

Identification of Hydrologic Model Change-Point for Middle Yangtze River*

Pan Liu¹, Chao Deng¹, Honggang Zhang², Yan Shen¹, Shengxiang Ding²

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

²Hydrology Bureau of Changjiang Water Resources Commission, Wuhan

Email: liupan@whu.edu.cn

Received: May 13th, 2013; revised: Jun. 7th, 2013; accepted: Jun. 21st, 2013

Abstract: It is very important to identify the model change-point for scientific understanding of the hydrologic simulation. Based on the Bayesian change-point estimation of the general linear model, a method has been set up to identify the change-point of hydrologic systematic model, especially a multiple inputs—single output linear model. With the cases study of the middle Yangtze River, two models have been set up to simulate the flows from Yichang to Shashi and Luoshan stations, respectively. It is indicated that no significant change has taken place in the studied area.

Keywords: Systematic Model; Change-Point of Model; Multiple Input Single Output; Bayesian

长江中游水文系统模型变异诊断研究*

刘攀¹, 邓超¹, 张洪刚², 沈艳¹, 丁胜祥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长江水利委员会水文局, 武汉

Email: liupan@whu.edu.cn

收稿日期: 2013年5月13日; 修回日期: 2013年6月7日; 录用日期: 2013年6月21日

摘要: 水文模型结构变异分析可为水文模拟提供科学依据。针对多输入单输出线性系统模型, 基于线性模型的贝叶斯变点估计方法, 诊断和识别水文系统中的模型结构变异。以长江中游水文系统为研究背景, 分别建立了宜昌-沙市、宜昌-螺山两个多输入单输出线性系统模型, 结果表明, 长江中游水文模型变异不显著, 模拟中可作为一致性序列处理。

关键词: 系统模型; 模型变异; 多输入单输出; 贝叶斯

1. 引言

变化环境下的水文模拟是当今水文科学热点、难点问题。受气候变化、人类活动影响, 水文序列的平稳性假定受到质疑和挑战。Merz 等指出, 模型参数和模型结构并不是一成不变的^[1]。此时, 如何识别模型

结构的时间变异点, 并分析变异的物理成因, 具有重要的理论与实践意义。

多输入单输出线性系统模型将流域、河段或者水库视为系统的输入或者输出单元, 通过数学工具映射输入输出之间的数学关系, 是应用广泛的水文模型, 适应于成因复杂的水文水资源模拟^[2,3]。本文基于多输入单输出系统模型, 采用线性模型的贝叶斯变点估计方法, 开展模型变异点诊断和识别研究。以长江中游

*基金项目: 教育部新世纪优秀人才支持计划(NCET-11-0401); 国家自然科学基金项目(51190094; 51249008)。

作者简介: 刘攀(1978-), 男(汉), 湖南湘潭人, 博士, 武汉大学教授, 主要研究方向为: 水文及水资源开发利用。

为例, 探讨模型结构变异问题。

2 线性模型的贝叶斯变点估计方法

多输入单输出线性系统模型可表述为^[2,3]: 设某一输出时间序列可用多个输入的线性模型表示, 写成矩阵形式为:

$$y = Gb + \varepsilon \quad (1)$$

式中: y 是一 N 维向量, 为自变量; G 是一 $N \times M$ 矩阵, 为因变量; b 是线性模型的参数向量, ε 是残差向量; N 为时间序列长度; M 为输入变量个数。

在估计线性模型的变点位置时, 可采用贝叶斯方法^[4-6]: 由残差向量服从正态分布 $N(0, \sigma^2)$ 且独立, 可写出似然函数如下:

$$L(\Phi|y) = p(y|\Phi) = p(\varepsilon|\Phi) = (2\pi\sigma^2)^{-N/2} \exp\left[-\frac{(y-Gb)^T(y-Gb)}{2\sigma^2}\right] \quad (2)$$

式中: $\Phi = [\sigma, b, \{\omega\}]$, ω 为变化点的位置。假定 b 、 ω 以及 $\log\sigma$ 服从均匀分布, 则可取 Φ 的先验分布为:

$$p(\Phi) \propto \frac{1}{\sigma} \quad (3)$$

由贝叶斯公式

$$p(\Phi|y) \propto L(\Phi|y)p(\Phi) = (2\pi\sigma^2)^{-N/2} \exp\left[-\frac{(y-Gb)^T(y-Gb)}{2\sigma^2}\right] \frac{1}{\sigma} \quad (4)$$

对参数 b 和 ε 进行积分, 最终可得到 ω 的概率是

$$p(\{\omega\}|y) \propto \frac{\left[y^T y - y^T G(G^T G)^{-1} G^T y\right]^{-(N-M)/2}}{|G^T G|^{1/2}} \quad (5)$$

为了计算标准化以后的任意一点可能是变点的概率, 可采用

$$p(\{\omega_j\}|y) = \frac{p(\{\omega_j\}|y)}{\sum_{i=1}^N p(\{\omega_i\}|y)} \quad (6)$$

在利用上面的通用模型进行变点分析的时候, 可通过构造不同的矩阵 G , 用来分析不同的线性模型。

在多输入单输出系统模型中, 如果模型在 ω 点发生变化, 则:

$$y_i = \begin{cases} h_1 x_i + h_2 x_{i-1} + \dots + h_m x_{i-m+1}, & i \leq \omega \\ h'_1 x_i + h'_2 x_{i-1} + \dots + h'_m x_{i-m+1}, & i > \omega \end{cases} \quad (7)$$

为了应用上述变点模型, 相应的可取(公式(9)见下):

$$b^T = [h_m, h_{m-1}, \dots, h_1, h'_m, h'_{m-1}, \dots, h'_1] \quad (8)$$

$$y^T = [y_1, y_2, \dots, y_N] \quad (10)$$

这样, 可以采用式(5)和(6)来推求系统模型的各点发生变化的概率, 取概率最大的为最终的模型变点估计 ω 。贝叶斯方法对于变点位置发生在两端的情况, 需要分析合理性; 并且理论上得到的变点位置, 仍需要从物理成因上进行分析其合理性。

3. 研究实例

选取长江中游干流的宜昌、沙市、螺山 3 个控制站、主要支流包括清江长阳站、沮漳河河榕站、荆南三河——松滋河、虎渡河、藕池河控制站, 以及洞庭湖水系四水——湘江、资水、沅水和澧水控制站的实测日平均流量资料, 宜螺区间水系概化图如图 1 所示。

采用算术平均得到宜昌 - 沙市区间、宜昌 - 螺山区间的日平均降水量。宜昌 - 沙市区间的流量和降雨资料序列为 1952 年 1 月 1 日~2010 年 12 月 31 日, 沙市站的观测流量资料序列为 1991 年 1 月 1 日~2010 年 12 月 31 日; 宜昌 - 螺山区间的流量和降水资料的资料序列为 1955 年 1 月 1 日~2010 年 12 月 31 日, 其中

$$G^T = \begin{bmatrix} x_1 & x_2 & \dots & x_{\omega-m+1} & 0 & 0 & \dots & 0 \\ x_2 & x_3 & \dots & x_{\omega-m+2} & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \dots & \vdots \\ x_m & x_{m+1} & \dots & x_\omega & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & x_{\omega-m+2} & x_{\omega-m+3} & \dots & x_{N-m+1} \\ 0 & 0 & \dots & 0 & x_{\omega-m+3} & x_{\omega-m+4} & \dots & x_{N-m+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & x_{\omega+1} & x_{\omega+2} & \dots & x_N \end{bmatrix} \quad (9)$$

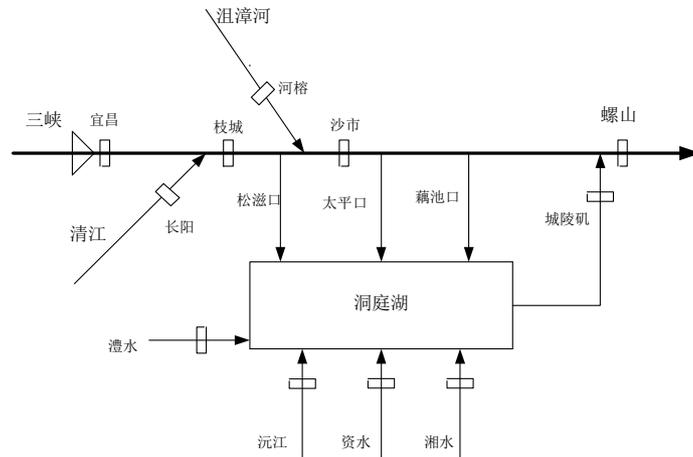


Figure 1. The sketch of the Middle Yantze River
图 1. 长江中游宜昌至螺山河段水系概化图

湘潭站缺 1987 年非汛期(1 月 1 日~4 月 30 日, 11 月 1 日~12 月 31 日)的流量数据。根据我国《水文情报预报规范》(GB/T 22482-2008)^[7], 计算采用模型效率系数 R^2 (确定性系数)和径流总量相对误差 RE 进行精度评定。

3.1. 宜昌 - 沙市水文模拟

3.1.1. 系统模型

宜昌至沙市的河道通畅, 流量、雨量资料齐全, 除干流流量以外, 主要有清江(控制站为长阳站)和沮漳河(控制站为河榕站)两大支流汇入, 如图 1 所示, 区间面积约为 26532 km²。根据《长江流域洪水预报方案汇编》, 以 0.361 倍的长阳站流量作为河榕站的流量过程, 同时, 多输入单输出模型要求各输出之间不能存在自相关关系, 因此将长阳、河榕合并为一个输入。模型的输入序列包括宜昌站(实测)、长阳站 + 河榕站、区间降雨(宜昌至沙市); 输出为沙市站的径流过程。

采用最小二乘法识别脉冲响应函数, 记忆长度参考洪水传播时间并采用试算法求得。模型率定期为 1991~2002 年, 检验期为 2003~2010 年, 各输入的记忆长度分别为 2、2 和 3, 对应的增益因子分别为 0.8713、0.6567 和 0.0448。模型的评价结果如表 1 所示, 模型效率系数在率定期和检验期分别到达 98.42% 和 97.89% 以上, 径流总量相对误差均不超过 5%。

3.1.2. 变异诊断

采用上述线性模型的贝叶斯变点估计方法, 对宜昌 - 沙市系统模型进行变异性分析, 得到宜昌 - 沙市

系统模型变异点概率分布如图 2 所示。变异时间发生在 1996 年 5 月 24 日, 与三峡大江截流时间 1997 年 11 月相对应, 可能为变异点。

根据这一变异点, 将 1991~2010 年的时间序列分为两段, 分别采用系统模型进行模拟计算, 参数识别结果如图 3 所示。可以看出, 经变异分割后, 分割前后的参数变化不大; 进一步采用模型对沙市进行径流模拟(表 2), 在增加变异点后, 模拟确定性系数由 98.29% 提高到 98.33%, 变化并不显著, 由于无显著物理成因, 可认为宜昌 - 沙市系统模型无需考虑模型变异。

为进一步论证宜昌 - 沙市系统模型无变异的结论, 将 1991~2010 年数据分为 4 段: 1991~1995、1996~2000、2001~2005 以及 2006~2010, 分别采用多输入单输出线性系统模型率定参数, 得到结果如图 4 所示。发现模型参数变异性不大, 支持了宜昌 - 沙市系统模型无变异的结论。

Table 1. The measurements for the Middle Yantze River systematic models
表 1. 长江中游系统模型精度评定

模型方案	时段	效率系数 R^2 (%)	径流总量相对误差 RE (%)
宜昌 - 沙市模型	率定期(1991~2002)	98.42	-3.76
	检验期(2003~2010)	97.89	-4.76
考虑“三口”宜昌 - 螺山模型	率定期(1955~1990)	99.20	-0.37
	检验期(1991~2010)	99.23	-0.13
不考虑“三口”宜昌 - 螺山模型	率定期(1955~1990)	90.95	-3.33
	检验期(1991~2010)	92.02	-3.96

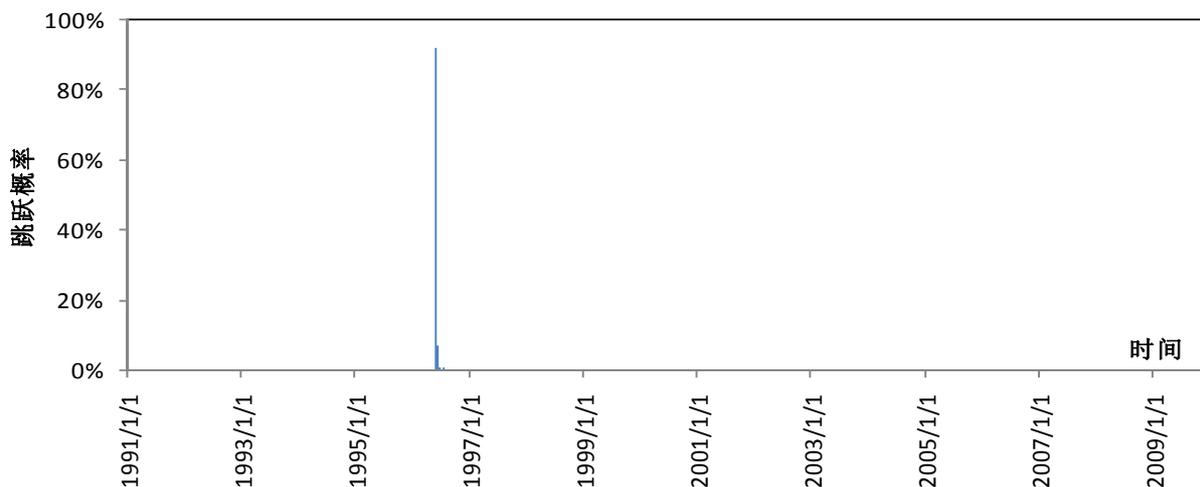


Figure 2. Change-point probability of the hydrologic model from Yichang to Shashi stations
图 2. 宜昌 - 沙市系统模型变异点概率分布图

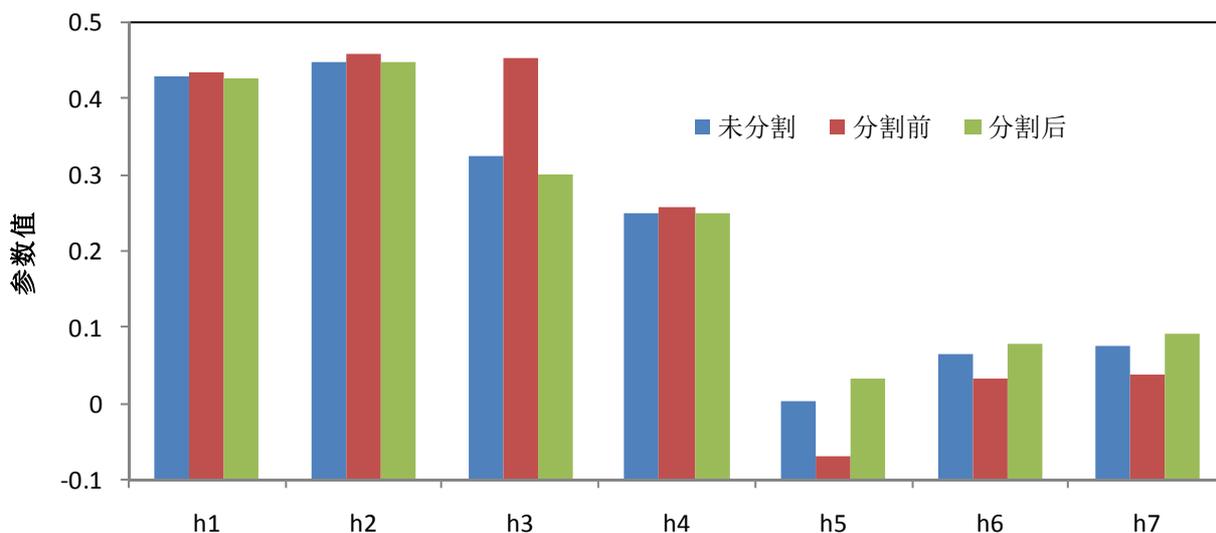


Figure 3. Comparison of the parameters with/without change-point for the hydrologic model from Yichang to Shashi stations
图 3. 宜昌 - 沙市系统模型参数结果比较图

Table 2. The measurements for systematic models with/without change-point
表 2. 考虑变异前后模型效率系数 R^2 (%)

模型	不考虑变异	考虑变异
宜昌 - 沙市	98.29	98.33
宜昌 - 螺山(考虑“三口”分流)	99.22	99.23
宜昌 - 螺山(不考虑“三口”分流)	91.23	91.31

3.2. 宜昌 - 螺山水文模拟

3.2.1. 系统模型

由于洞庭湖流域与长江干流存在着复杂的河湖关系, 在雨量及流量资料的处理上存在一定的困难。采用两种演进方案:

1) 考虑“三口”分流。即在宜昌站的径流量中减去松滋口和藕池口的流量, 在城陵矶的流量中减去太平口的分流; 孙晓红等曾对比了 2 种方案^[5]: a) 在宜昌站中减去松滋口和藕池口的流量, 城陵矶中减去太平口的流量; b) 宜昌站中减去松滋口的流量, 城陵

矾中减去藕池口和太平口的流量。经过反复计算分析比较, 方案 a) 的精度优于方案 b), 因此本文采用方案 a)。考虑洞庭湖“三口”分流的影响, 各输入为: 宜昌 1(宜昌 - 松滋口 - 藕池口)、长阳 + 河榕、城陵矾 1(城陵矾 - 太平口)、区间降雨(宜昌至螺山), 输出为螺山站的径流过程。

2) 不考虑“三口”分流。即将洞庭四水和湖区降雨作为区间入流, 当作模型的独立输入量。输入序列包括宜昌、长阳 + 河榕、桃江、桃源、石门、湘潭、区间降雨(宜昌至螺山), 输出为螺山站的径流过程。

模型率定期为 1955~1990 年, 检验期为 1991~2010 年, 结果见表 1 和图 5。

3.2.2. 变异诊断

同样采用线性模型的变点估计方法, 对宜昌~螺山系统模型开展变异分析, 得到考虑“三口”分流和不考虑“三口”分流两种情形下的变异点概率分布如图 6 所示。变异时间分别发生在 2006 年 4 月 18 日和 1959 年 2 月 19 日, 均无显著物理成因, 且发生在序列的两端, 不考虑为变异点。

根据这一变异点, 将 1955~2010 年的时间序列分为两段, 分别采用系统模型进行模拟计算。如表 2 所示, 在增加变异点后, 模拟确定性系数变化并不显著, 为简单起见, 可认为宜昌 - 螺山系统模型无需考虑模型变异。

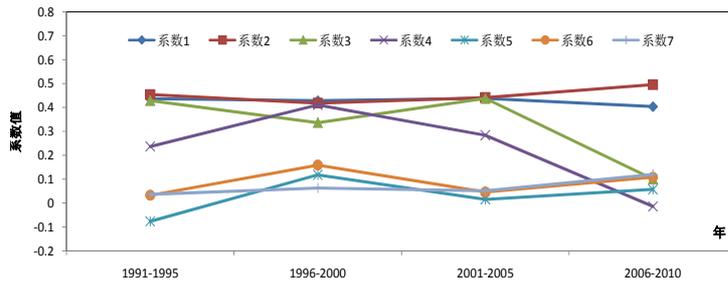


Figure 4. Parameters with separated time periods for the hydrologic model from Yichang to Shashi stations
图 4. 宜昌 - 沙市分段率定模型参数结果图

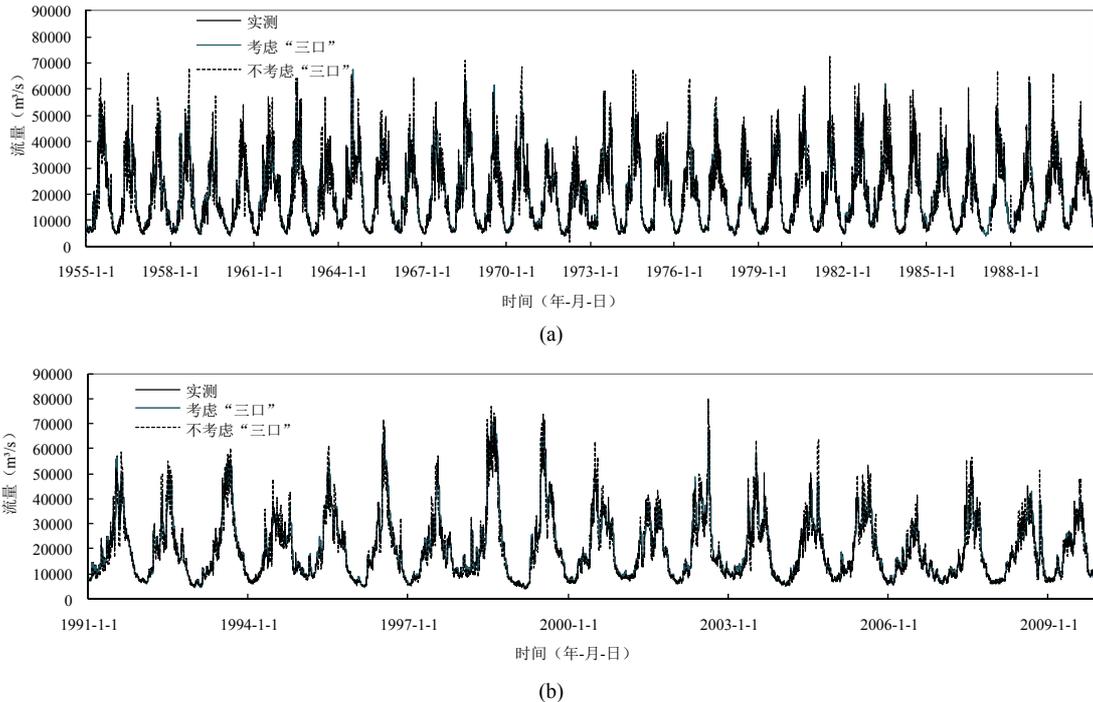


Figure 5. Comparison of observed and simulated flow hydrographs; (a) 1955-1990 years; (b) 1991-2010 years
图 5. 宜昌至螺山区间模拟与实测过程; (a) 1955~1990 年; (b) 1991~2010 年



Figure 6. Probability of the change-point model from Yichang to Luoshan stations

图 6. 宜昌 - 螺山系统模型变异点概率分布图

4. 结语

基于线性模型的贝叶斯变点估计方法, 诊断线性系统模型变异, 提出了识别水文系统中模型结构变异的方法。以长江中游水文系统为研究背景, 分别建立了宜昌 - 沙市、宜昌 - 螺山多输入单输出线性系统模型, 结果表明, 长江中游水文模型变异不显著, 模拟中可作为一致性序列处理。

参考文献 (References)

- [1] MERZ, R., PARAJKA, J. and BLOÖSCHL, G. Time stability of catchment model parameters: Implications for climate impact analyses. *Water Resources Research*, 2011, 47(2): W02531.
- [2] 孙晓红, 郭生练, 郭海晋, 徐高洪, 陈桂亚. 长江洞庭湖和宜昌至汉口流域洪水过程模拟研究[J]. *水电能源科学*, 2001, 19(3): 46-48.
SUN Xiaohong, GUO Shenglian, GUO Haijing and CHEN

- Guiya. Flood simulation study in the Dongtinghu and Yichang-Hankou basins of the Yangtze River. *International Journal Hydroelectric Energy*, 2001, 19(3): 46-48.
- [3] 沈艳, 刘攀, 刘玥. 三峡水库调蓄后长江中游径流演变趋势分析[J]. *人民长江*, 2013, 44(4): 25-27.
SHEN Yan, LIU Pan and LIU Yue. Analysis on runoff evolution trend of middle Yangtze River after impoundment of Three Gorges Reservoir. *Yangtze River*, 2013, 44(4): 25-27.
- [4] RASMUSSEN, P. Bayesian estimation of change points using the general linear model. *Water Resources Research*, 2001, 37(11): 2723-2731.
- [5] SEIDOU, O., ASSELIN, J. J. and OUARDA, T. B. M. J. Bayesian multivariate linear regression with application to change point models in hydrometeorological variables. *Water Resources Research*, 2007, 43(8).
- [6] BEAULIEU, C., SEIDOU, O., OUARDA, T. B. M. J. and ZHANG, X. Intercomparison of homogenization techniques for precipitation data continued: Comparison of two recent Bayesian change point models. *Water Resources Research*, 2009, 45(8):
- [7] GB/T 22482-2008, 水文情报预报规范[S]. 北京: 中国标准出版社, 2009.
GB/T 22482-2008, The hydrological forecasting standard. Beijing: China Standard Press, 2009.