

Scoring of Multiple Integrated Forecast Methods in Area Rainfall Forecast of Chaohu Lake Basin

Wenben Wang, Shejun Jin, Hanwu Liu

Chaohu Lake Meteorological Service, Hefei Anhui
Email: wangwenben_1@163.com

Received: Jun. 27th, 2019; accepted: Jul. 17th, 2019; published: Jul. 29th, 2019

Abstract

Based on the European Centre (ECMWF), Japan Meteorological Agency (JMA), WRF precipitation forecast products and observed data from June to August of 2014 and 2015 around Chaohu Lake basin, using the multiple linear regression (MLR) and the Principle Component Regression methods (PCR), 6 sub-units integrated forecast equations (CF) of Chaohu Lake basin were established respectively. Data from June to August of 2016, Precipitation Correctness Test, Average absolute error and TS score were used to test the effect of three single modes and two integrated forecasting equations. The following conclusions were drawn: 1) Comparing the effect of integrated forecast with single model forecast, the score of integrated forecast is higher than single model. On the level of light and moderate rain, two integrated methods have higher advantages than the single model of 24 h, 48 h, 72 h; on the level of heavy rainfall, two integrated methods have a slight advantage of 24 h, 48 h. 2) Compared with the two integrated forecast methods, the PCR method scored higher than MLR, and was more stable. 3) With the prolongation of forecasting time, the score of single model was significantly lower than the integrated forecast model.

Keywords

Area Rainfall, Integrated Forecast Methods, Chaohu Lake Basin

多种集成方法在巢湖流域面雨量预报中的效果检验

王文本, 金社军, 刘汉武

安徽省巢湖气象局, 安徽 合肥
Email: wangwenben_1@163.com

收稿日期: 2019年6月27日; 录用日期: 2019年7月17日; 发布日期: 2019年7月29日

摘要

利用2014年和2015年6~8月欧洲中心(ECMWF)、日本气象厅(JMA)和中小尺度天气预报模式WRF不同时效预

作者简介: 王文本, 男, 工程师, 主要从事数值预报方面研究工作。

报产品,结合巢湖流域对应的实况降水资料。采用多元线性回归方法(MLR)和主成分回归方法(PCR),分别建立巢湖流域6个子单元面雨量集成预报方程(CF)。并利用2016年6~8月实况降水数据,通过正确率、平均绝对误差和TS评分几种检验方法对3个单模式和两种集成预报方程在不同子流域不同时效面雨量预报效果进行对比检验。有以下结论:1)集成预报与单模式预报效果比较,集成预报较单一模式预报评分更高。小雨和中雨量级上,24 h、48 h、72 h时效两种集成方法较单一模式有明显优势;大雨量级上,24 h、48 h时效两种集成方法较单一模式略有优势;暴雨量级上,48 h、72 h时效两种集成方法较单一模式有明显优势。2)比较两种集成方法预报效果,PCR方法在不同量级不同时效的面雨量预报中较MLR方法得分更高。3)随预报时效(24 h、48 h、72 h)延长,各模式对不同量级面雨量预报效果逐渐下降,但单一模式预报效果较集成预报效果下降更明显。

关键词

面雨量,集成预报,巢湖流域

Copyright © 2019 by author(s) and Wuhan University.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

面雨量预报是洪水预报和河湖生态补偿非常重要的参考对象,是一定范围区域内面积上的平均降水量,能较客观地反映整个区域降水空间和时间分布情况[1][2]。各流域面雨量预报产品是各级政府制定防汛抗洪、水库调度和水体生态补偿决策方案的重要依据。因此在面临汛期防汛任务严峻形势下,水污染防治环保政策压力下,怎样提高各个子流域面雨量的预报水平[3],为政府各类决策气象服务提供科学依据,成为业务和科研人员必须面对的技术难题。常规面雨量的预报方法主要有天气学方法、数值预报模式降水输出客观化的面雨量产品方法及以数值预报产品为基础的多模式集成预报方法[4][5]。

集成预报是从多种预报结果中获得一个较为确定的预报结论的方法,该方法将多种数值预报模式的预报结果进行集成,以提高预报准确率为目的。集成预报的方法有很多种,总数学模型为输入-输出的函数关系。将其应用于面雨量预报,一定程度上克服了单模式的不稳定性、随机性带来的偏差,并体现了每种单模式在不同强度和不同预报时效上的相对优势。随着计算机性能增强、预报模式的增加,基于大数据的神经网络方法、动态变权集成方法能更好地建立集成预报模型。如盛飞等将主成分分析方法引入到用神经网络作集成预报中[6],根据各个主成分的方差贡献的大小来确定所需主成分的个数,对集成预报的原始因子进行重构,来解决样本间的共线性问题,对方程降维拟合分析,结合神经网络学习算法,达到加快计算速度和提高预报准确率的目的。严明良等探讨了几种基于超级集合思想的多模式数值预报动态变权集成处理方法[7],根据多个模式在不同时间段内的预报能力进行动态加权,能动态反映各类单模式的预报能力变化,在一定程度上提高了不同时间、不同区域的精细化预报水平和数值产品的利用效率。在样本数量不足的情况下,叶金印给出了用信息矩阵方法进行面雨量预报[8],克服了降水集成预报常常面临样本序列时间短问题,为模式数量较少的业务平台提供了一种集成方法。

结合以上各类集成算法相对优势,考虑到现实中地市级单位无法提供多种类的预报模式,受样本数量的限制,结合本地业务性质,我们选取了多元回归和主成分回归算法,并以巢湖流域面雨量为预报对象,进行集成应用试验。采用气象部门常用的降水分级检验评分(TS)统计评价指标、预报正确率指标和平均绝对误差指标对集成预报产品进行评分验证,得出了这两类集成预报在一定程度上提高了面雨量预报的准确率。

2. 资料与方法

2.1. 资料说明

2.1.1. 巢湖流域子单元划分

巢湖流域位于安徽省中部，长江下游左岸水系。流域面积为 1.35 万 km²。流域内山区面积 3123 km²，丘陵区面积 6280 km²，平原水网区面积 2729 km²，巢湖、黄陂湖等水面面积共计 808 km²，出入河流 33 条。根据安徽省水文局资料将其分为 6 个子单元，如图 1 分别为：1) 店埠河南淝河派河子单元；2) 杭埠河丰乐河子单元；3) 马尾河西河上游子单元；4) 西河下游清溪裕溪河子单元；5) 柘皋河子单元；6) 巢湖湖面子单元。

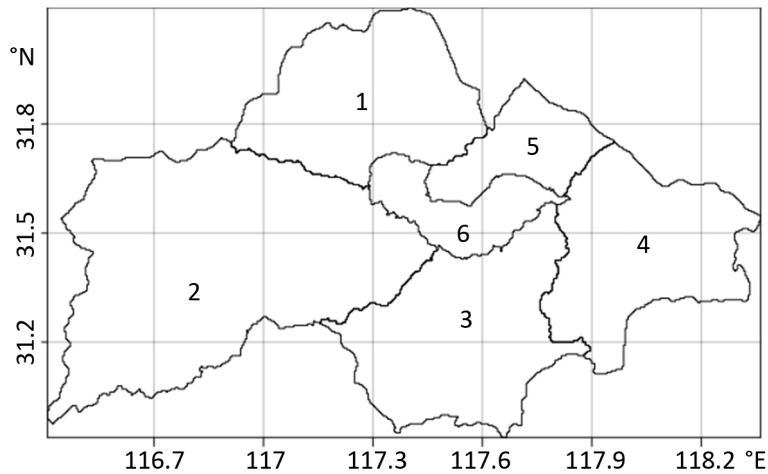


Figure 1. Distribution map of 6 sub-units in Chaohu Lake basin
图 1. 巢湖流域 6 个子单元划分示意图

2.1.2. 降水资料

利用 2014 年和 2015 年 6~8 月份巢湖流域内自动雨量站的日降水资料。及其对应时段 ECMWF、JMA、WRF 数值预格点场数据，预报时效为 24 h、48 h、72 h 降水预报产品。

2.2. 面雨量的计算方法

2.2.1. 面雨量定义

面雨量是指一次降雨过程中，整个流域面上的平均降雨量。可表示为(1)式。

$$\bar{P} = \frac{1}{A} \int_A P dA \quad (1)$$

式中： \bar{P} 为面雨量， A 为特定区域的面积， P 为有限元 dA 上的雨量。最常用的计算方法有天气学方法、网格插值法、等雨量线法、算术平均法、泰森多边形法等。

2.2.2. 实况及模式面雨量计算方法

在流域实况面雨量计算方法上，采用泰森多边形法。泰森多边形法又叫垂直平分法或加权平均法。面雨量表达式为(2)：

$$\bar{P} = f_1 P_1 + f_2 P_2 + \dots + f_n P_n \quad (2)$$

式中： f_1, f_2, \dots, f_n 分别为各雨量站多边形面积计算的权重系数； $P_1, P_2, \dots, P_n, \bar{P}$ 分别为各测站同时期降雨量和流域平均雨量。

模式面雨量计算方法。ECMWF 和 WRF 模式资料精度较高, 面雨量方法采用平均值法。JMA 产品精度相对较低, 分辨率为 $0.5^\circ \times 0.5^\circ$, 因此选取各子流域相邻格点资料, 然后采用等权重法输出面雨量。

2.2.3. 面雨量的预报方法

所用的面雨量集成预报方法有多元线性回归(MLR)和主成分回归(PCR)两种方法。多元线性回归就是对某一预报量 y , 研究多个因子与它的定量统计关系。假定预报量 y 与 p 个因子的关系是线性的, 为研究它们之间的联系作 n 次抽样, 每一次抽样可能发生的预报量之值为 y_1, y_2, \dots, y_n , 它们是 n 个随机变量。而第 i 次观测的因子值记为 $x_{i1}, x_{i2}, \dots, x_{ip}$ ($i=1, 2, \dots, n$)。因此多元线性回归方程如(3)式:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + e_i \quad (3)$$

式中: $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ 为 $p+1$ 个待估计参数, x_1, x_2, \dots, x_p 是 p 个一般变量, 设 e_1, e_2, \dots, e_p 是 n 个相互独立的且遵从同一正态分布的随机变量。

主成分回归(PCR), 是在自变量和因变量中各自找到一些互相独立的主成分, 然后按照计算主成分时得到的特征值大小来选取部分主成分, 主成分是独立的, 用这些主成分代替原来的变量进行回归。通过以下步骤实现, 首先利用主成分分析方法将多个变量通过线性组合求取独立变量; 其次找出前几个主成分的方差贡献, 方差贡献超过 80%即可作为新的变量; 最后利用因变量和新的自变量进行回归分析。通过主成分贡献进一步确定预报因子的方法可以减少集成方程的不稳定性, 使得预报方程更加稳定。

2.3. 面雨量预报检验方法

2.3.1. 面雨量等级划分

采用 2006 年中国气象局发布的《江河流域面雨量等级》国家标准, 按 24 h 的面雨量值划分为小雨(0.1~5.9 mm)、中雨(6.0~14.9 mm)、大雨(15.0~29.9 mm)、暴雨(30.0~59.9 mm)、大暴雨(60.0~150.0 mm)和特大暴雨(>150.0 mm) 6 个等级[1] [2]。

2.3.2. TS 评分

参考 2005 年中国气象局《中短期天气预报质量检验办法(试行)》提供的方法[3], 检验面雨量预报效果采用如下公式。

$$TS = \frac{NA_k}{NA_k + NB_k + NC_k} \times 100\% \quad (4)$$

式中: NA 为等级降水正确(即预报等级与实况等级相同)的天数, NB 为空报站天数(即预报等级大于实况等级, 记录为等级空报), NC 为漏报站天数(即预报等级小雨实况等级, 记录为实况等级漏报), 其中 k 为面雨量降水等级。

2.3.3. 正确率检验(PC)

$$PC = \frac{NA + ND}{NA + NB + NC + ND} \times 100\% \quad (5)$$

式中: ND 为无降水预报正确的站天数。

2.3.4. 平均绝对误差(AE)

指预报值和实况值的平均绝对误差, 其计算式为:

$$Ea = \frac{1}{n} \sum_{i=1}^n |R_f - R_0| \quad (6)$$

式中： n 为有雨预报正确的天数， R_f 为有雨且预报正确时的面雨量预报值， R_0 为有雨且预报正确时的面雨量实况值。本文仅统计实况有雨且预报也有雨时的误差。

3. 面雨量预报方程建立与预报效果检验

3.1. 利用多元线性回归和主成分回归建立方程

利用 2014 年和 2015 年 6~8 月巢湖流域逐日降水资料，以及 2014 年和 2015 年 6~8 月 ECMWF、JMA 和 WRF 数值预报模式 24 h、48 h 和 72 h 面雨量降水资料。首先利用多元线性回归建立面雨量预报的回归方程，并对各个因子进行 F 检验，建立具体的多元线性回归方程组，方程如式(7)。

$$CF_{k,h} = a * EC_{k,h} + b * JMA_{k,h} + c * WRF_{k,h} \quad (7)$$

式中： $CF_{k,h}$ 为第 k 个流域、预报时效为 h 的集成预报值， $EC_{k,h}$ 、 $JMA_{k,h}$ 、 $WRF_{k,h}$ 分别为第 k 个流域、预报时效为 h 的模式预报值， a 、 b 、 c 分别为 EC、JMA、WRF 模式预报的权重系数。

Table 1. The first principal component of areal rainfall in sub-units of Chaohu Lake basin

表 1. 巢湖流域各个子单元不同预报时效第 1 主成分线性组合

	24 h	48 h	72 h
店埠河南淝河派河	pc1 = 0.46EC + 0.43WRF	pc1 = 0.53EC + 0.3WRF	pc1 = 0.75EC
杭埠河丰乐河	pc1 = 0.46EC + 0.53JMA	pc1 = 0.44EC + 0.34WRF	pc1 = 0.62EC
马尾河西河上游	pc1 = 0.33EC + 0.42JMA	pc1 = 0.26EC + 0.22WRF	pc1 = 0.45EC
西河下游清溪裕溪河	pc1 = 0.54EC + 0.4WRF	pc1 = 0.61EC	pc1 = 0.45EC + 0.17WRF
柘皋河子单元	pc1 = 0.42EC + 0.74WRF	pc1 = 0.87EC	pc1 = 0.77EC
巢湖湖面子单元	pc1 = 0.50EC + 0.37WRF	pc1 = 0.56EC	pc1 = 0.37EC + 0.28 WRF

Table 2. The variance contribute rate of the first principal component

表 2. 巢湖流域各个子单元不同预报时效第 1 主成分方差贡献率

	24 h	48 h	72 h
店埠河南淝河派河	0.8	0.88	0.9
杭埠河丰乐河	0.79	0.87	0.96
马尾河西河上游	0.85	0.82	0.9
西河下游清溪裕溪河	0.76	0.95	0.93
柘皋河子单元	0.88	0.85	0.79
巢湖湖面子单元	0.94	0.93	0.81

利用 3 种单模式预报产品做主成分分析，根据上述确定主成分个数的原则和方法。表 1 列出了不同子流域不同预报时效的第 1 主成分组合，得到不同子流域不同预报时效下的第 1 主成分组合方程。表 2 列出了不同子流域不同时效下的第 1 主成分方差贡献率。由于本地业务中仅有 3 个单模式预报产品，变换后仅有 2 个主成分，通过方差贡献验证后，第 1 主成分方差贡献率均超过 80%，可以作为新的自变量与各子流域不同时效下时实况面雨量进行一元回归分析，即可得到基于 3 个模式的主成分回归的集成预报模型。

3.2. 正确率检验

图 2 给出了 2016 年 6~8 月 ECMWF、JMA、WRF 3 种单一模式和两种集成预报产品对巢湖流域 6 个子单元面雨量预报正确率检验结果。分析发现，随着预报时效的延长，单模式预报和集成预报的正确率都呈下降趋势，单一模式较集成预报模式下降更快。主成分回归方法 24 h 预报(75.6%)和 72 h 预报(71.9%)的正确率略高于单一模式和多

元线性回归集成预报, 综合预报效果较单一模式更好。对比两种集合预报, 主成分回归得分高于多元线性回归。

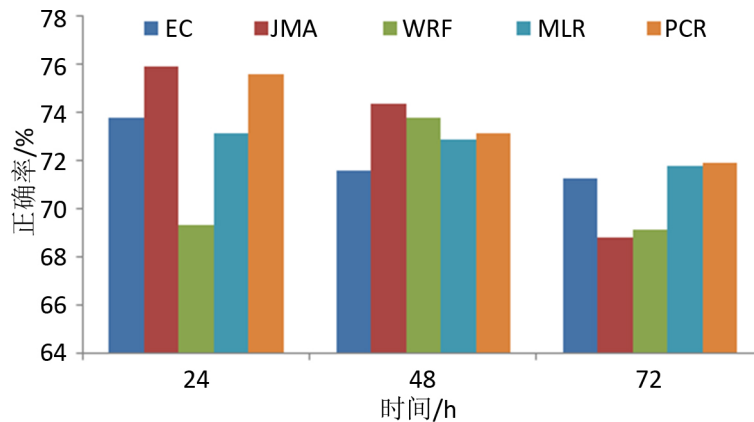


Figure 2. 24, 48, 72 h forecast accuracy of ECMWF, JMA, WRF and MLR, PCR
图 2. 24、48、72 h 的 ECMWF、JMA、WRF 单模式和 MLR、PCR 集成预报正确率

3.3. 平均绝对误差检验

图 3 给出了 2016 年 6~8 月 ECMWF、JMA、WRF 3 种单一模式和两种集成预报对巢湖流域 6 个子单元面雨量预报平均绝对误差检验结果。分析发现, 随着预报时效的延长, 各模式预报和集成预报的平均绝对误差都呈增大趋势。24 h、48 h 时效内, 两种集成预报平均绝对误差略高于 WRF 单一模式, 小于其他单一模式; 72 h 时效内, JMA 预报平均绝对误差最低, 主成分回归方法平均绝对误差略高于 JMA 单一模式, 优于其他单一模式和多元线性回归集成预报方法。

图 4 给出了 2016 年 6~8 月 ECMWF、JMA、WRF 3 种单一模式和两种集成预报对巢湖流域 6 个子单元面雨量预报 TS 评分检验结果。分析发现, 随着降水等级的增大, 随着时效的延长, 24 h 以后各种模式 TS 评分逐渐下降。在 TS 评分中, 小雨级别, 24 h 预报 JMA 评分最高(55.7), 两种集成预报分数与其接近, 均高于其他 2 种模式, 48、72 预报时效内, 集成预报优于单一模式, 主成分回归由于多元回归模型; 中雨级别, 集成预报得分高于其他单模式, 主成分回归以效果好于多元回归; 大雨级别, JMA 效果较好, 集成预报均高于其他 2 种单模式; 暴雨级别, 48、72 预报时效内, 两种集成预报方法均高于 3 种单一模式, 主成分回归 TS 评分最高。TS 评分总体评价, 两种集成预报模型较单一模式效果好, 主成分回归较多元线性回归更好。

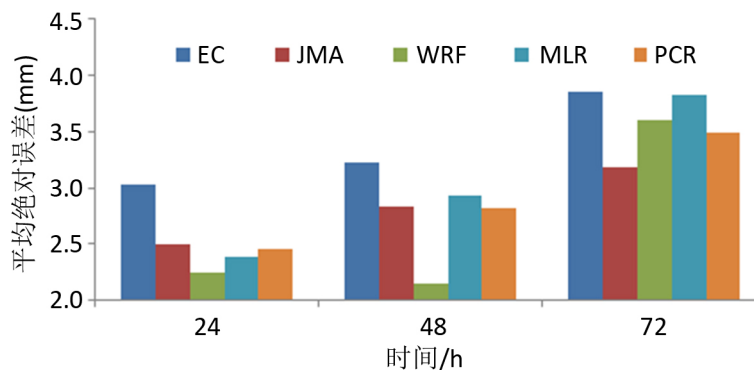


Figure 3. 24, 48, 72 h average absolute errors of ECMWF, JMA, WRF and MLR, PCR

图 3. 24、48、72 h 的 ECMWF、JMA、WRF 单模式和 MLR、PCR 集成预报平均绝对误差

3.4. TS 评分检验

24、48、72 h 的 ECMWF、JMA、WRF 单模式和 MLR、PCR 集成预报的 TS 评分，见图 4。

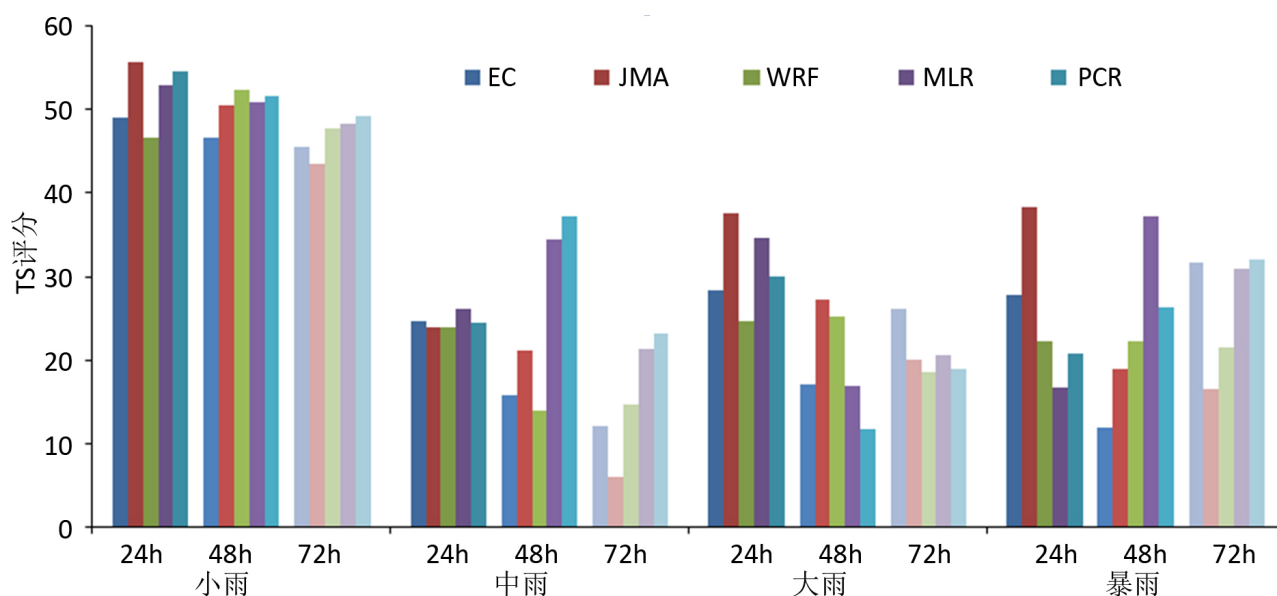


Figure 4. 24, 48, 72 h TS score of ECMWF, JMA, WRF and MLR, PCR

图 4. 24、48、72 h 的 ECMWF、JMA、WRF 单模式和 MLR、PCR 集成预报的 TS 评分

4. 结论与讨论

通过对 ECMWF、JMA、WRF 3 种模式尝试集成预报效果进行检验，得到：

1) 从 TS 评分和平均绝对误差可以看出，集成预报虽未改变预报性能随预报时效延长而降低的趋势，但随着预报时效延长，集成模型较单一模式预报效果更好。

2) 从 24 h、48 h、72 h 预报时效面雨量预报的正确率评估可以看出，两种集成方法均从整体上提高了预报的准确率，主成分回归较多元线性回归有进一步的提高。从平均绝对误差可以看出，主成分回归集成方法降低了平均绝对误差随预报时效增加而增大的速度，且平均绝对误差较单一模式稳定。

3) 从 24 h、48 h、72 h 面雨量预报时效的 TS 评分可以看出，小雨和中雨级别，24 h、48 h、72 h 三个时效面雨量预报中，两种集成预报方法 TS 评分有明显的优势，主成分回归较多元线性回归优势更明显。大雨级别，24h 集成预报优势明显。暴雨级别，48 h、72 h 面雨量预报时效内，两种集成预报效果较单一模式有明显优势。

4) 从总体评估效果可以得出，集成预报方法较单一模式有明显优势。比较两种集成方法，主成分回归较多元线性回归优越。

参考文献

- [1] 徐晶, 林建, 姚学祥, 等. 七大江河流域面雨量计算方法及应用[J]. 气象, 2001, 27(11): 13-16.
XU Jing, LIN Jian, YAO Xuexiang, et al. Calculating method of area rainfall over seven river valleys and its application. Meteorological Monthly, 2001, 27(11): 13-16. (in Chinese)
- [2] 中国气象局. 全国七大江河流域面雨量监测和预报业务规定[S]. 北京: 气象出版社, 2005.
China Meteorological Administration. The monitoring and forecast rules of areal rainfall on seven major river basins. Beijing: Meteorology Press, 2005. (in Chinese)
- [3] 中国气象局. 中短期天气预报质量检验办法[S]. 北京: 气象出版社, 2005.
China Meteorological Administration. The method of quality inspection on short-term weather forecast. Beijing: Meteorology Press, 2005. (in Chinese)

Press, 2005. (in Chinese)

- [4] 朱红芳, 王东勇, 朱鹏飞, 等. GRAPES 模式在淮河流域面雨量预报中的应用[J]. 气象, 2007, 33(3): 76-82.
ZHU Hongfang, WANG Dongyong, ZHU Pengfei, et al. Application of GRAPES model to area-rainfall forecast in the Huaihe River basin. Meteorological Monthly, 2007, 33(3): 76-82. (in Chinese)
- [5] 刘静, 叶金印, 等. 淮河流域汛期面雨量多模式预报检验评估[J]. 暴雨灾害, 2014, 33(1): 60-64.
LIU Jing, YE Jinyin, et al. The performance evaluation of the multi-model forecasting of areal rainfall for Huaihe River basin during flood season. Torrential Rain and Disasters, 2014, 33(1): 60-64. (in Chinese)
- [6] 盛飞. 主成分分析法在神经网络集成预报中的应用[J]. 气象科学, 2005, 25(4): 362-368.
SHENG Fei. Application of principal component analysis method in neural network integrative forecasting. Scientia Meteorologica Sinica, 2005, 25(4): 362-368. (in Chinese)
- [7] 严明良, 缪启龙, 沈树勤. 基于超级集合思想的数值预报产品变权集成方法探讨[J]. 气象, 2009, 35(6): 19-25.
YAN Mingliang, MIAO Qilong and SHEN Shuqin. Exploration on ensemble model of numerical forecasting based on variable-weight super-ensemble method. Meteorological Monthly, 2009, 35(6): 19-25. (in Chinese)
- [8] 叶金印, 刘静, 等. 信息矩阵方法在面雨量预报集成中的应用[J]. 自然灾害学报, 2014, 23(4): 127-133.
YE Jinyin, LIU Jing, et al. Application of information matrix method to consensus forecast of areal rainfall. Journal of Natural Disasters, 2014, 23(4): 127-133. (in Chinese)