

组合灰色预测模型在入库流量预测中的应用

郭 华¹, 陈 勇², 马耀光^{1*}

(1. 西北农林科技大学水利与建筑工程学院, 陕西 杨凌 712100; 2. 南京市水利规划设计院, 江苏 南京 210006)

摘 要: 本文将灰色 GM(1,1) 模型、BP 人工神经网络和马尔柯夫链相结合, 利用历年入库流量及干河径流量建立组合模型对入库流量进行预测。GM(1,1) 模型主要预测趋势, 其前半部分与实测值拟合较好, BP 神经网络模型后半部有波动部分与实测值拟合较好, 二者结合使相对误差最小建立组合模型, 同时运用马尔柯夫链预测入库流量的变化范围。预测 2001 和 2002 年的入库流量对模型进行检验; GM(1,1) 模型预测的相对误差分别为 0.359 和 -0.017; BP 神经网络预测的相对误差分别为 0.032 和 -0.251, 组合模型相对误差分别为 0.164 和 0.117, 组合预测值在预测区间之内, 该组合模型预测结果合理有效, 能更精确预测冯家山水库入库流量。

关键词: 灰色 GM(1,1) 模型; BP 神经网络; 马尔柯夫链; 预测模型; 入库流量; 冯家山水库

中图分类号: TV213.9 **文献标识码:** A **文章编号:** 1000-7601(2012)03-0096-05

水库是储存水资源的特有方式, 具有防洪、灌溉、养鱼、工业用水、发电、人畜饮用、旅游等多种功能, 不少水库已成为城市供水的重要水源。入库流量的预测可为水资源合理开发、优化利用和合理调度提供必要的依据。目前国内径流量预测的方法主要有时间序列法、模糊分析法、小波分析法、灰色系统法、人工神经网络法、马尔柯夫链、信息熵分析法等^[1]。

灰色 GM(1,1) 模型要求数据序列成指数规律变化, 且在建模前通过累加处理使数据的随机性弱化。因此, 该模型不适合长期的、随机性和波动性较大的数据序列预测, 通常 GM(1,1) 模型用来揭示数据的发展趋势^[2-6]。人工神经网络^[7]是对人类大脑系统的一阶特性的一种描述。它是一个数学模型, 可以用电子线路来实现, 也可以用计算机程序来模拟, 是人工智能研究的一种方法。神经网络适宜于处理具有残缺结构和含有错误成分的模式, 能够在信息源信息含糊、不确定、不完整、存在矛盾及假象等复杂环境中处理模式。而马尔可夫随机性过程理论指出: 系统将来所处的状态只与现在系统状态有关, 而与系统过去状态无关即无后效性, 它可以描绘一个随机变化的动态系统, 根据状态之间的转移概率, 来推测一个系统未来的发展变化, 适合描述随机波动性较大的预测问题, 通常利用该模型来确定状态的转移规律^[8-12]。因而, 可将三个模型结合, 弥补各模型的不足, 并将其应用于水库入库流量的预测。

1 模型简介

1.1 GM(1,1) 模型^[2]

在系统论中把一个既含有已知信息又含有未知信息或非确定信息的系统称之为灰色系统。在年径流量预测中, 整个流域系统内有已知的也有未知的信息, 所以年径流量预测可看做灰色系统的预测。

给定的原始时间序列为 $x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(t), \dots, x^{(0)}(n)$, 将原始数据进行累加, 得到新的数据序列 $x^{(1)}(1), x^{(1)}(2), x^{(1)}(3), \dots, x^{(1)}(t), \dots, x^{(1)}(n)$ 。

其中 $x^{(1)}(j) = \sum_{k=1}^j x^{(0)}(k), j = 1, 2, \dots, n$ (1)

对累加后的数据建立 GM(1,1) 模型的一阶微分方程:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u \quad (2)$$

根据微分方程, 运用最小二乘法, 求得 a, u , 得到微分方程

$$\hat{x}^{(1)}(t+1) = (x^{(0)}(1) - \frac{u}{a})e^{-at} + \frac{u}{a} \quad (3)$$

由于预测方程是对累加数据系列的预测方程, 因而进行累减还原, 可以得到数据系列的预测值。还原模型为:

$$\hat{x}^{(0)}(t+1) = \hat{x}^{(1)}(t+1) - \hat{x}^{(1)}(t) \quad (4)$$

实测值与预测值的残差为:

$$\zeta^{(0)}(t) = x^{(0)}(t) - \hat{x}^{(0)}(t) \quad (5)$$

收稿日期: 2012-02-27

基金项目: 国家自然科学基金(50879071)

作者简介: 郭 华(1986—), 女, 山西临汾人, 研究生, 主要从事水文与水资源研究。E-mail: guohua12@yahoo.com.cn。

* 通讯作者: 马耀光(1957—), 男, 陕西武功人, 硕士生导师, 主要从事水资源与水环境保护方面的教学工作。E-mail: myg0609@yahoo.com.cn。

相对误差为:

$$q = \zeta^{(0)}(t)/x^{(0)}(t) \times 100\% \quad (6)$$

为检验模型群中单个模型的拟合程度,可进行后验差检验。原始数据的均方差与预测值的均方差比值为后验差比值,即后验差比值

$$C = S_2/S_1 \quad (7)$$

小误差概率

$$P = \{|\zeta^{(0)}(t) - \bar{\zeta}^{(0)}| < 0.6745 S_1\} \quad (8)$$

模型的精度由上述后验差比值和小误差概率共同确定,因此 C 和 P 为检验精度指标。 $GM(1,1)$ 预测精度等级如表 1 所示,当后验差检验精度不合格时,可通过建立残差模型来对原模型进行修正。

表 1 $GM(1,1)$ 预测精度等级

Table 1 Prediction precision grade of $GM(1,1)$

等级 Grade	好 Excellence	合格 Pass	勉强 Reluctance pass	不合格 No pass
C	< 0.35	< 0.50	< 0.65	≥ 0.65
P	> 0.95	> 0.80	> 0.7	≤ 0.7

1.2 人工神经网络^[12]

人工神经网络是理论化的人脑神经网络的数学模型,是基于模仿大脑神经网络结构和功能而建立起来的一种信息处理系统。是把算法和结构统一为一体的系统,是一种软件和硬件的混合体,在某种程度上模拟大脑的结构,所以更有智慧,可能有更快的计算速度。

在时间序列预测中,BP 神经网络是最常使用的网络。BP 模型把一组样本的 I/O 问题变为一个非线性优化问题,它使用的是优化中最普通的梯度下降法。如果把神经网络看成输入到输出的映射,则这个映射是一个高度非线性映射。

建模步骤如下:(1) 确定神经网络的结构参数:包括定义输入层、隐含层和输出层的神经元个数,确定最小训练速率、动态参数、允许误差、迭代次数和 Sigmoid 参数;(2) 对输入层样本值进行归一化处理;(3) 网络权值初始化;(4) 从前向后计算各隐含层到输出层单元的输出向量;(5) 计算输出层的误差信号;(6) 从后向前计算隐含层的误差信号;(7) 计算并保存各权值修正量;(8) 修正权值并保存;(9) 计算误差。

利用经训练所保存的网络权值,根据网络输入的各项指标值,运行网络向前计算,即可得实际输出值。根据网络的实际输入确定其性能优劣,并确定其是否具有很好的泛化能力。

1.3 马尔柯夫预测模型

马尔柯夫链预测^[7] 是根据初始的状态概率向

量和状态概率转移矩阵来推出某一时期所处状态的一种方法,其理论基础是马尔柯夫过程。它描述的是一个随机事件序列的动态变化过程。

1.3.1 状态划分 根据马尔柯夫链将数据序列划分为多个不同状态,用 $1, 2, 3, 4, \dots, n$ 分别来表示;并且状态转移只在 $t_1, t_2, t_3, \dots, t_n$ 等可数时刻发生。

1.3.2 状态转移概率矩阵 马氏链由状态 i 经过 k 步转移到状态 j 的转移概率用 $p_{ij}^{(k)}$ 表示。其中一部状态转移概率矩阵用 $P^{(1)}$ 表示。可以证明^[7],一部转移概率矩阵与 k 步转移概率矩阵存在以下关系:

$$P^{(k)} = [P^{(1)}]^k \quad (9)$$

1.3.3 预测模型 令时刻 t 的无条件概率分布或边际概率分布为 $P_t = [p_t(1), p_t(2), \dots, p_t(m)]$, 其中 $p_t(j)$ 是概率 $P[X(t) = a_j]$ 。若时刻 t 已发生,则 P_t 已知。那么, $t+1$ 时刻的条件分布为

$$P_{t+1} = P_t P^{(1)} \quad (10)$$

$$\text{以此类推,有 } P_{t+1} = P_t [P^{(1)}]^{t+1} \quad (11)$$

其中 P_0 为开始时刻的无条件概率分布。

2 组合模型预测方法

2.1 建模原理

建模的基本思想:首先建立 $GM(1,1)$ 灰色预测模型并检验,得出预测值序列;建立 BP 神经网络模型,得出预测值,再与灰色预测进行最优组合得到组合预测值,计算其与实测序列的相对残差序列,依其均值和方差进行状态区间的划分。根据落入各状态区间的点计算出转移概率矩阵,预测未来状态,就可以找到预测值的区间,即是要求的预测范围值。

2.2 建模步骤

灰色组合预测模型的具体步骤如下:

(1) 建立 $GM(1,1)$ 灰色模型并检验该模型。

(2) 运用 BP 神经网络预测时间序列的值,运用最优组合得到组合预测值。

(3) 划分状态。根据预测对象的具体情况,划分若干状态区间。由组合模型预测序列 $\hat{X}$ 与实测序列 $X(t)$ 残差相对值序列 $q(t)$ 的均值 $\bar{q}$ 和方差 s_q ,建立指标分级标准,即相当于确定马尔柯夫链的状态空间,一般以 $\bar{q}$ 为中心,以 s_q 为标准分组。可将数据序列分成以下 5 组: $(-\infty, \bar{q} - s_q], (\bar{q} - s_q, \bar{q} - 0.5s_q], (\bar{q} - 0.5s_q, \bar{q} + 0.5s_q], (\bar{q} + 0.5s_q, \bar{q} + s_q], (\bar{q} + s_q, \infty)$ 。

(4) 计算转移概率矩阵,计算方法同 1.3。

(5) “马氏性”检验。通常离散序列的马氏链可用 χ^2 统计量来检验。

$\chi^2 = 2 \sum_{i=1}^m \sum_{j=1}^m f_{ij} \left| \log \frac{p_{ij}}{p_{.j}} \right|$ 统计量 χ^2 服从自由度为 $(m-1)^2$ 的 χ^2 分布。给定显著性水平 α , 若 $\chi^2 > \chi^2[(m-1)^2]$, 则满足马氏性。其中:

$$p_{.j} = \frac{\sum_{i=1}^m f_{ij}}{\sum_{i=1}^m \sum_{j=1}^m f_{ij}}$$

(6) 预测。由式(9)可得 k 步以后的状态向量, 以此判断后步所处的状态区间, 从而得到最终预测结果。

3 结果与分析

3.1 冯家山水库年入库径流量预测计算过程

本研究为简洁说明此方法, 仅选取了冯家山水库 1983—2000 年入库径流量为样本(表 2), 预测 2001—2002 的径流量。验证该方法有效性。

表 2 冯家山水库入库年径流量预测及误差 (10^7 m^3)

Table 2 Prediction and error of flow into Fengjiashan reservoir

序号 Sequence number	年份 Year	实测值 Measured value	GM(1,1)预测值 Predicted value of GM(1,1)	相对误差 Relative error	BP 预测值 Predicted value of BP	相对误差 Relative error
1	1983	76.152	76.1520	0.466989	40.589857	0
2	1984	66.839	55.7541	0.393712	40.523679	0.165845
3	1985	50.343	51.2180	0.207130	39.915454	-0.01738
4	1986	25.085	47.0509	-0.107710	27.786970	-0.87566
5	1987	21.569	43.2229	-0.054080	22.735529	-1.00394
6	1988	56.020	39.7063	0.281267	40.263442	0.291212
7	1989	47.396	36.4758	0.164108	39.617934	0.230403
8	1990	57.837	33.5082	0.302613	40.334783	0.420644
9	1991	19.941	30.782	-0.01326	20.205342	-0.54365
10	1992	33.317	28.2776	-0.06174	35.374011	0.151256
11	1993	37.474	25.9769	0.002574	37.377525	0.306802
12	1994	20.088	23.8635	-0.01746	20.438738	-0.18795
13	1995	6.498	21.9219	0.047481	6.1894688	-2.37364
14	1996	18.945	20.1384	0.017611	18.611355	-0.06299
15	1997	5.092	18.4999	-0.14791	5.8451821	-2.63313
16	1998	19.382	16.9948	0.00345	19.315129	0.123166
17	1999	20.416	15.6121	-0.02677	20.962541	0.235301
18	2000	15.868	14.3419	0.001288	15.847565	0.096175

由表 2 和图 1 可知, 灰色模型主要预测其趋势, 前半部分与实测值拟合较好, BP 神经网络模拟在后半部有波动部分与实测值拟合较好, 两种模型对 2001—2002 年的预测结果与实测值都拟合较好。运用两种模型对历年入库流量的预测结果, 使用 MATLAB 进行线性规划求解, 使得相对误差最小, 最后确

定灰色模型和 BP 神经网络的权重系数分别为 0.8, 0.2。但是, 由图 1 可以直观看出, 两种模型的预测结果越来越贴近实测值, 不分伯仲, 在随后两年的预测中, 确定其权重系数分别为 0.5 与实际情况较为符合^[13-15]。

3.1.1 GM(1,1) 预测模型 以 1983—2000 年径流量作为原始序列 $X^{(0)}(t)$, 由前面的建模步建立并检验 GM(1,1) 灰色模型预测模型为:

$$X(t+1) = -685.2818 \exp(-0.08486t) + 761.4338$$

经过分析计算: $P = 94.4\%$, $C = 0.374$ 。由预测等级表可知, 该灰色模型精度为“合格”, 不必对模型进行修正。

3.1.2 灰色 BP 神经网络预测^[11] 千河径流量与入库流量相关性较大, 以其作为预报因子, 运用 DPS 数据处理系统 BP 神经网络模型预测出冯家山水库的入库流量。进入 BP 神经网络训练时, 输入最小训练速率取 0.1, 动态参数 0.7, Sigmoid 参数为 0.9, 允许误差 0.00001, 最大迭代次数 1 000。并对输入节点的数值进行标准化转换。学习样本的拟合值和实际观察值, 灰色模型的预测值及误差列于表 2。

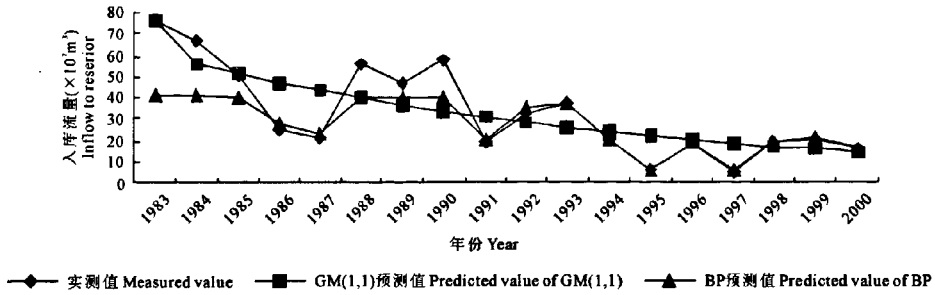


图1 GM(1,1)与BP预测比较

Fig.1 Predicted value comparison of GM(1,1) and BP

3.1.3 状态划分 由前面的计算步骤,可得径流量预测序列 $\hat{X}$ 与实测序列 $X(t)$ 残差相对值序列 $q(t)$ 的均值 $\bar{q}$ 和方差 $s_q: \bar{q} = 0.539389, s_q = 0.5867$ 。根据数据量不多的实际情况,只划分4个状态: $(-0.04735, 0.246018], (0.246018, 0.832759], (0.832759, 1.12613], (1.12613, 3]$, 此时 $m = 4$ 。

3.1.4 “马氏性”检验 计算出统计量: $\chi^2 = 2 \sum_{i=1}^m \sum_{j=1}^m f_{ij} \left| \log \frac{p_{ij}}{p_{.j}} \right| = 5.8818$; 取置信度 $\alpha = 0.05$, 查表得 $\chi^2[(m-1)^2] = 3.325$, 可知, $\chi^2 > \chi^2[(m-1)^2]$, 则可认为该径流量序列符合“马氏性”。

3.1.5 计算状态转移矩阵 求得一步状态转移矩阵为:

$$P^{(1)} = \begin{bmatrix} 0.44 & 0.22 & 0.11 & 0.22 \\ 0.75 & 0.25 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$
$$P^{(2)} = \begin{bmatrix} 0.59 & 0.21 & 0.10 & 0.10 \\ 0.52 & 0.23 & 0.08 & 0.17 \\ 0.375 & 0.375 & 0.25 & 0 \\ 0.44 & 0.22 & 0.11 & 0.22 \end{bmatrix}$$

2000年入库径流量为 $P_{2000} = [1, 0, 0, 0]$ 。由式(1)可得 $P_{2001} = [0.44, 0.22, 0.11, 0.22]$ 。

再由公式

$$q^{(i)} = \frac{X^{(0)}(t) - \hat{X}^{(0)}(t)}{\hat{X}^{(0)}(t)}, \text{ 得 } \hat{X}^{(0)}(t) =$$

$$\frac{\hat{X}^{(0)}(t)}{1 - q}$$

式中, $\hat{X}^{(0)}(t)$ 为组合预测值, $X^{(0)}(t)$ 为还原后的区间分界值, q 为原状态区间的分界值。

由 $\hat{X}^{(0)}(19) = 17.186$, 可得出待测年预测值区间 $(4.297, 8.083], (8.083, 16.409], (16.409, 22.794]$ 和 $(22.794, 102.762]$, 待测年预测值处于上述区间的概率分别为 $0.11, 0.22, 0.44, 0.22$ 。可知, $X^{(0)}(t)$ 处于 $(16.409, 22.794]$ 的概率最大, 所以取 $(16.409, 22.794]$ 为待测值的预测区间。同理, 由 $\hat{X}^{(0)}(20) = 10.504$, $X^{(0)}(t)$ (2002年) 处于待测年预测值区间 $(2.626, 4.940], (4.940, 10.029], (10.029, 13.931]$ 和 $(13.931, 62.808]$ 的概率分别为 $0.1, 0.1, 0.59, 0.21$, 其中处于 $(10.029, 13.931]$ 的概率最大, 所以取其为待测值的预测区间。

3.2 结果分析

通过对冯家山水库入库年径流量 2001—2002 年的预测(表3)对模型进行检验, 灰色 GM(1,1) 模型预测的相对误差分别为 0.359 和 -0.017 ; BP 神经网络预测的相对误差为 0.032 和 -0.251 , 将两种模型分别以 0.5 的权重组合, 预测结果与实测值更加接近, 更稳定, 相对误差为 0.164 和 0.117 , 并且组合预测值在预测区间之内, 所以组合模型预测结果合理有效。两种模型组合通常只是给出了一个点值, 显然这不太妥当, 灰色马尔柯夫给了预测值的波动区间和概率, 更具说服力, 更易于接受。

表3 各预测模型与实测值比较 (10^7m^3)

Table 3 Comparison of predicted and measured values of each model

年份 Year	实测值 Measured value	GM(1,1)预测值 Predicted value of GM(1,1)	相对误差 Relative error	BP 预测值 Predicted value of BP	相对误差 Relative error	组合预测 Combined predicted value	相对误差 Relative error	预测区间 Range of prediction
2001	20.549	13.175	0.359	21.197	0.032	17.186	0.164	(16.409, 22.794)
2002	11.897	12.103	-0.017	8.906	-0.251	10.505	0.117	(10.029, 13.931)

4 结 论

本研究运用灰色理论、马尔柯夫理论及 BP 人工神经网络建立了径流量组合预测模型。应用该方法能充分利用时间序列给予的信息,同时避免了自然地理条件、流域下垫面状况、水文气象变化、人为影响等多种因素的干扰,同时又结合与入库流量相关性密切的因素补充预测,提高了预测精度,又考虑到径流量在不同年份下的状态。与单一灰色预测方法相比,不仅给出了预测变量所处的准确状态区间,而且给出了处于各个区间的分布概率,具有较强的科学性和实用性。

经冯家山水库入库年径流量预测验证,灰色理论、马尔柯夫理论及 BP 人工神经网络组合模型方法可用于径流量预测,具有很好的效果。与已有灰色模型预测相比,该组合模型将马尔柯夫链应用于河川径流量的预测。提出了灰色模型的修正法和“马氏性”检验法,使该预测模型更完善,更具说服力。但是,在状态划分中,有时会因分组不同得出不同的结果,虽然理论上状态划分越多,预测精度越高,如果状态过多,又会使各状态样本减少。如何根据现有资料合理分组,确定待测变量所处的状态,还需进一步的探讨。

参 考 文 献:

- [1] 李 彬,袁 鹏,常 江.GM(1,1)改进模型在年径流量预测上的应用[J].东北水利水电,2006,(2):28-30.
- [2] 邓聚龙.灰色预测与决策[M].武汉:华中理工大学出版社,1992.
- [3] 梁学玉,张 鑫,孙天青.组合灰色预测模型在城市用水量预测中的应用[J].人民黄河,2010,32(4):79-80.
- [4] 赵宇哲,武春友.灰色振荡序列 GM(1,1)模型及在城市用水中的应用[J].运筹与管理,2010,19(5):155-160.
- [5] 蒋绍阶,江崇国.灰色神经网络最优权组合模型预测城市需水量[J].重庆建筑大学学报,2008,30(2):113-115.
- [6] 佟春生,杨建霞,王义民.系统工程的理论与方法概论[M].北京:国防工业出版社,2005.
- [7] 王文圣,丁 晶,金菊良.随机水文学[M].北京:中国水利水电出版社,2008.
- [8] Bluementhal R M,Getoor R K. Markov Process and Potential Theory[M]. New York: Academy Press, 1986.
- [9] 蒋承仪.灰色马尔柯夫预测模型[J].重庆建筑大学学报,1996,(3):116-121.
- [10] 孙才志,张 戈,林学钰.加权马尔柯夫链在降水丰枯状况预测中的应用[J].系统工程理论与实践,2003,(4):100-105.
- [11] Liu S F, Deng I L. The rang suitable for GM(1,1)[J]. Journal of Greysystem, 1999, 11(1): 34-37.
- [12] 唐启义. DPS 数据处理系统:实验设计、统计分析及数据挖掘[M].北京:科学出版社,2010.
- [13] 金菊良,程吉林,魏一鸣,等.确定区域水资源分配权重的最小相对熵方法[J].水力发电学报,2007,26(1):28-32.
- [14] 罗贤琴,夏靖波,陈天平.网络性能评估中客观权重确定方法比较[J].计算机应用,2009,29(10):2624-2631.
- [15] 陶菊春,吴建民.综合加权评分法的综合权重确定新探[J].系统工程理论与实践,2001,(8):43-48.

Application of combined grey forecasting model in the prediction of incoming flow

GUO Hua¹, CHEN-Yong², MA Yao-guang¹*

(1. College of Water Resources and Architectural Engineering, Northwest A & F University, Yangling, Shaanxi 712100, China;

2. Nanjing Hydraulic Planning Design Institute, Nanjing, Jiangsu 210006, China)

Abstract: With the combination of gray GM(1, 1) model, BP artificial neural network and Markov chain, the incoming flow to a reservoir was predicted by using the data of past inflow and the Qian river runoff establish combination model. GM(1,1) model mainly predict trend, the first half and measured data is better, the BP neural network model second half part have volatility and measured data is better, making relative error smallest to establish combination model, and using markov chain predict incoming flow range change. Forecasting inflow of 2001 and 2002 to inspection flow model: the relative error of GM(1,1) model were 0.359 and 0.017; The relative error of BP neural network were 0.032 and 0.251, the relative error model portfolios were 0.164 and 0.117, combined prediction value within the range. This combination forecasting model is reasonable and effective, and can be more precise forecasts Fengjiashan reservoir inflow.

Keywords: GM(1,1); BP artificial neural network; Markov chain; prediction model; inflow; Fengjiashan reservoir