

Mid and Long-Term Hydrological Forecasting Using Optimal Combined Model*

Yue Liu¹, Pan Liu¹, Huankun Huang², Liping Li¹, Yunfei Yu², Anbiao Zhang²

¹State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan

²The Administration Office of Feilaixia Water Conservancy Project of Guangdong Province, Qingyuan
Email: liuyue_me@sina.com

Received: Sep. 30th, 2013; revised: Nov. 20th, 2013; accepted: Nov. 26th, 2013

Abstract: This paper applies six models, including the autoregressive model, the seasonal autoregressive model, the threshold autoregressive model, the nearest neighbor bootstrap regressive model, the artificial neural network model and the support vector machine model into the mid and long-term hydrological forecasting. Based on Feilaixia reservoir project, the results show that the artificial neural network model is able to time series very well. The support vector machine model has the powerful ability of not only the simulation but also the forecasting. The results of those models were combined by the optimal combined forecasting model. The mean absolute error and the mean square error are selected as the measurements. Relying on the merits of each single model, the results of the optimal combined forecasting model work very well and are very well in robustness.

Keywords: Feilaixia Reservoir; Optimal Combined Forecasting Model; Inflow; Hydrological Forecasting

中长期水文预报的最优组合模型研究*

刘 玥¹, 刘 攀¹, 黄焕坤², 李立平¹, 虞云飞², 张安标²

¹武汉大学水资源与水电工程科学国家重点实验室, 武汉

²广东省飞来峡水利枢纽管理处, 清远

Email: liuyue_me@sina.com

收稿日期: 2013年9月30日; 修回日期: 2013年11月20日; 录用日期: 2013年11月26日

摘 要: 本文建立了径流中长期预报的自回归、季节性自回归、门限自回归、最近邻抽样回归、人工神经网络、支持向量机等六种模型, 并对这些模型结果进行综合, 开展最优组合预报。以飞来峡水库为研究实例, 选取平均绝对误差和均方误差作为评价指标, 发现人工神经网络模型模拟精度较高; 支持向量机模型模拟精度高、且具有最好的预报性能; 最优组合预报模型综合各单一预报模型的优点, 结果稳健、通用性强。

关键词: 飞来峡水库; 最优组合预报模型; 入库径流; 水文预报

1. 引言

中长期水文预报^[1]是指对未来一段时期的水文要素进行科学的预测, 通常预见期在 3d 以上。它不仅

*基金项目: 教育部新世纪优秀人才支持计划(NCET-11-0401); 水利部公益性行业科研专项经费项目(201101003)。
作者简介: 刘玥(1990-), 女, 硕士研究生。

能为水利枢纽提供安全保障, 而且可最大化兴利效益。目前, 应用较为广泛的中长期水文预报方法是基于实测资料, 采用数理统计方法(如线性平稳自回归模型、季节性自回归模型等)或者智能拟合方法(如人工神经网络模型、支持向量机模型等)实现。针对单一预

报模型不确定性因素, 组合预报模型能综合各单一预报模型的有用信息, 使模型具有更高的预报精度, 增强预报的稳定性^[2]。

本文采用自回归模型、季节性自回归模型、门限自回归模型、最近邻抽样回归模型、神经网络模型和支持向量机模型等6种模型进行模拟, 对所得结果利用最优组合预报模型进行综合预报。以飞来峡水库为研究实例, 开展中长期水文预报研究。

2. 最优组合预报模型的构建

2.1. 最优组合预报模型原理

组合预报对多个单一预报结果进行综合, 可有效减小模型不确定性。其中算术平均法对每个单一预报方法分配相同的权重, 是最简单的组合预报方法。最优组合预报对单一预报结果赋予最优权重, 进行综合, 即最优加权法^[2]。

最优加权法^[2]依据最优准则构造目标函数, 在满足约束条件的前提下求得各单一模型的加权系数。设 $\{x_t\}, t=1, 2, \dots, N$ 为观察序列, N 为观察序列长度, $\{x_t(j)\}, j=1, 2, \dots, J$ 为 J 个数学模型模拟所得预报结果, e_{jt} 为第 t 时刻第 j 种单一预报模型的预报误差, 则最优加权模型的组合权系数 w_j 是以下规划问题的解(最小二乘法则):

$$\begin{cases} \min Z & Z = e'e = W'EW \\ \text{s.t.} & R'W = 1 \end{cases} \quad (1)$$

式中: W 为权数向量, E 为误差阵。 R 为元素均为1的 n 维列向量, 用 Lagrange 乘子法得最优权及最小 Z 值的解析表达式:

最优权向量:

$$W^* = (R'E^{-1}R)^{-1} E^{-1}R \quad (2)$$

最小误差平方和值:

$$Z^* = (R'E^{-1}R)^{-1} \quad (3)$$

2.2. 单一预报模型

2.2.1. 自回归模型

自回归模型^[3-5]可用来表示前后变量的相依性, 其表达式如下:

$$x_t = \mu + \varphi_1(x_{t-1} - \mu) + \varphi_2(x_{t-2} - \mu) + \dots + \varphi_p(x_{t-p} - \mu) \quad (4)$$

式中: μ 为序列 x_t 的均值; $\varphi_1, \varphi_2, \dots, \varphi_n$ 为自回归的权重系数, 称为自回归系数; p 为阶数。

2.2.2. 季节性自回归模型

径流过程也具有季节性为非平稳过程。考虑到径流的这种特性, 将模型形式表述为:

$$Q_{t,\tau} = \varphi_{0,\tau} + \varphi_{1,\tau}Q_{t,\tau-1} + \varepsilon_{t,\tau} \quad (5)$$

式中: t 代表年份, τ 代表月份; $Q_{t,\tau}$ 为第 t 年第 τ 月的月径流量; $\varphi_{0,\tau}$ 和 $\varphi_{1,\tau}$ 为第 τ 月的两个回归参数; $\varepsilon_{t,\tau}$ 为第 t 年第 τ 月的独立随机项。

2.2.3. 门限自回归模型

对于时间序列 $\{Z_t\}$, 在 $\{Z_t\}$ 的取值范围内引入 $L-1$ 个门限值 $r_j (j=1, 2, \dots, L-1)$, 将该范围分成 L 个区间, 并根据延迟步数 d 将 $\{Z_t\}$ 按 $\{Z_{t-d}\}$ 分配到不同的门限区间内, 再对不同区间内的 Z_t 采用不同的自回归模型来描述, 完成对时序 $\{Z_t\}$ 的整个非线性动态系统的描述。

门限自回归模型^[4,5]的一般形式为:

$$Z_t = \begin{cases} \varphi_0^1 + \sum_{i=1}^{p_1} \varphi_i^1 Z_{t-i} + \varepsilon_t^1 & Z_{t-d} \leq r_1 \\ \varphi_0^2 + \sum_{i=1}^{p_2} \varphi_i^2 Z_{t-i} + \varepsilon_t^2 & r_1 < Z_{t-d} \leq r_2 \\ \vdots & \\ \varphi_0^L + \sum_{i=1}^{p_L} \varphi_i^L Z_{t-i} + \varepsilon_t^L & r_{L-1} \leq Z_{t-d} \end{cases} \quad (6)$$

式中: r_1, r_2, \dots, r_{L-1} 为门限值; d 为门限滞时; L 为门限区间数; $\varphi_0^j, \varphi_1^j, \dots, \varphi_{p_j}^j$ 为第 j 区间自回归系数; p_j 为第 j 区间模型阶数 ($j=1, 2, \dots, L$); ε_t^j 是独立随机变量。

2.2.4. 最近邻抽样回归模型

最近邻抽样回归模型^[6]的基本思想是: 在已有的特征矢量 $D_t (t=P+1, P+2, \dots, n)$ 中, 总可得到 K 个特征矢量 $D_{1(i)}, D_{2(i)}, \dots, D_{K(i)}$ 与当前特征矢量 D_i 按欧式距离计算(公式(7))最近邻, 其对应的后续值分别为 $X_{1(i)}, X_{2(i)}, \dots, X_{K(i)}$ 。则距离越小, D_i 与 D_j 越近邻, $X_{j(i)}$ 对 X_i 的贡献越大。

$$r_{i(i)} = \left[\sum_{j=1}^K (d_{ij} - d_{ij})^2 \right]^{1/2} \quad (7)$$

式中: $r_{i(i)}$ 表示 D_i 与 D_j 间的欧氏距离; d_{ij}, d_{ij} 分别为

D_j, D_i 的第 j 个元素; P 为特征矢量维数。

最近邻抽样回归模型的基本形式为:

$$X_i = \sum_{j=1}^K W_{j(i)} X_{j(i)} \quad (8)$$

式中: K 称为最近邻数; $W_{j(i)}$ 为 $X_{j(i)}$ 的抽样权重;

$$\sum_{j=1}^K W_{j(i)} = 1.0$$

2.2.5. 人工神经网络模型

人工神经网络^[5,6](Artificial Neural Networks)是一种应用类似于大脑神经突触联接的结构进行信息处理的数学模型。它的基本要素是人工神经元, 由输入层, 隐含层, 输出层三层组成。它主要模拟两个过程:

- (1) 对每个节点进行处理, 以确定其权值进行组合;
- (2) 确定其输出函数, 即激活函数。神经元模型的输入输出关系为

$$s_j = \sum_{i=0}^n w_{ji} x_i - \theta_j = \sum_{i=1}^n w_{ji} x_i \quad (x_0 = \theta_j, w_{j0} = -1) \quad (9)$$

$$y_j = f(s_j) \quad (10)$$

式中: w_{ji} 为连接权系数; $f(\cdot)$ 为输出激活函数。

2.2.6. 支持向量机模型

支持向量机是在约束条件(12)下求解下面的优化问题^[7-10]:

$$\min_{w, b, \xi, \xi^*} J = \frac{1}{2} w^T w + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (11)$$

约束条件:

$$s.t. \begin{cases} y_i - w^T \varphi(x_i) - b \leq \varepsilon + \xi_i \\ w^T \varphi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \quad i = 1, 2, \dots, l \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (12)$$

式中: ξ_i, ξ_i^* 为松弛变量; ε 为误差。常数 $C > 0$, 为惩罚函数, 它控制对超出误差 ε 的样本的惩罚程度。

3. 实例研究

飞来峡水利枢纽位于北江干流中下游, 广东省清远市辖区内, 上游距英德市 50 km, 下游距清远市 33 km, 坝址控制流域 34,097 km², 占北江流域面积的 73%, 占石角水文站控制面积 38,363 km² 的 88%。对飞来峡水库枢纽而言^[11], 准确的中长期水文预报不仅

能为水利枢纽提供安全保障, 而且可最大化兴利效益, 具有重要的作用。

本文选用广东省飞来峡水库 1953~2011 年入库径流资料, 选取自变量为飞来峡水库流量指标, 对飞来峡水库年、月、最大洪峰流量按照上述六种模型进行水文预报, 对所得结果进行最优权重分配得到最优组合模型预报结果, 将各模型所得结果进行列表比较。选取平均绝对误差 MAE 和均方误差 MSE 作为评价指标:

$$MAE = \frac{\sum_{i=1}^n |x'_i - x_i|}{n} \quad (13)$$

$$MSE = \sqrt{\frac{\sum_{i=1}^n (x'_i - x_i)^2}{n}} \quad (14)$$

式中: x'_i 为模拟所得预报值, x_i 为实测流量值, n 为计算时间序列的长度。

3.1. 年平均流量预报结果

取飞来峡水库 1953~1999 年入库年径流资料作为率定期建模, 2000~2011 年径流资料用于预报检验。各模型所得结果如表 1, 最优组合预报模型结果如图 1。

根据图 1 和表 1 结果可知:

1) 各单一预报模型中支持向量机模型预报精度最高, 模型的率定期和检验期的预报误差小于 20% 的分别为 89.4% 和 58.3%。支持向量机模型有较好的泛化性, 模拟精度高。

2) 其次精度较高的是人工神经网络模型, 模型的率定期和检验期的预报误差小于 20% 的分别为 70.6% 和 48%。人工神经网络模型可以以任意精度逼近任意函数, 容错能力较高, 可靠性强。

3) 最优组合预报模型精度高于各单一预报模型, 模型的率定期和检验期的预报误差小于 20% 的分别为 97.9% 和 91.7%。各单一预报模型只考虑了水文现象单方面的不确定性因素, 而最优组合预报模型将各单一预报模型的不确定性进行综合考虑, 得到的结果具有更高的精度和更强的稳定性。

3.2. 月平均流量预报结果

取飞来峡水库 1953~1999 年入库月径流资料作为率定期建模, 2000~2011 年月径流资料用于预报检验。各模型所得结果如表 2, 最优组合预报模型拟合结果

Table 1. Comparisons of Feilaixia reservoir annual average inflow models

表 1. 飞来峡水库入库年平均流量模型预报误差

模型	误差	<20%	<30%	MAE	MSE
自回归模型	率定期	50	82.4	218.7	288.2
	检验期	44	68	244.8	282.6
门限自回归模型	率定期	26.4	34.3	448.1	467.1
	检验期	20	40	434.4	426.4
最近邻抽样回归模型	率定期	74.47	80.84	73.2	101.74
	检验期	16.67	41.67	267.06	347.69
人工神经网络模型	拟合期	70.6	91.2	144.7	220
	检验期	48	72	269.4	342.8
支持向量机模型	率定期	89.4	97.9	67.3	95.3
	检验期	58.3	83.3	192.8	231.4
最优组合预报模型	率定期	97.9	97.9	62.9	76.5
	检验期	91.7	91.7	68.2	98.2

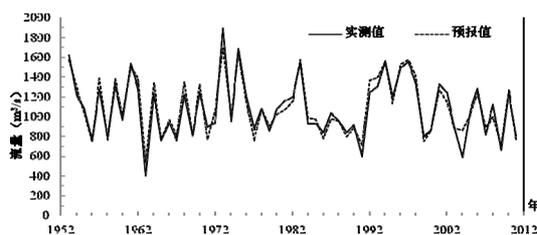


Figure 1. Comparisons of the optimal combination forecast model and the measured average annual inflow for the Feilaixia reservoir
图 1. 飞来峡水库入库年平均流量最优组合预报模型预报值与实测值比较图

Table 2. Comparisons of Feilaixia reservoir monthly inflow models
表 2. 飞来峡水库入库月平均流量模型预报误差

模型	误差	<20%	<30%	MAE	MSE
自回归模型	率定期	27.48	36.52	478.57	727.53
	检验期	25.69	38.19	450.67	638.61
季节性自回归模型	率定期	14.36	20.92	867.64	1142.39
	检验期	20.14	24.31	654.59	924
门限自回归模型	率定期	24.18	38.12	463.33	840.07
	检验期	30.46	41.67	498.42	763.03
最近邻抽样回归模型	率定期	48.87	77.66	234.74	374.97
	检验期	36.81	42.78	418.62	674.87
人工神经网络模型	拟合期	15.07	21.28	832.26	1151.91
	检验期	18.06	22.92	789.12	1073.82
支持向量机模型	率定期	92.2	98.76	37.14	48.94
	检验期	20.83	30.56	545.64	745.91
最优组合预报模型	率定期	84.75	93.44	79.5	116.37
	检验期	20.83	33.33	465.42	636.88

如图 2。

根据图 2 和表 2 所得结果可知:

1) 所用各单一模型中支持向量机模型精度最高, 模型的率定期和检验期的预报误差小于 20% 的分别为 92.2% 和 20.83%。支持向量机模型有较好的泛化性能, 结果精度最高。

2) 人工神经网络模型预报精度最低, 模型的率定期和检验期的预报误差小于 20% 分别为 15.07% 和 18.06%。因飞来峡水库入库月平均流量数据量较大, 出现局部极大值现象, 从而导致精度降低。

3) 最优组合预报模型的率定期和检验期预报误差小于 20% 的分别为 84.75% 和 20.83%, 与支持向量机模型相比精度较低, 但高于其余各单一预报模型。因最优组合预报模型选取的目标函数为误差方最小, 而评判指标选取的是误差小于 20% 的合格率, 故导致最优组合预报模型所得精度低于支持向量模型。

3.3. 最大洪峰流量预报结果

取飞来峡水库 1953~1999 年入库最大洪峰流量资料作为率定期建模, 2000~2012 年最大洪峰流量资料用于预报检验。各模型所得结果如表 3, 最优组合预报模型拟合结果如图 3。

根据图 3 和表 3 所得结果可知:

1) 各单一预报模型中支持向量机模型精度最高, 模型的率定期和检验期误差小于 20% 的分别为 80.85% 和 46.15%。支持向量机模型有较好的泛化性能, 所得结果精度较高。

2) 其次精度较高的是人工神经网络模型, 其率定期和检验期模型误差小于 20% 为 68.09% 和 38.46%。人工神经网络模型可以以任意精度逼近任意函数, 且在最大洪峰预报中数据量较少, 容错能力较高, 可靠性强。

3) 最优组合预报模型精度最高, 模型的率定期和检验期的预报误差小于 20% 的分别为 91.49% 和 46.15%。因各单一预报模型所得的误差均较小, 故最优组合预报模型所得精度最高。

4. 结论

本文选取飞来峡水库 1953~2011 年入库径流资料, 用自回归模型、季节性自回归模型、门限自回归模型、最近邻抽样模型、人工神经网络模型以及支持

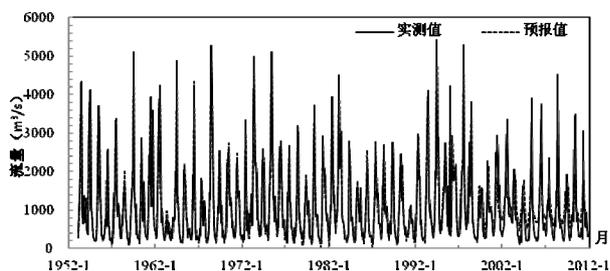


Figure 2. Comparisons of the optimal combination forecast model and the measured monthly inflow for the Feilaixia reservoir
图 2. 飞来峡水库入库月平均流量最优组合预报模型预报值与实测值比较图

Table 3. Comparison of Feilaixia reservoir maximum flood inflow models

表 3. 飞来峡水库入库最大洪峰流量模型预报误差

模型	误差	<20%	<30%	MAE	MSE
自回归模型	率定期	48.94	70.21	2220.09	2960.76
	检验期	30.77	76.92	2633.1	3330.84
门限自回归模型	率定期	29.79	44.68	3764.41	4725.06
	检验期	23.08	38.46	3918.29	4343.92
最近邻抽样回归模型	率定期	48.94	68.09	2579.26	3458.74
	检验期	30.77	53.85	3131.12	3858.42
人工神经网络模型	拟合期	68.09	85.11	1544.72	2115.58
	检验期	38.46	53.85	3104.42	3677.44
支持向量机模型	率定期	80.85	89.36	1322.55	1929.64
	检验期	46.15	69.23	2808.87	3318.64
最优组合预报模型	率定期	91.49	91.49	685.69	1042.75
	检验期	46.15	61.54	2613.01	3158.93

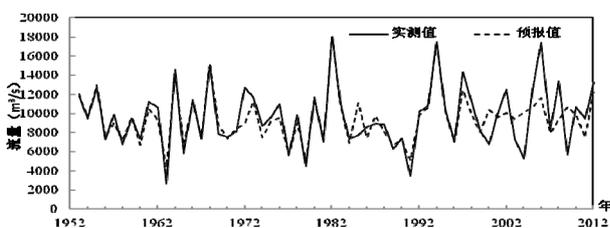


Figure 3. Comparisons of the optimal combination forecast model and the measured maximum flood inflow for the Feilaixia reservoir
图 3. 飞来峡水库入库最大洪峰流量最优组合预报模型预报值与其实测值对照图

向量机模型进行预报, 将所得结果进行最优组合预报, 所得主要结论如下:

1) 人工神经网络模型适用于年平均流量和最大洪峰流量预报, 其模型精度高, 可靠性强; 但数据量大时, 计算时间较长, 同时可能出现“过拟合”和局部极值现象, 导致精度降低。

2) 支持向量机模型适用性高, 因能以任意精度逼近任意函数, 有较好的泛化性能, 故模型精度高, 适合处理非线性和高维问题。

3) 最优组合预报模型同样适用于各种时间尺度的水文预报, 模型能综合各单一预报模型的有用信息模型和不确定性因素, 故结果稳定性好, 通用性强。

5. 致谢

感谢第十一届中国水论坛的推荐!

参考文献 (References)

- [1] 王雪玉. 中长期水文预报研究现状[J]. 黑龙江科技信息, 2012, 5: 37.
WANG Xueyu. Medium and long-term hydrological forecast research status. Heilongjiang Science and Technology Information, 2012, 5: 37. (in Chinese)
- [2] 商勇, 丁咏梅. 最优组合预测方法评析[J]. 统计与决策, 2005, 9: 122-123.
SHANG Yong, DING Yongmei. The elucidating of optimal combined forecasting method. Statistics and Decision, 2005, 9: 122-123. (in Chinese)
- [3] 丁晶, 邓育仁. 随机水文学[M]. 四川: 成都科技大学出版社, 1988.
DING Jing, DENG Yuren. Stochastic hydrology. Sichuan: Chengdu University of Science and Technology Press, 1988.
- [4] 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.
LI Jiren, LIU Deping. Hydrological time series model and prediction methods. Nanjing: Hohai University Press, 1991. (in Chinese)
- [5] 汤成友, 官学文, 张世明. 现代中长期水文预报方法及其应用[M]. 北京: 中国水利水电出版社, 2008.
TANG Chengyou, GUAN Xuewen and ZHANG Shiming. The advanced methods for mid-long term hydrological forecasting and its application. Beijing: China Water and Power Press, 2008. (in Chinese)
- [6] 朱聪. 径流中长期预测模型研究[D]. 四川: 四川大学, 2005.
ZHU Cong. Study on medium-and-long-term hydrologic prediction of runoff. Sichuan: Sichuan University, 2005. (in Chinese)
- [7] 黄鹤, 卢文喜, 伊燕平, 罗建男. 基于支持向量机模型的水质评价研究[J]. 节水灌溉, 2012, 2: 57-63. (in Chinese)
HUANG He, LU Wenxi, YI Yanping and LUO Jiannan. Assessment of water quality based on support vector machine mode. Water Saving Irrigation, 2012, 2: 57-63. (in Chinese)
- [8] Vapnik, V. The nature of statistical learning theory. New York: Springer, 1995.
- [9] YOON, H., JUN, S.-C., HYUN, Y.J., BAE, G.-O. and LEE, K.-K. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. Journal of Hydrology, 2011, 396(1-2): 128-138.
- [10] WANG, Y., GUO, S.L., CHEN, H. and ZHOU, Y.L. Comparative study of monthly inflow prediction methods for the Three Gorges Reservoir. Stochastic Environmental Research and Risk Assessment (Online).
- [11] 虞云飞. 飞来峡水库水文预报浅析[J]. 广西水利水电, 2002, 3: 36-39.
YU Yunfei. Simple analysis of the hydrologic prediction of the Feilaixia reservoir. Guangxi Water Resources and Hydropower Engineering, 2002, 3: 36-39. (in Chinese)