

# 分布式水文模型与气象遥相关分析相结合的丹江口 水库月入库径流预测

何小刚 赵铜铁钢 杨大文

(清华大学水利水电工程系 水沙科学与水利水电工程国家重点实验室 北京 100084)

**摘要:** 流域初始条件和气象条件是决定流域未来径流变化的重要因素,因此,确定符合实际情况的初始条件和输入可靠的未来降雨预测值是提高长期径流预测精度的有效途径。本研究将分布式水文模型和气象遥相关分析相结合,采用水文模型提供初始条件,利用遥相关关系确定气象输入,开展了汉江流域上游丹江口水库入库径流的长期预测研究。主要工作包括:①通过随机森林模型分析水文-气象遥相关关系用于挑选气象因子,根据气象因子的相似性从历史气象序列中筛选未来可能的降水序列;②以前期实测气象资料为输入,采用分布式水文模型确定流域初始条件,然后以筛选出的气象序列为输入,采用分布式水文模型对未来径流进行预测。结果表明,随机森林模型能帮助选择有效的气象因子用于筛选历史气象序列,分布式水文模型能够较好确定流域实时下垫面条件并充分利用筛选出来的气象序列进行集合预报,从而提高了长期径流预测的精度。

**关键词:** 水文学;长期径流预测;分布式水文模型;气象遥相关关系;随机森林模型;丹江口水库

中图分类号: P338<sup>+</sup>.9

文献标识码: A

## Prediction of monthly inflow to the Danjiangkou reservoir by distributed hydrological model and hydro-climatic teleconnections

HE Xiaogang, ZHAO Tongtiegang, YANG Dawen

(Tsinghua University, Department of Hydraulic Engineering, State Key Laboratory of  
Hydro-science and Engineering, Beijing 100084)

**Abstract:** Watershed initial conditions and future meteorological prediction are two important aspects that affect long-term streamflow prediction. This study combines a distributed hydrological model and hydro-climatic teleconnection analysis to predict the monthly streamflow of the Han River. We first applied a random forests model to analyze hydro-climatic teleconnections for selecting the dominant climate factors, and then selected future possible precipitation scenarios from the historical meteorological records based on similarity of the selected climate factors. The distributed hydrological model was used to obtain initial conditions and predict the long-term streamflow using the selected meteorological scenarios. Results show that this random forests model is effective for analyzing hydro-climatic teleconnections, and the distributed hydrological model could do rational prediction of long-term streamflow by using appropriate initial conditions and meteorological scenarios.

**Key words:** hydrology; long-term streamflow prediction; distributed hydrological model; hydro-climatic teleconnection; random forests model; the Danjiangkou reservoir

## 0 研究背景

汉江流域是南水北调中线工程水源区,流域内的丹江口水库是中线调水的源头工程。研究丹江口水库长期

收稿日期: 2011-09-29

基金项目: 水利部公益项目“南水北调中线水源区中长期径流预报技术研究”(201001004);“水资源优化配置数字化技术及示范研究”(201101004);水沙科学与水利水电工程国家重点实验室开放课题(2011-KY-2)

作者简介: 何小刚(1988-)男,本科生

通信作者: 杨大文(1966-)男,教授. E-mail: yangdw@tsinghua.edu.cn

入库径流预测对于中线工程的调度规划和运行管理等具有重要指导意义<sup>[1-3]</sup>。近年来,集合预测方法(也称集合预报方法)被广泛用于长期径流预测<sup>[4,5]</sup>,其基本原理是以流域当前土壤含水量和河道径流量作为初始状态(采用当年气象输入驱动水文模型得到),将降水、气温等历史序列作为未来气象输入(由历史序列代表未来气象条件),从而得到多组入流情景预测未来径流<sup>[4,6]</sup>。集合预测方法的优势在于可以充分利用流域的初始状态信息(土壤含水量、河道径流量、气象信息以及下垫面条件等),机理性更强。然而,传统集合预测对各历史年份气象条件赋予相同权重,从水文气象角度看,这种方法并不合理。例如,某年发生厄尔尼诺时,当年的气象条件应该与历史上厄尔尼诺年份的气象条件更为接近<sup>[7,8]</sup>。长期径流预测中,厄尔尼诺等气象要素与目标流域径流间的遥相关关系构成了径流预测的物理基础<sup>[2,3]</sup>。

本研究综合“传统集合预测方法能提供流域下垫面初始条件”和“长期径流预测方法能发掘水文气象遥相关关系”这两大特点,将分布式水文模型 GBHM(获取初始下垫面条件)和随机森林模型(分析水文气象遥相关关系)相结合,对汉江流域月径流进行预测研究。

## 1 基于分布式水文模型的传统集合预测

### 1.1 采用的数据及水文模型

本文基于分布式水文模型 GBHM(Geomorphology-Based Hydrological Model)构建汉江上游地区(丹江口水库以上,流域面积约 95200 平方公里)径流集合预测模型。汉江上游地区年降雨为 700~900 毫米,年陆地蒸发量为 500~600 毫米。海拔高程在 75~3556 米之间,地形起伏大,山区特征显著。在 GBHM 模型中,汉江上游流域被概化为“山坡-河沟”系统,从而反映下垫面条件空间变化对流域产汇流的影响<sup>[9-11]</sup>。

GBHM 建模需要地理和气象水文两方面的数据。地理数据中的数字高程数据和土地利用数据均来源于美国联邦地质调查局(USGS),空间分辨率分别为 100m 和 1km;植被数据来源于 GIMMS 卫星遥感影像,空间分辨率为 8km;土壤数据来源于国际粮农组织(FAO)全球土壤分类数据库,空间分辨率为 10km。气象水文数据主要包括气象数据、雨量数据以及流量数据。其中,气象数据主要用于计算流域蒸散发,其来源于汉江流域上游及其周边地区的约 30 个国家气象站点;模型降雨输入除了采用气象站实测降雨外,还采用了汉江上游 42 个雨量站数据。实测径流资料主要用于模型的率定和验证。上述气象水文资料均为日尺度的观测数据,时间跨度均为 1970-2000 年。

在气象遥相关分析过程中,本研究采用了国家气候中心提供的 74 项气象特征量,包括印缅槽强度、东亚槽强度、亚洲区极涡面积指数等<sup>[3]</sup>。受径流资料限制,本研究截取了 1970-2000 年间的水文气象数据。

### 1.2 基于历史气象系列的月径流集合预测

传统集合预测方法的适用性主要取决于水文模型模拟精度和流域初始状态<sup>[4]</sup>。本文评价了 GBHM 模型对汉江上游流域 5 个站点的模拟精度,如表 1 所示。从相对误差和 NS 系数来看,率定期、验证期模拟精度相似,相对误差在  $\pm 10\%$  以内,NS 系数在 0.8 左右。这说明 GBHM 模型对汉江流域上游的径流模拟是可靠的。

表 1 GBHM 模型在汉江上游的逐日径流模拟精度评价

Table 1 Evaluation of GBHM based streamflow prediction in the upper Han River

干支流 水文站	汇流面积/ km <sup>2</sup>	率定期(1981-1983)		验证期(1984-1986)	
		相对误差 RE/(%)	NS 系数 NSE	相对误差 RE/(%)	NS 系数 NSE
石泉	23805	8	0.79	1	0.86
安康	38625	-0.4	0.80	-0.2	0.82
白河	59115	7	0.74	8	0.74
黄家港	95217	4	0.85	6	0.84
荆紫关	7086	9	0.86	3	0.86

在此基础上,本研究采用集合预测方法<sup>[4,6]</sup>对汉江上游径流进行逐月滚动预测。预测中,月初土壤含水量和河道径流量(模型初始条件)由前期实测气象条件驱动水文模型进行动态更新,模型气象输入为 1971-2000 年同期气象条件(共 30 组历史气象序列)。

考虑到集合预测精度受初始条件和气象条件两方面的制约,本研究以白河站(丹江口水库入库径流控制站)为例,采用丛树铮教授介绍的回顾性检验方法<sup>[4]</sup>对初始条件、气象条件的影响进行评价。图 1 给出了白河站 1 月份(非汛期典型月份)和 8 月份(汛期典型月份)的回顾性集合预测结果评价。其中,横坐标 1~30 表示按照雨量

由少到多对 30 组气象条件进行重新排列的序数; 纵坐标 1~30 为按照流域湿润程度由干到湿对 30 组初始条件重新排列的序数; 图中条带则反映了径流量随初始条件、气象条件的变化情况。对比发现: 1 月份径流等值线平缓, 这说明非汛期气象条件对径流量影响很小, 径流主要取决于初始条件; 8 月份径流等值线陡峭, 这说明汛期的气象条件对流域径流起决定作用, 初始条件影响很小。

由图 1 可知, GBHM 模型能根据前期气象资料为径流预测提供有效的流域下垫面初始条件信息。对于非汛期径流, 由于其主要受初始条件影响, 采用 GBHM 模型和传统集合预测方法能达到良好的预测结果。对于汛期径流, 由于其受气象条件制约, 单纯根据历史气候资料给出的未来径流集合预测效果并不理想。本研究将在接下来的部分采用随机森林模型进行水文气象遥相关分析和降雨趋势预测, 从而对汛期径流预测进行改进。

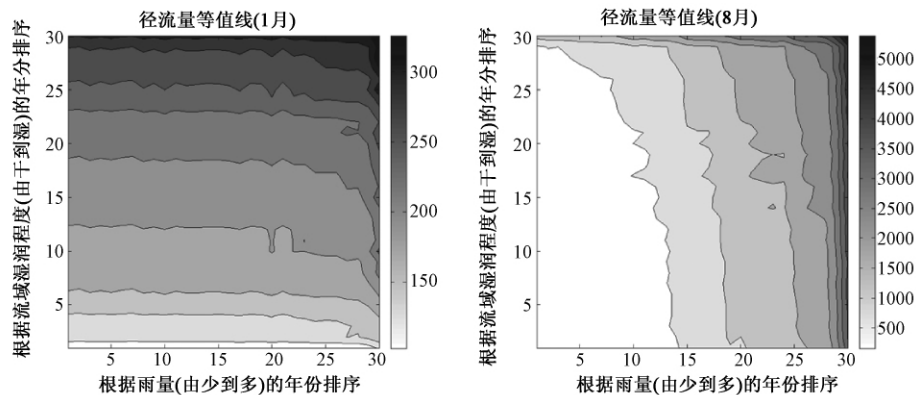


图 1 白河站不同月径流对初始条件和气象条件的敏感性比较(1月和8月)

Fig. 1 Sensitivity comparison of monthly runoff at the Baihe station to the initial condition and climate input (January and August)

## 2 基于气象遥相关分析对传统集合预测方法的改进

本研究将随机森林模型<sup>[12-14]</sup>用于发掘汉江流域径流与气象因素间的遥相关关系, 从而预测降雨趋势。随机森林模型(RF, Random Forests)<sup>[12]</sup>是一种数据挖掘模型, 其可以挖掘目标变量(径流)与解释变量(气象要素)间关联关系。随机森林模型采用自举法(Bootstrapping)随机选取多组训练样本集训练多个回归树模型(Regression Tree)进行预测, 从而抵消随机误差带来的过拟合对预测值的影响<sup>[12-14]</sup>。赵钢铁等(2011)将这个模型用于发掘水文-气象遥相关关系, 从而预测长江上游枯水期径流; 本研究采用随机森林模型分析未来降雨趋势, 从而筛选历史气象条件以改进传统集合预测方法。随机森林模型包含两个重要参数, 回归树节点划分待选变量数  $N$  (刻画子预测模型间的差异性) 和子预测模型数  $M$  (反映随机森林模型的过拟合效应)。 $N$  越大, 子预测模型越相似;  $M$  越大, 过拟合效应越小。根据相关文献<sup>[15]</sup>, 本文将  $N$  取为 250,  $M$  取为 2000。

随机森林模型建模的第一步是预测因子选择<sup>[3]</sup>。以 7 月份为例, 预测因子选取步骤为: 以 7 月份径流作为目标变量, 将过去一年 1-12 月份和当年 1-4 月份 74 项气象因素作为待选预测变量(共  $(12+4) \times 74 = 1184$  个); 对 1970-2000 年共 31 组“目标变量-待选预测变量”观测样本, 采用随机森林模型随机划分训练、检验样本集合, 评价待选预测因子对目标变量的重要性<sup>[3]</sup>。本研究选取前 5 个最重要的气象因素, 白河站汛期 5-10 月份径流的预测因子选取结果如表 2 所示。

本文选择的预测因子(见表 2)与冯小冲等针对同一流域的相关研究<sup>[16]</sup>所选预测因子有较好的一致性, 主要包括亚洲经向环流指数、北美区极涡强度指数、西太平洋副高北界等均被用作汉江流域夏季径流预测因子。

在表 2 所选预测因子基础上, 本研究引入模糊数学中欧几里得贴近度这一指标来刻画历史年份与目标年份气象条件的相似性, 从而进行降雨趋势预测。贴近度计算公式为:

$$N_H(A, B) = 1 - \sqrt{\frac{1}{n} \sum_{i=1}^n (A_i - B_i)^2}$$

式中:  $A$  和  $B$  表示 2 个样本;  $A_i$  和  $B_i$  为样本属性(即表 2 中气象要素);  $N_H(A, B)$  为样本  $A$  与  $B$  的贴近度。 $N_H(A, B)$  越大, 表示  $A, B$  样本的气象条件越相似。

根据  $N_H(A, B)$  判断的相似性, 即可筛选气象条件与目标年份最为相似的前  $N$  个年份, 从而构建改进的集合预报模型。本研究采用连续概率排位分数指标(Continuous Ranked Probability Scores, CRPS)评价  $N$  (集合成员数

目,也即筛选得到的相似年份的数目)对预测结果的影响。CRPS 的具体计算方法可以参考文献 [17],其值越小表示预测误差越小。

表 2 丹江口水库上游(白河站)汛期气象预测因子

Table 2 Predictors for the monthly streamflow at the Baihe station

预测月份	预测因子名称及对应月份
当年 5 月份	北非副高脊线(去年 3 月) 东亚槽位置(去年 12 月) 东亚槽强度(当年 2 月) 北非大西洋北美副高强度指数(去年 5 月) 亚洲径向环流指数(去年 4 月)
当年 6 月份	南海副高脊线(去年 8 月) 南海副高北界(去年 8 月) 北美区极涡面积指数(去年 12 月) 太平洋区极涡强度指数(去年 12 月) 太平洋区极涡面积指数(当年 2 月)
当年 7 月份	西太平洋副高脊线(去年 8 月) 太平洋副高北界(去年 8 月) 登陆台风(去年 8 月) 西太平洋副高北界(去年 8 月) 西藏高原(去年 8 月)
当年 8 月份	大西洋副高面积指数(当年 3 月) 大西洋副高强度指数(当年 3 月) 大西洋副高北界(当年 3 月) 北美副高强度指数(当年 3 月) 南海副高强度指数(去年 12 月)
当年 9 月份	太平洋区极涡面积指数(去年 11 月) 欧亚纬向环流指数(去年 2 月) 亚洲区极涡面积指数(当年 1 月) 西太平洋副高北界(去年 5 月) 北美区极涡面积指数(当年 1 月)
当年 10 月份	北美区极涡强度指数(去年 4 月) 亚洲纬向环流指数(当年 2 月) 大西洋欧洲环流型 E(去年 4 月) 大西洋欧洲环流型 C(去年 4 月) 大西洋欧洲区极涡面积指数(去年 6 月)

### 3 考虑未来降雨趋势的长期径流预测

本研究比较了汛期(5 月-10 月)各月径流预测中  $N$  对预测精度的影响,如图 2 所示。由图可知,相比于传统的集合预测方法(集合成员数目为 29 的情况),对历史气象系列进行筛选后,预测结果有一定程度的改善。并且,CRPS 值与集合成员数目大体呈现单调递增的关系,集合成员数目为 5 时,预测结果改善明显。可能的原因在于,在汛期,降雨的变异性更大,因而径流过程的变异性也更为明显,集合成员数目越多,误差可能也会越大。降

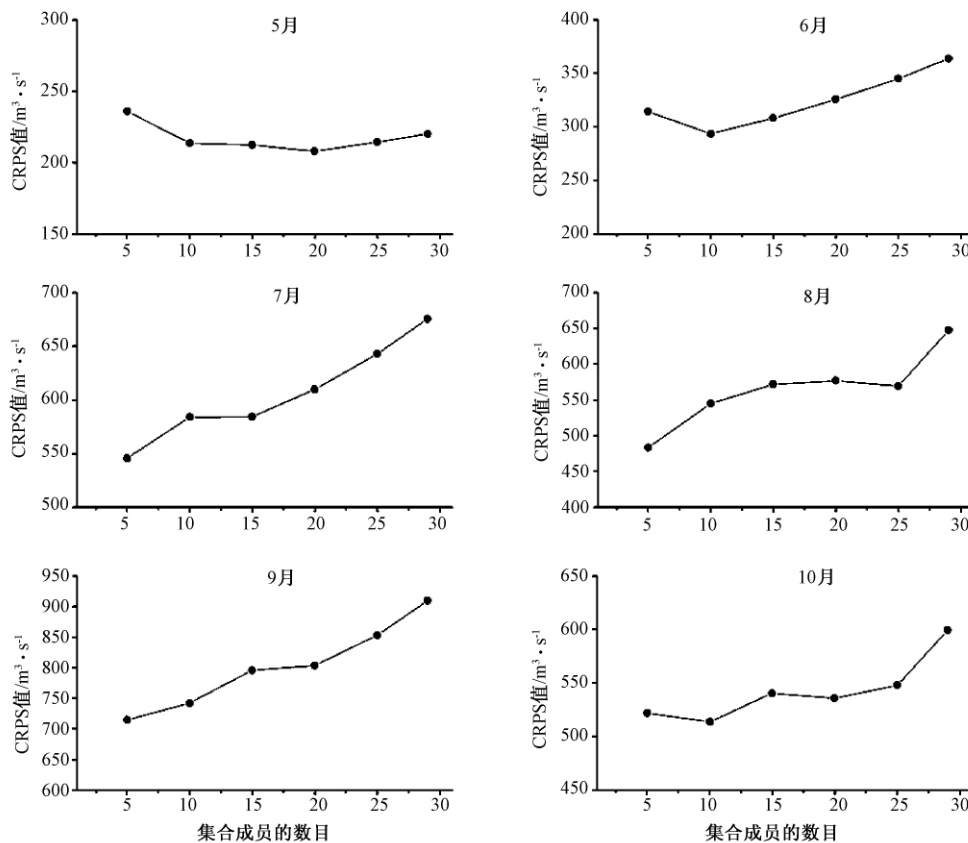


图 2 集合成员数目与 CRPS 间关系

Fig. 2 Relationship of the CRPS index vs. ensemble number

雨趋势的引入排除了非相似年份的干扰,因而预测结果更加准确。整体上,集成员数目为5时,预测精度较高。

在图2基础上,本研究将 $N$ 取为5,对白河站7、8月份的径流进行预测并与传统集合预测方法进行比较,结果如图3所示。其中,基于贴近度指标从历史气象系列中选取5年的数据作为水文模型的输入,以模拟得到的流量序列中最大、最小值作为集合预测的上、下限。对比图3中传统和改进后的集合预测方法的预测结果,可知:改进的集合预测方法均值与实测值更加接近,能够更好的反映径流丰枯趋势;此外,集成员数目减少使得径流预测不确定性大大降低,并且集合预测范围仍然能够包含大多数的实测样本。这说明采用水文气象遥相关关系能够有效改进传统集合预测。

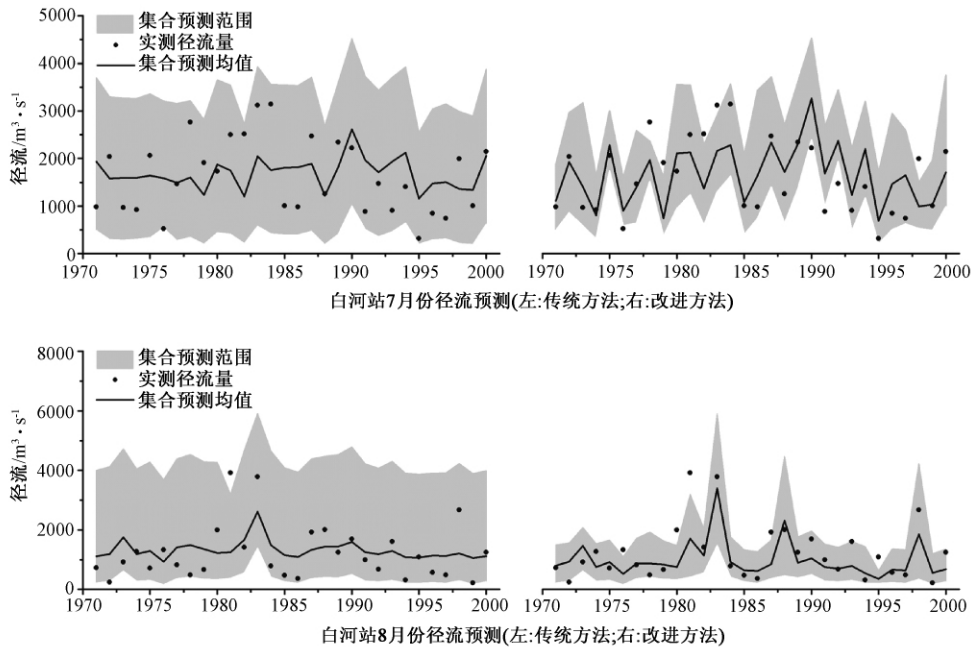


图3 传统集合预测方法与改进后的集合预测方法在白河站的预测结果

Fig. 3 Predictions of monthly streamflow at the Baihe station using conventional and improved extended streamflow prediction methods

## 4 总结与展望

采用合适的水文模型,利用前期气象数据模拟确定流域初始条件,根据过去的实测气象系列进行长期径流集合预测的方法,在非汛期具有较好的预测精度,但在汛期的预测不确定性很大。究其原因,非汛期的降雨变化相对较小,径流的变化主要取决于初始条件,因此传统集合预测方法在非汛期能够达到较好的预测结果;汛期的径流变化主要取决于气象条件,而且汛期的降雨变化较大,因此单靠改善初始条件的传统集合预测效果并不理想。针对汛期气象条件不确定性大的问题,本研究通过水文气象遥相关分析对未来降雨条件进行趋势预测,然后从历史降雨系列中挑选相似的降雨系列作为分布式水文模型的输入,这样大大降低了径流预测的不确定性,提高了长期径流预测精度。

值得指出的是,长期径流预测,尤其是汛期径流预测的最大不确定性来自于气象输入条件。从水文-气象遥相关关系入手,挑选出合适的气象因子进行降雨趋势预测是降低气象输入不确定性的有效途径。本文采用基于随机森林模型的水文-气象遥相关分析尝试了未来降雨的长期趋势预测,并取得了预期的结果。进一步提高长期径流预测精度的途径是,应用气象学相关研究成果,构建统计学模型或动力学模型,改进水文-气象遥相关分析方法,提高未来降雨预测的准确性。

### 参考文献:

- [1] 林剑艺,程春田. 支持向量机在中长期径流预报中的应用[J]. 水利学报, 2006, 37(6): 681-686.  
LIN Jianyi, CHENG Chuntian. Application of support vector machine method to long-term runoff forecast [J]. Journal of Hydraulic Engineering, 2006, 37(6): 681-686. (in Chinese)
- [2] 赵钢铁,杨大文,李明亮. 超越概率贝叶斯判别分析方法及其在中长期径流预报中的应用[J]. 水利学报, 2011, 42(6):

692-698.

ZHAO Tongtiengang, YANG Dawen, LI Mingliang. The exceedance probability method for mid-term and long-term streamflow prediction [J]. *Journal of Hydraulic Engineering*, 2011, 42(6): 692-698. (in Chinese)

- [3] 赵钢铁, 杨大文, 蔡喜明. 基于随机森林模型的长江上游枯水期径流预报研究[J]. *水力发电学报*, 2012 (3): 18-24 38.  
ZHAO Tongtiengang, YANG Dawen, CAI Ximing. Predicting seasonal low-flow in the upper Yangtze river using the random forests model [J]. *Journal of Hydroelectric Engineering*, 2012 (3): 18-24 38. (in Chinese)
- [4] 丛树铮. 水科学技术中的概率统计方法[M]. 北京: 科学出版社, 2010.  
CONG Shuzheng. Probability and statistical methods in water science and technology [M]. Beijing: Science Press, 2010. (in Chinese)
- [5] Cloke H L, Pappenberger F. Ensemble flood forecasting: A review [J]. *Journal of Hydrology*, 2009, 375(3-4): 613-626.
- [6] 李岩, 胡军, 王金星 等. 河流集合预报方法(ESP)在水资源中长期预测中的应用研究[J]. *水文*, 2008, 28(1): 25-27.  
LI Yan, HU Jun, WANG Jinxing, et al. Application of ensemble streamflow prediction (ESP) to medium-and-long-term water resources prediction [J]. *Journal of China Hydrology*, 2008, 28(1): 25-27. (in Chinese)
- [7] 刘永强, 丁一汇. ENSO事件对我国季节降水和温度的影响[J]. *大气科学*, 1995, 19(2): 200-208.  
LIU Yongqiang, DING Yihui. Reappraisal of the influence of ENSO events on seasonal precipitation and temperature in China [J]. *Scientia Atmospherica Sinica*, 1995, 19(2): 200-208. (in Chinese)
- [8] 王群英, 龚道溢. 华北降水资源的变化及其与厄尔尼诺的关系[J]. *自然资源学报*, 1999, 14(2): 103-108.  
WANG Qunying, GONG Daoyi. Variation of rainfall over north China and its possible connection to El Nino [J]. *Journal of Natural Resources*, 1999, 14(2): 103-108. (in Chinese)
- [9] Yang D W, Herath S, Musiak K. Development of a geomorphology-based hydrological model for large catchments [J]. *Annual Journal of Hydraulic Engineering, JSCE*, 1998, 42: 69-174.
- [10] 王磊, 王忠静, 尹航 等. GBHM模型原理及其在中尺度流域的应用[J]. *冰川冻土*, 2006, 28(2): 256-261.  
WANG Lei, WANG Zhongjing, YIN Hang, et al. A distributed hydrological model-GBHM and its application in middle-scale catchment [J]. *Journal of Glaciology and Geocryology*, 2006, 28(2): 256-261. (in Chinese)
- [11] 高冰, 杨大文, 刘志雨 等. 雅鲁藏布江流域的分布式水文模拟及径流变化分析[J]. *水文*, 2008, 28(3): 40-44.  
GAO Bing, YANG Dawen, LIU Zhiyu, et al. Application of a distributed hydrological model for the Yarlung Zangbo river and analysis of the river runoff [J]. *Journal of China Hydrology*, 2008, 28(3): 40-44. (in Chinese)
- [12] Breiman L. Random forests [J]. *Machine Learning*, 2001, 45(1): 5-32.
- [13] Peters J, Baets B D, Verhoest N E C, et al. Random forests as a tool for ecohydrological distribution modeling [J]. *Ecological Modelling*, 2007, 207(2-4): 304-318.
- [14] Carlisle D M, Falcone J, Wolock D M, et al. Predicting the natural flow regime: models for assessing hydrological alternation in streams [J]. *River Research and Applications*, 2010, 26(2): 118-136.
- [15] Liaw A, Wiener M. Classification and regression by random Forest [J]. *R News*, 2002, 2(3): 18-22.
- [16] 冯小冲, 王银堂, 刘勇 等. 基于物理统计方法的丹江口水库月入库径流预报[J]. *河海大学学报*, 2011, 39(3): 242-247.  
FENG Xiaochong, WANG Yintang, LIU Yong, et al. Monthly runoff forecast for Danjiangkou reservoir based on physical statistical methods [J]. *Journal of Hohai University*, 2011, 39(3): 242-247.
- [17] 李明亮, 杨大文, 陈劲松. 基于采样贝叶斯方法的洪水概率预报研究[J]. *水力发电学报*, 2011, 30(3): 27-33.  
LI Mingliang, YANG Dawen, CHEN Jinsong. Probabilistic flood forecasting by a sampling-based Bayesian model [J]. *Journal of Hydroelectric Engineering*, 2011, 30(3): 27-33. (in Chinese)