T_{il} \quad (10)$$

3 实例应用

本文以贵州省普定后寨河流域为例,利用1980—1989年(经对该记录数据的合理性及可靠性分析,数据质量较好)逐日降雨、冒水坑地下暗河径流观测资料,采用神经网络权重分析法确定网络隐含层数,选取权重贡献度最大的预报因子,建立ANN模型。

3.1 神经网络隐含层数确定

ANN模型建立的关键在于隐含层数和隐含层神经元数的确定。通常凭经验选取隐含层数,认为两个隐含层足以表示输入图形的任意输出函数^[12,13]。

本文在神经网络权重分析方法的基础上,考虑模型收敛性、稳定性两个指标,选取最佳神经网络隐含层数,为确定神经网络结构提供分析手段。

由于ANN模型随机给定神经元初始权重值,不同网络结构,模型收敛情况不一。为选择模型收敛性最好的网络结构,以降雨系列 P_t, P_{t-1}, P_{t-2} ,以及冒水坑前期流量系列 $Q_{t-1}, Q_{t-2}, Q_{t-3}$ 作为模型输入,分别建立单隐层、双隐层、三隐层网络结构。随机给定不同的权重初始值,模型训练次数取2 000次(模拟误差达到误差定值 E_0 ,神经元权重修正量很小,权重系数稳定),模型运行100次,从中选择效率系数(NSC)最大的16次单层、双层及三层网络结构的权重系数(图3)。可以看出,单隐层网络的权重系数与输入因子之

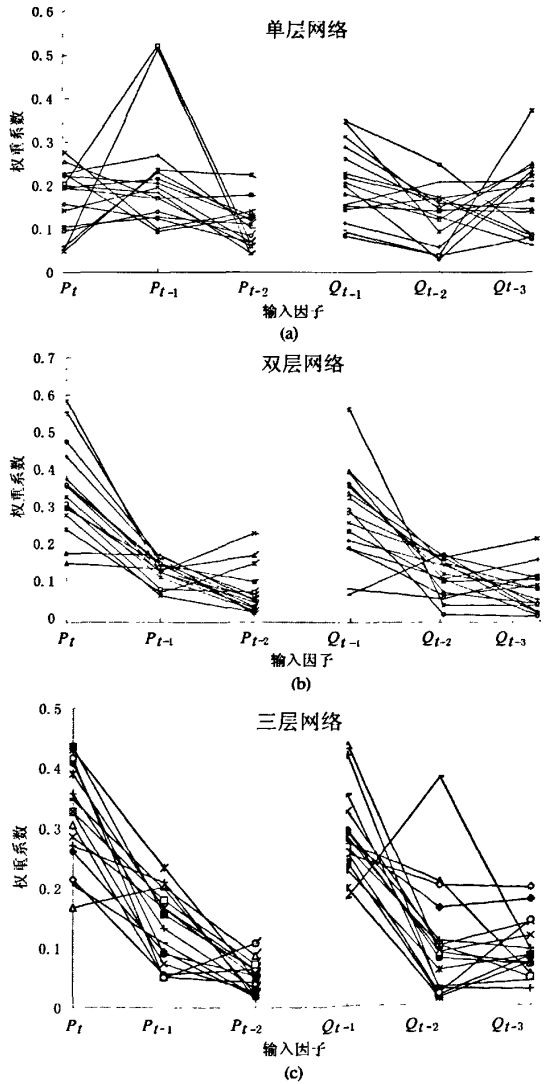


图3 权重系数随输入因子变化图

Fig. 3 Variation in weight coefficients with input factors

间关系紊乱, P_{t-1} 的权重系数比 P_t 大, Q_{t-3} 的权重系数比 Q_{t-2} 大, 不能反映喀斯特小流域降雨—径流陡涨、陡落水文特征及降雨—径流响应关系。相对而言, 双隐层网络权重系数最为稳定, 因此模型采用双隐层网络结构。

以降雨系列 P_t 、 P_{t-1} 、 P_{t-2} 以及前期流量系列 Q_{t-1} 、 Q_{t-2} 、 Q_{t-3} 作为模型输入, 模型运行 100 次的效率系数 (NSC) 统计值如表 1, 双隐层网络 NSC 大于 0.8 和 0.9 分别为 85 和 76 次, 比单隐层和三隐层 NSC 大, 进一步说明双隐层网络结构模拟精度最高、稳定性最好。

表 1 不同隐层结构的 ANN 模型效率系数统计表(单位: 次)
Tab. 1 Statistical results of NSC for ANN model with different hidden layers

效率系数	单隐层	双隐层	三隐层
$NSC \geq 0.8$	75	85	78
$NSC \geq 0.9$	65	76	64
$NSC \geq 0.95$	16	35	33
$NSC \geq 0.96$	4	11	11

3.2 模型输入变量选择

选取 P_t 、 P_{t-1} 、 P_{t-2} 、 Q_{t-1} 、 Q_{t-2} 、 Q_{t-3} 6 个因子作为 ANN 模型初选预报因子, 基于双隐层 ANN 模型, 进一步采用神经网络权重分析法(式 8~10)计算得出 6 个输入因子的权重系数(表 2)。由表 2 可知, 初选因子中 P_t 、 P_{t-1} 、 Q_{t-1} 、 Q_{t-2} 权重贡献度大, 且 P_t 权重贡献度大于 Q_{t-1} , 说明当日降雨 (P_t) 对模型输出 (Q_t) 影响最大, 其它因子权重贡献度依次是 Q_{t-1} 、 P_{t-1} 、 Q_{t-2} 、 P_{t-2} 。

表 2 ANN 模型影响因子权重系数
Tab. 2 Weight coefficients of impact factors for ANN model

影响因子	P_t	P_{t-1}	P_{t-2}	Q_{t-1}	Q_{t-2}	Q_{t-3}
平均权重系数	0.355	0.123	0.063	0.275	0.106	0.079

按权重大小, 选取降雨—径流因子中权重重大的前 3 个因子 P_t 、 P_{t-1} 、 Q_{t-1} 和前 4 个因子 P_t 、 P_{t-1} 、 Q_{t-1} 、 Q_{t-2} 为模型输入, 模型运行 100 次, 效率系数 NSC 分布如表 3 所示。 NSC 大于 0.9 模拟的径流所占次数从 6 个输入因子的 76 次, 到 4 个因子增加到 82 次, 3 个因子的 85 次(表 3)。表明因子数太多会发生信息冗余, 影响模型稳定性。

相关分析表明, 预报因子 Q_{t-1} 、 Q_{t-2} 与 Q_t 互相关系数分别为 0.936、0.842, 而 Q_{t-2} 与 Q_{t-1} 因子之间互相关系数达 0.936, P_t 、 P_{t-1} 之间互相关系数仅为 0.168, 说明 Q_{t-2} 、 Q_{t-1} 因子间相关程度高、重复信息多, 对模

表 3 双隐层 ANN 模型不同影响因子效率系数统计值(单位: 次)
Tab. 3 Statistical results of NSC for ANN model of double hidden layers with different input factors

输入因子	$P_t, P_{t-1}, P_{t-2}, Q_{t-1}, Q_{t-2}, Q_{t-3}$	$P_t, P_{t-1}, Q_{t-1}, Q_{t-2}$	P_t, P_{t-1}, Q_{t-1}
$NSC \geq 0.8$	85	89	91
$NSC \geq 0.9$	76	82	85
$NSC \geq 0.95$	35	68	80
$NSC \geq 0.96$	11	23	64

型预报不利。根据预报因子之间的关系越小越好, 预报因子与预报对象的关系越大越好原则^[14], 且考虑到 Q_{t-2} 、 Q_{t-3} 的权重贡献度很小且增加 Q_{t-2} 对 NSC 影响有限, 最终选取 P_t 、 P_{t-1} 、 Q_{t-1} 作为预报因子, 其地下暗河径流的 NSC 值大于 0.95 和 0.96 的次数为 80 和 64 次, 模型收敛性好, 计算结果可靠。

3.3 降雨径流模拟结果分析

为了进一步验证所选择的 ANN 模型结构及输入变量对研究区降雨径流模拟结果的稳定性, 本文利用 1980—1989 年日降雨、流量资料, 分别选取三种不同年份的训练样本及三种不同的验证样本(包括一组随机选取的验证样本 1982 年和 1986 年)对模型进行交叉训练和验证, 交叉验证结果如表 4 所示。

表 4 ANN 模型交叉验证结果
Tab. 4 The cross-validation results of ANN Model

验证方案	训练及验证时间	NSC	$RMSE/m^3/s$	$MPE/\%$
方案 1	训练期: 1980—1987	0.926	0.155	8.25
	验证期: 1988—1989	0.965	0.102	7.41
方案 2	训练期: 1982—1989	0.940	0.145	7.90
	验证期: 1980—1981	0.933	0.146	6.63
方案 3	训练期: 1980—1981、1983—1985、1987—1989	0.944	0.141	7.71
	验证期: 1982、1986	0.881	0.193	9.66

由表 4 可以看出, 不同验证期, 模型计算结果精度不一。以 1980—1987 年作为模型训练样本、1988—1989 年作为验证样本的计算精度最高, 效率系数 NSC 均在 0.9 以上, 达到了《SL250—2000 水文情报预报规范》^[15] 中的甲等标准(甲等标准为大于 0.90)。以 1982、1986 年为验证期的计算精度最低, 但 NSC 也达 0.881, 达到了乙等标准(乙等标准为 0.70~0.90)。从 NSC 、 $RMSE$ 、 MPE 三个指标可以看出, 采用神经网络权重分析法选取预报因子, 模型的预报精度高, 稳定性、收敛性好, ANN 模型能较好地应用于西

南喀斯特流域水文过程模拟。由此可见神经网络权重分析法是选取水文预报因子的一种实用方法,为水文预报因子选择提供了一种有效的非线性方法。

4 结 论

本文选择普定后寨河流域,利用1980—1989年实测降雨、地下暗河流量资料,研究了神经网络权重分析方法确定喀斯特流域水文预报因子与预报对象之间的定量关系。根据BP模型模拟的径流效率系数随神经网络隐含层数和输入变量个数的变化,确定BP模型神经网络结构。模型计算结果表明,在普定后寨河喀斯特流域,可采用双隐层、三输入(与预报对象同时期及前一时期的降雨量及同时期径流量)的网络结构。经交叉验证的降雨-径流模拟结果表明,模型训练期NSC达0.92以上,验证期NSC也达0.881,说明神经网络分析方法,能较好地确定喀斯特流域神经网络结构和模拟降雨径流过程。

致 谢:感谢贵州省普定岩溶地质与生态研究综合试验站提供资料。

参考文献

- [1] 杨明德. 峰丛洼地形成动力过程与水资源开发利用[J]. 中国岩溶, 2000, 19(1): 44—51.
- [2] 洪勇, 日浦启全, 栾茂田. Tank模型在模拟滑坡地下水位中的应用[J]. 吉林大学学报(地球科学版), 2006, 36: 69—74.
- [3] Chen Xi, ed al. Simulation of rainfall—underground outflow responses of a karstic watershed in southwest China with an artificial neural network [J]. Water Science and Engineering, 2008, 1(2): 1—9.
- [4] 高仁祥, 张世英, 刘豹. 基于神经网络的变量选择方法[J]. 系统工程学报, 1998, 13(2): 32—37.
- [5] 陈孝源. 因子的可预报性和预报模型适用性研究探讨[J]. 大气科学, 1994, 18(1): 122—126.
- [6] 王腊春, 杨晓轩. 喀斯特流域空间结构与地貌瞬时单位线应用——以贵州后寨地下河流域为例[J]. 南京大学学报, 1995, 31(4): 649—656.
- [7] Maier H R, Dandy G C. Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications [J]. Environmental Modeling and Software, 2000, 15: 101—23.
- [8] 陈喜, 吴净禄, 王玲. 人工神经网络模型预测气候变化对博斯腾湖流域径流影响[J]. 湖泊科学, 2005, 17(3): 207—212.
- [9] 张利平, 王德智, 夏军, 等. 基于气象因子的中长期水文预报方法研究[J]. 水电能源科学, 2003, 21(3): 4—6.
- [10] 黄洪钟, 黄文培, 王金诺. 前馈神经网络的一种权重分析方法及其在机械结构分析中的应用[J]. 机械科学与技术, 2004, 6: 114—118.
- [11] 陈守煜, 冀鸿兰. 冰凌预报模糊优选神经网络BP方法[J]. 水利学报, 2004, 6: 114—118.
- [12] Flood I & Kartam N. Neural networks in civil engineering I: principles and understanding [J]. Journal of Computing in Civil Engineering, 1994, 8(2): 131—139.
- [13] Flood I & Kartam N. neural networks in civil engineering II: system and application [J]. Journal of Computing in Civil Engineering, 1994, 8(2): 140—148.
- [14] 阎清元. 天气预报相关因子的选择和使用[J]. 内蒙古气象, 1997, 5: 12—15.
- [15] 中华人民共和国水利部. SL250—2000水文情报预报规范[S]. 2000: 18—20.

Analysis and simulation to the relation between rainfall and runoff in karst basin with artificial neural network model

CHEN Cai, CHEN Xi, ZHANG Zhi-cai, WEI Lin-na

(State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, Jiangsu 210098, China)

Abstract: Rainfall-runoff response in karst basin is a non-linear process. Determination of the major factors influencing underground river flow by proper non-linear analysis method is very important for simulating karst hydrological processes. In this study, the observed rainfall and flow discharge data in the Houzhai catchment of Puding County was used. The BP model structure of two hidden layers and three inputs in the study catchment was determined by the neural network weight analysis. This structure is able to keep stability of the rainfall-runoff simulation. Cross training and validation results of the BP model show that efficient coefficient (NSC) is over 0.9 during the training period, and NSC is over 0.88 during the validation period. Therefore, the neural network weight analysis can be used to determine the relationship between forecast object and its influencing factors. The artificial neural network model offers an efficient way for rainfall-runoff simulation in karst basin.

Key words: karst; neural network; weight analysis; rainfall-runoff simulation; model structure; Houzhai catchment in Puding; Guizhou