

韩焱红, 矫梅燕, 陈静, 等. 2013. 基于贝叶斯理论的集合降水概率预报方法研究. 气象, 39(1): 1-10.

# 基于贝叶斯理论的集合降水概率 预报方法研究<sup>\*</sup>

韩焱红<sup>1</sup> 矫梅燕<sup>2</sup> 陈 静<sup>3</sup> 陈法敬<sup>3</sup>

1 南京信息工程大学大气科学学院, 南京 210044

2 中国气象局, 北京 100081

3 中国气象局数值预报中心, 北京 100081

**提 要:** 将贝叶斯理论应用到集合降水概率预报方法研究中。采用集合预报资料和历史观测资料, 通过建立贝叶斯产品处理技术(Bayesian Processor of output, BPO)降水概率预报模型, 将一组集合成员降水确定预报值修订为一组贝叶斯降水概率分布或概率密度的预报, 并获得表征每个集合成员预报能力有效信息评分(Informativeness Score, IS)。基于 IS 值对集合成员概率预报信息融合, 得到集成贝叶斯降水概率预报, 并采用连续等级概率评分(Continuous Ranked Probability Score, CRPS)方法检验试验结果。结果表明, 基于 BPO 方法得到的集成贝叶斯降水概率预报可靠性高于由集合预报得到的直接概率预报。

**关键词:** 贝叶斯理论, 降水集合预报, 概率预报, 试验

**中图分类号:** P456

**文献标识码:** A

**doi:** 10. 7519/j. issn. 1000-0526. 2013. 01. 001

## Study on the Method of Rainfall Ensemble Probability Forecast Based on Bayesian Theory and Its Preliminary Experiments

HAN Yanhong<sup>1</sup> JIAO Meiyang<sup>2</sup> CHEN Jing<sup>3</sup> CHEN Fajing<sup>3</sup>

1 School of Atmospheric Science, Nanjing University of Information Science and Technology, Nanjing 210044

2 China Meteorological Administration, Beijing 100081

3 CMA Numerical Prediction Centre, Beijing 100081

**Abstract:** The paper applies BPO (Bayesian Processor of Output) method based on Bayesian theory to the method of rainfall ensemble probability forecast. Using ensemble prediction data and historical observational data, we develop a rainfall probability forecast model, and then revise a set of precipitation predicted value into a set of Bayesian precipitation probability forecast in the form of continuous probability distribution or continuous probability density. Besides, we obtain a group value of Informativeness Score (IS), which can express the prediction ability of each ensemble member. Furthermore, we fuse the probability forecast results of each member into an integration Bayesian precipitation probability forecast on the basis of IS and test the results with Continuous Ranked Probability Score (CRPS). Experiment results show that the reliability of integration Bayesian precipitation probability forecast is higher than ensemble direct probability forecast.

**Key words:** Bayesian theory, precipitation ensemble forecast, probability forecast, experiments

\* 国家自然科学基金(41075035)资助

2012年4月25日收稿; 2012年9月3日收修定稿

第一作者: 韩焱红, 主要从事集合预报应用研究. Email: hanyanhong.1988@163.com

通信作者: 矫梅燕, 主要从事天气预报研究. Email: jiaomy@cma.gov.cn

## 引言

天气预报存在着不可避免的不确定性(Lorenz, 1963; 1965), 概率是表达预报不确定性的一种方式(王东海等, 2011), 集合预报(杜钧等, 2010; 矫梅燕, 2010; 陈静等, 2002)则是获得概率预报的一个有效途径。集合预报产品的释用是集合预报系统必不可少的一部分, 是实现模式结果实际应用价值的重要过程。目前的集合预报释用方法主要以模式预报值为样本, 统计得到一系列概率预报结果, 如天气要素概率预报图、邮票图和面条图等。由于集合预报系统在获得不确定性来源过程中仍存在一定缺陷, 由此得到的直接概率预报结果并不能完整地量化表达不确定性, 因此, 近年来国内外气象学者逐渐研究基于贝叶斯理论的概率预报方法来解决这一问题。

贝叶斯理论是统计学中的一个重要分支, 在水文预报及气象预报等领域有所应用。从 20 世纪 90 年代开始, 基于贝叶斯理论的水文模拟不确定性估计方法被广泛地应用于模拟方法、参数估计和水文预报等方面, 成为水文不确定性研究的主流方向。90 年代末, 气象学者将贝叶斯理论应用于单一数值预报产品的概率化预报中, 如 Krzysztofowicz (1983; 1999a) 提出了贝叶斯产品处理技术(Bayesian Processor of output, BPO), 采用模式预报值作为预报因子, 通过建立 BPO 预报模型对预报量先验概率修订, 得到预报量累积概率分布或概率密度的预报。随着数值预报技术的发展, 集合预报逐渐发展成熟, 近年来贝叶斯理论也逐渐应用于该研究领域, Raftery(2005)提出了贝叶斯模式平均(Bayesian Model Averaging, BMA)的方法, 利用地面温度历史集合预报资料, 将单个集合成员预报结果修订为概率密度函数形式的预报, 但由于模式集合平均与气候平均有很大差别, 该方法对于极端事件的预报不够准确。Bishop(2008)对上述方法进一步地改进, 将气候分布与贝叶斯理论结合得到既适用于非极端事件又适于极端事件的概率预报。中国气象局也于 2010 年开始这方面的研究, 赵琳娜等(2011)采用淮河流域历史降水资料及集合预报资料利用 BMA 方法对中国气象局(CMA)集合预报 15 个成员的定量降水预报进行了集成与订正, 得到有预测效果的概率密度函数, 使得观测降水真值包含在有效区间预报内的可能性更大, 获得预报能力高于确

定预报的概率预报。陈法敬等(2011)以连续预报量——温度为例, 对 BPO 方法在集合预报中的应用进行了初步试验验证, 将一组集合成员预报值修订为一组概率预报并对其预报结果合理融合得到预报能力高于单一集合成员的集成贝叶斯概率预报。

由前人基于贝叶斯理论的概率预报结果可以看出, 采用数值模式产品提供的有效预报信息对预报量的气候(先验)概率进行修订, 得到模式预报信息与气候信息最佳融合的概率预报可以提高预报准确性。降水作为离散型预报量, 其概率预报形式一般有两种: 降水有无的分类概率预报和降水量等级概率预报。实际上, 降水量在  $[0, \infty)$  内的各个连续值处都存在一定概率, 获得该范围内连续的概率分布或概率密度预报可以更加完整地体现预报不确定性。本文以 24 小时降水为预报量, 集合成员预报值为确定预报值, 根据 BPO 方法(Krzysztofowicz et al, 2001; Krzysztofowicz, 1999)建立降水概率预报模型, 获得一组在预报范围内连续变化的降水概率分布或概率密度预报, 并按照预报能力对各集合成员贝叶斯降水概率预报信息融合, 获得集成贝叶斯降水概率预报。

## 1 方法和资料

### 1.1 方法

#### 1.1.1 贝叶斯方法

贝叶斯方法是基于贝叶斯定理而发展起来用于系统地阐述和解决统计问题的方法(茆诗松, 1999; 吴喜之, 2000)。该方法的核心为贝叶斯公式, 其基本形式如下:

$$P(Y | X) = \frac{P(Y)P(X | Y)}{P(X)} \quad (1)$$

式中,  $P(Y)$  是未知数  $Y$  的先验概率, 它是由已知的先验信息获取的最初概率, 反映了人们在抽样前对  $Y$  的认知;  $P(X|Y)$  则是样本值  $X$  的抽样分布密度, 其综合了未知数  $Y$  的样本信息和总体信息(综合称为抽样信息);  $P(X)$  为随机变量  $X$  的边缘分布。公式表达了通过抽取样本  $X$ , 利用抽样信息对未知数  $Y$  的先验概率进行修订, 得到重新估计的条件概率即后验概率  $P(Y|X)$  的算法。

#### 1.1.2 BPO 方法

BPO 方法是一种将贝叶斯理论运用到气象领

域,通过融合预报量先验信息,将单一数值模式预报值修订为概率预报的技术。本文选取降水作为预报量  $Y$ ,将单一集成员降水预报值作为预报因子  $X$ 。由于离散型预报量降水总体上可以分为有无降水两种情况,而在有降水条件下的降水量(以下简称为条件降水量)的分布在  $(0, \infty)$  范围内是连续的。因此为了得到降水在  $[0, \infty)$  预报值范围内连续的概率分布预报,分别对有降水的概率和条件降水量的概率分布修订,得到有降水的后验概率  $\pi$  及条件降水量的后验概率分布  $\Phi(y)$  或后验密度函数  $\phi(y)$ 。基于 BPO 方法的降水概率预报数学模型如式(2)或(3)所示,式中  $x$  为预报因子  $X$  的值, $y$  为  $\geq 0$  的任意实数, $\delta(y)$  为  $y$  的狄拉克函数。由公式可以看出,基于 BPO 方法得到的降水概率预报由两部分构成:无降水的概率和条件降水量的概率分布。

概率分布形式:

$$P(Y \leq y | X = x) = (1 - \pi) + \pi\Phi(y), y \geq 0 \quad (2)$$

概率密度形式:

$$p(y | X = x) = (1 - \pi)\delta(y) + \pi\phi(y), y \geq 0 \quad (3)$$

降水概率预报模型中有降水的后验概率  $\pi$  由公式(4)获得,式中  $g$  为由历史降水观测资料统计得到的有降水先验概率(降水阈值为 0.1 mm)。 $f_0$ 、 $f_1$  则分别为无降水和有降水时,由降水观测值、预报值构成的联合样本获得的预报值  $x$  的条件密度函数即似然函数。

$$\pi = \left[ 1 + \frac{1 - g}{g} \frac{f_0(x)}{f_1(x)} \right]^{-1} \quad (4)$$

模型中条件降水量的后验概率分布  $\Phi(y)$  和后验概率密度  $\phi(y)$  则分别由式(5)和(6)获得。公式中  $Q$  代表标准正态分布函数; $G(y)$ 、 $K(x)$  则是分别由降水历史观测值样本、预报值样本估计得到的条件降水量  $y$  的先验概率分布、预报值  $x$  的边缘分布; $c_0$ 、 $c_1$  和  $T$  为通过建立似然模型得到的后验参数。

概率分布函数形式:

$$\Phi(y) = Q \left\{ \frac{1}{T} [Q^{-1}G(y) - c_1 Q^{-1}K(x) - c_0] \right\} \quad (5)$$

概率密度函数形式:

$$\phi(y) = \frac{1}{T} \exp \left( \frac{1}{2} \{ [Q^{-1}G(y)]^2 - [Q^{-1}\Phi(y)]^2 \} \right) g(y) \quad (6)$$

似然模型是为了获得表征预报量和预报因子之间依赖关系的似然参数而建立的数学模型。由于降水观测值与预报值并不服从标准正态分布,其分布特征使得两者之间的依赖关系难以用解析函数来刻画,因此本文采用 Kelly 等(1997)和 Krzysztofowicz (1997)提出的亚高斯似然模型来获得似然参数。其主要思想是先通过正态分位数转换(Normal Quantile Transform, NQT)将预报量  $y$ 、预报因子  $x$  变换为正态分布变量  $u$ 、 $z$ ,如式(7)。

$$u = Q^{-1}G(y), z = Q^{-1}K(x) \quad (7)$$

式中, $Q^{-1}$ 表示标准正态分布函数反函数,因此转换后的变量均完全服从标准正态分布(高斯分布)。在转换空间内建立的似然模型则为亚高斯似然模型,根据  $u$ 、 $z$  的均值  $\mu_0$ 、 $\mu_1$ , 方差  $\sigma_0^2$ 、 $\sigma_1^2$  以及协方差  $\sigma_{10}$  通过式(8)得到似然参数  $a$ 、 $b$  和  $\sigma^2$ ,以此来描述 NQT 转换量  $u$ 、 $z$  之间的线性回归关系,如式(9)所示。

$$a = \frac{\sigma_{10}}{\sigma_0^2}, b = \mu_1 - \frac{\sigma_{10}}{\sigma_0^2}\mu_0, \sigma^2 = \sigma_1^2 - \frac{\sigma_{10}^2}{\sigma_0^2} \quad (8)$$

$$E(Z | U = u) = au + b, \text{Var}(Z | U = u) = \sigma^2 \quad (9)$$

式中, $E$  为降水观测值  $Y$  的 NQT 转换量  $U$  值为  $u$  时,预报值  $X$  的转换量  $z$  的期望, $\text{Var}$  则为其方差。同时,通过建立似然模型分别由式(10)和(11),获得表征预报因子有效信息评价指标的有效信息评分(Informativeness Score, IS)(Krzysztofowicz, 1992)及 BPO 预报模型中的后验参数。

$$IS = |\gamma| = \left[ \left( \frac{a}{\sigma} \right)^2 + 1 \right]^{-\frac{1}{2}} \quad (10)$$

$$c_1 = \frac{a}{a^2 + \sigma^2}, c_0 = \frac{-ab}{a^2 + \sigma^2}, T = \left( \frac{\sigma^2}{a^2 + \sigma^2} \right)^{\frac{1}{2}} \quad (11)$$

## 1.2 资料

本文选取了 5 个不同气候区的代表站点:广州、南京、武汉、成都和北京站。将站点 24 h 降水量作为预报量  $Y$ ,集合预报系统中单一集成员 24 h 降水预报值(预报时效为 24、72 和 120 h)作为预报因子  $X$ 。

其中降水的先验概率由国家气象信息中心提供的全国基准站 1952—2007 年 6 月逐日 20—20 时 24 h 降水观测资料获得。为了研究 BPO 方法在多

模式集合预报中的应用,本文预报因子  $X$  是由交互式全球大集合预报系统 (THORPEX Interactive Grand Global Ensemble, TIGGE) 提供的中国气象局 (CMA) 与美国环境预报中心 (NCEP) 逐日 24、72 和 120 h 预报时效的全球格点降水预报资料利用双线性插值方法得到广州、南京、武汉、成都和北京站 5 个观测站的降水预报, 试验时段为 2008 年 6 月 1—30 日期间。

## 2 基于确定性预报的降水概率预报模型试验

### 2.1 降水概率预报模型的应用实例

以 CMA 和 NCEP 集合预报资料中集合控制预报值作为确定性预报值, 对基于 BPO 方法的降水概

率预报模型进行应用试验。

分别对广州、南京、武汉、成都和北京站建立 BPO 降水概率预报模型, 得到 5 个站点的 6 月逐日有降水的后验概率  $\pi_k$ , 其与历史观测资料统计得到的先验概率  $g_k$  比较结果如图 1 所示 ( $k$  为 6 月中的第  $k$  天)。

由图 1 可以看出, 根据降水历史观测资料统计得到的站点有降水先验概率  $g_k$  逐日变化平缓, 其中广州站的  $g_k$  值在 0.45~0.75 之间 (如图 1a), 武汉和成都站的  $g_k$  值在 0.3~0.7 之间 (如图 1b 和 1d), 南京和北京站的  $g_k$  值在 0.2~0.5 之间 (如图 1c 和 1e)。而通过建立 BPO 降水概率预报模型修订得到的后验概率  $\pi_k$  逐日变化起伏较大, 这说明数值预报值作为预报因子, 提供的预报因子有效信息发挥了一定修订作用。

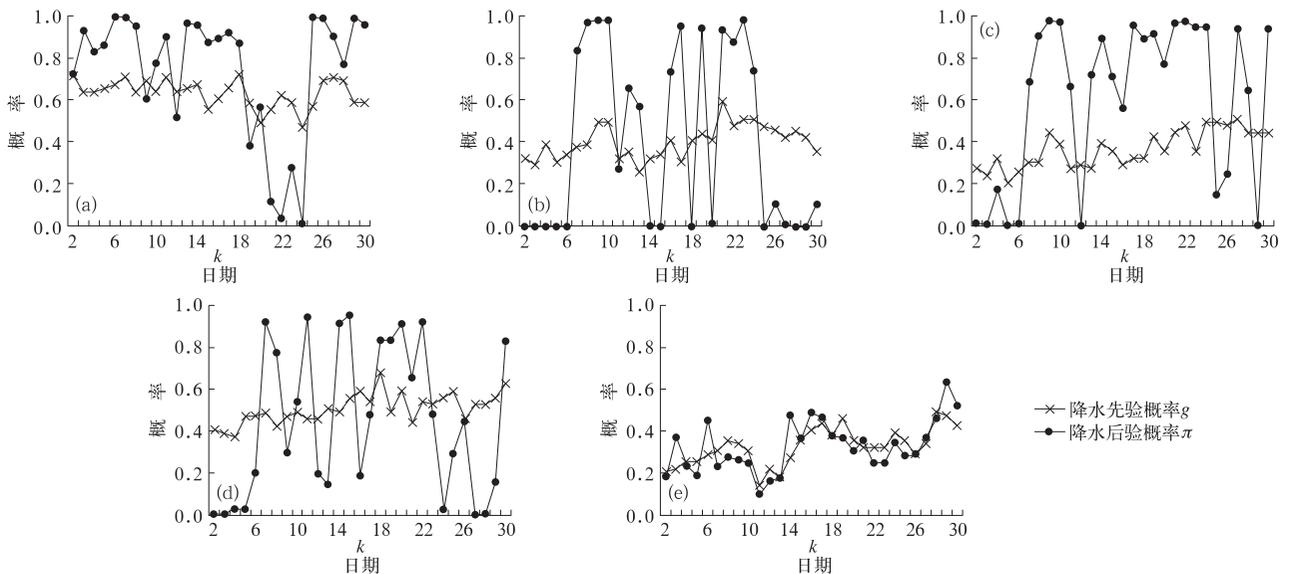


图 1 广州 (a)、武汉 (b)、南京 (c)、成都 (d) 和北京 (e) 站点 6 月逐日有降水先验概率  $g_k$  与后验概率  $\pi_k$

Fig. 1 The prior probability ( $\times$ ) of precipitation occurrence  $g_k$  and posterior probability  $\pi_k$  ( $\bullet$ ) of each day in June at stations

Guangzhou (a), Wuhan (b), Nanjing (c), Chengdu (d), and Beijing (e)

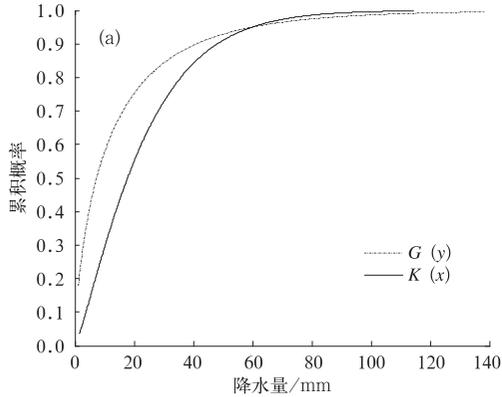
将历史观测资料中有降水时的降水观测值作为条件降水量的历史观测样本, 与其对应的集合预报控制预报值作为历史预报值样本, 通过统计分析发现广州、南京、武汉、成都和北京站的 6 月逐日条件降水量近似服从威布尔分布。因此, 本文采用威布尔分布函数估计条件降水量的先验分布  $G(y)$  和预

报值的边缘分布  $K(x)$ 。

$$G(y | \alpha, \beta) = 1 - e^{-\left(\frac{y}{\alpha}\right)^\beta}, y > 0 \quad (12)$$

以广州站 2008 年 6 月 29 日为例, 条件降水量的先验分布  $G(y)$  及预报值边缘分布  $K(x)$  如图 2a 所示。可以看出, 条件降水量  $Y$ 、预报值  $X$  的威布尔分布形状与尺度存在差异, 两者之间的统计关系

难以统一刻画。分别对其进行 NQT 转换,得到条件降水量、预报值的 NQT 转换量  $U$ 、 $Z$ ,其分布形式如图 2b 所示。转换量  $U$ 、 $Z$  均服从标准正态分布,分布曲线完全重合,便于似然模型的建立。



以 CMA 和 NCEP 集合预报为例,在转换空间内分别建立 5 个研究站点的似然模型,得到后验参数  $c_0$ 、 $c_1$  和  $T$ ,结果如表 1 和表 2 所示。

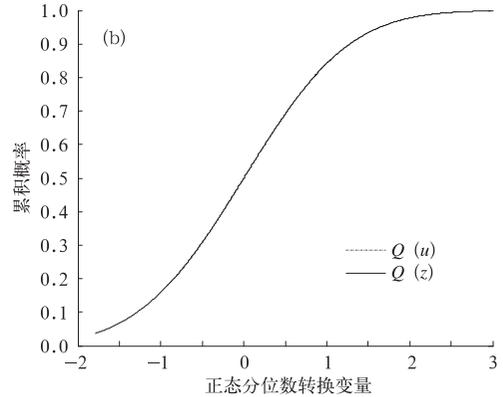


图 2 条件降水量的边缘分布  $G(y)$  及预报值的边缘分布  $K(x)$  (a) 和 NQT 转换量  $U$ 、 $Z$  的标准正态分布  $Q(u)$  及  $Q(z)$  (b) (以广州 2008 年 6 月 29 日为例)

Fig. 2 The marginal distribution  $G$  of predictand  $y$  and marginal distribution  $K$  of predictor  $x$  (a) and the standard normal distribution of NQT transformed  $u$  and  $z$  (b) (Take Guangzhou 29 June 2008 as example)

表 1 站点后验参数值(以 CMA 集合成员为例)  
Table 1 Posterior parameters at five stations (Take CMA ensemble member as example)

站点	$c_0$	$c_1$	$T$
广州	0.3062	0.5915	0.8110
武汉	0.0293	0.3855	0.8918
南京	0.0114	0.6107	0.6887
成都	0.2330	0.6369	0.6788
北京	0.0114	0.2775	0.9513

表 2 站点后验参数值(以 NCEP 集合成员为例)  
Table 2 Posterior parameters at five stations (Take NCEP ensemble member as example)

站点	$c_0$	$c_1$	$T$
广州	0.3503	0.6142	0.7857
武汉	0.0247	0.0247	0.5979
南京	-0.1987	-0.1283	0.5909
成都	0.1862	0.5119	0.8263
北京	0.0117	0.1900	0.9796

利用表 1 中广州站的后验参数值,由式(6)得到给定任意 4 个预报因子值( $X=10$ 、 $20$ 、 $35$  和  $55$  mm)时的条件降水量后验概率密度函数,如图 3 所示。

由图 3 可见,预报因子值分别取 10、20、35 和 55 mm 时,概率密度峰值对应的降水量分别为 1、5、

15 和 35 mm。同时,先验概率密度  $g(y)$  的峰值位于条件降水量小值区(倒 J 型分布),而预报因子  $x$  值愈大,修订得到的后验密度函数  $\phi(y)$  愈圆滑,峰值也随之向降水量大值区移动,而体现了预报因子

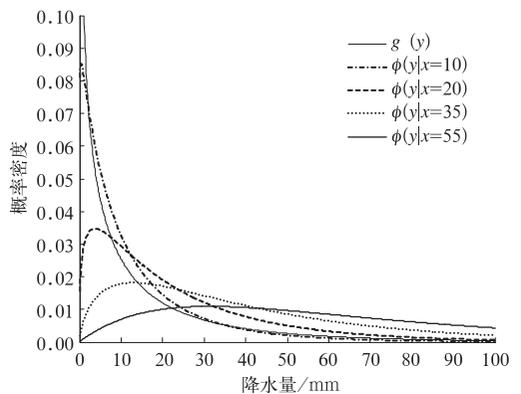


图 3 条件降水量  $Y$  的先验概率密度  $g(y)$  及给定任意 4 个预报因子  $X$  值

( $X=10$ 、 $20$ 、 $35$  和  $55$  mm) 的后验概率密度  $\phi(y|x=10)$ 、 $\phi(y|x=20)$ 、 $\phi(y|x=35)$  和  $\phi(y|x=55)$

Fig. 3 The prior density function  $g(y)$  and posterior density functions  $\phi(y)$  based on four different predictor realizations  $x=10, 20, 35$  and  $55$  mm

对概率的修订作用。

## 2.2 基于确定性预报的降水概率预报结果分析

分别将 CMA 和 NCEP 集合预报控制预报值作为确定性预报,对广州、南京、武汉、成都和北京站 2008 年 6 月逐日 24 h 降水建立 BPO 降水概率预报

模型,得到各站点的逐日降水概率分布或概率密度预报。其中,以 CMA 控制预报对广州 2008 年 6 月 30 日降水,NCEP 控制预报对武汉 2008 年 6 月 21 日降水的预报为例,得到确定性预报的概率化结果——单一集合成员贝叶斯降水概率预报的结果如图 4 所示。

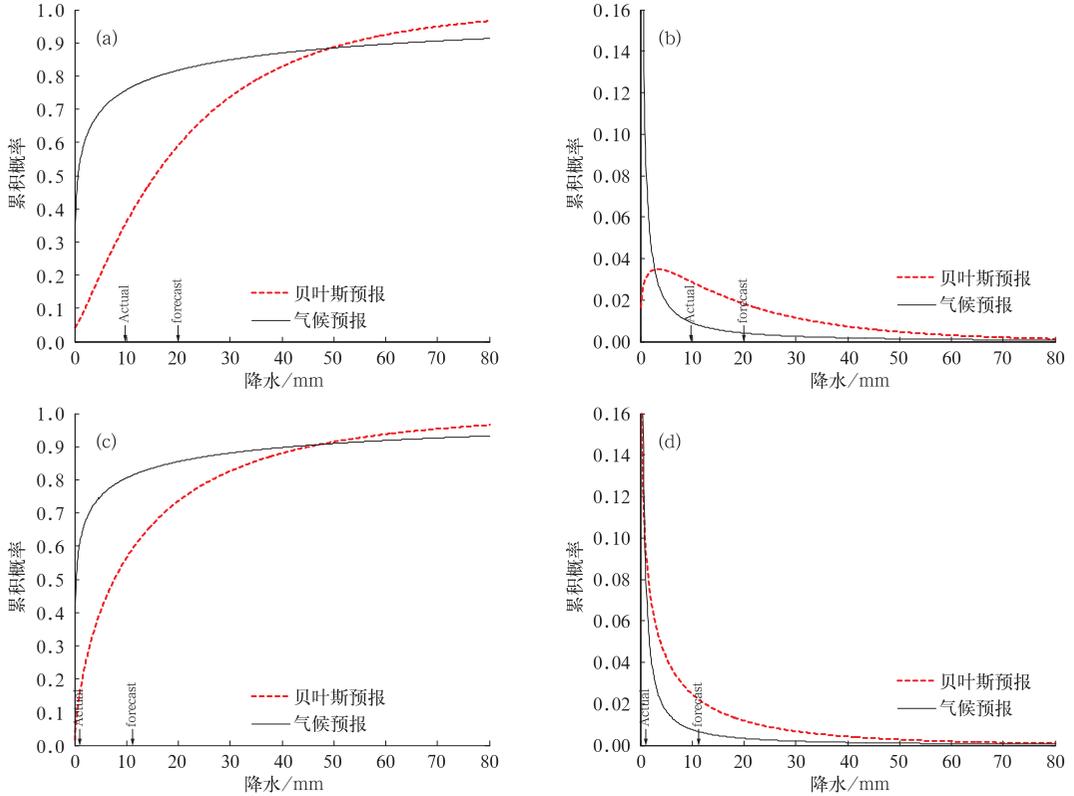


图 4 贝叶斯降水概率预报与气候概率预报试验结果对比

- (a) 基于 CMA 控制预报对广州 2008 年 6 月 30 日的降水预报得到的概率分布预报,  
 (b) 同上,但为概率密度预报(Actual:降水实况 8.9 mm,forecast:集合预报控制预报值 19.3 mm),  
 (c) 基于 NCEP 控制预报对武汉 2008 年 6 月 21 日的降水预报得到的概率分布预报,  
 (d) 同上,但为概率密度预报  
 (Actual:降水实况 0.4 mm,forecast:集合预报控制预报值 10.93 mm)

Fig. 4 Comparisons of Bayesian forecast and climate forecast based on deterministic prediction:

- (a) cumulative probability prediction taking CMA precipitation forecast on 30 June, 2008 at Guangzhou for example, (b) the same as above but for probability density (actual observation: 8.9 mm, forecast: value 19.3 mm), (c) cumulative probability prediction taking NCEP precipitation forecast on 21 June, 2008 at Wuhan for example, and (d) the same as above but for probability density  
 (Actual observation: 0.4 mm, forecast: 10.93 mm)

由图 4 可以看出,广东站的实际观测值为 8.9 mm,CMA 控制预报结果为 19.3 mm,修订得到的概率密度峰值位于观测值附近(如图 4b),预报准确性有所提高;武汉站的实际观测值为 0.4 mm,

NCEP 控制预报结果为 10.93 mm,修订得到的概率密度峰值同样位于观测值附近(如图 4d),且预报确定度较高。同时,得到的贝叶斯降水概率预报与确定性预报相比,给出了预报范围内更多的预报信息,

并以概率的形式定量化、连续地表达了降水预报不确定性。

本文选取的成都、南京和北京站的试验个例得到的贝叶斯降水概率预报形式与以上广州、武汉站的个例相似,不再赘述。

### 3 基于集合预报的降水概率预报模型应用

#### 3.1 集合成员的预报模型应用结果

前面阐述了基于确定性预报建立 BPO 降水概率预报模型,获得单一集合成员贝叶斯降水概率预报的方法。试验结果表明,得到的概率预报具有一定的预报能力。而集合预报的预报不确定性是所有集合成员预报不确定性的完整体现,因此有必要研究一种合理的集合成员概率预报信息融合方法。将集合预报结果包含的一组集合成员预报值 $(x_1, \dots, x_n)$ ( $n$ 为集合成员数)中每个集合成员预报值视为确定预报数值模式产品,基于 BPO 方法建立降水概率预报模型,获得一组概率分布预报 $\{P(Y \leq y | X = x_i) | i = 1, \dots, n\}$ 。以广州站 2008 年 6 月 4 日的 24 小时降水预报为例,图 5 给出 CMA 集合预报第

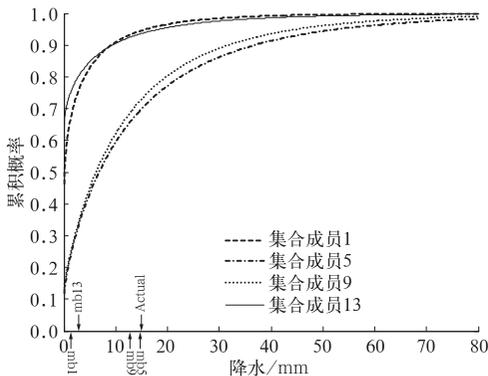


图 5 由 CMA 集合预报第 1、5、9、13 集合成员对广州 2008 年 6 月 7 日降水量 24 h 预报值修订得到的降水概率分布预报 (Actual:降水观测值 14.4 mm, mb1、mb5、mb9 和 mb13 分别表示集合成员 1、5、9 和 13 的预报值 1.82、15.26、13.49 和 2.37 mm)

Fig. 5 The probability distribution based on the 1st, 5th, 9th, 13th member of CMA 24 h forecast at Guangzhou on 7 June, 2008 (Actual observation: 14.4 mm; mb1, mb5, mb9, mb13 stand for the forecasts of member 1, 5, 9, 13 about 1.82, 15.26, 13.49, 2.37 mm)

1、5、9 和 13 集合成员贝叶斯降水概率预报。

由图 5 可以看出,降水观测值为 14.4 mm,选定的 CMA 集合成员预报值均小于实况。概率化后的集合成员 1 和集合成员 13 的贝叶斯降水概率分布预报较相似,集合成员 5 和集合成员 9 的概率分布预报结果近乎重合。这表明,集合成员的降水预报值存在差异,所包含的不确定信息不同,因此其概率分布预报也有所区别。

#### 3.2 基于集合成员贝叶斯降水概率预报结果的集成方法

由于集合成员为预报提供的有效预报信息存在差异,其预报能力也有所不同。本文采用在亚高斯似然模型中获得的预报因子有效信息评分  $IS$  值表征其预报能力。 $IS$  值的范围为 $[0, 1]$ ,且  $IS$  值越大,预报因子包含的有效预报信息越多,集合成员的预报能力越高。分别基于 CMA 和 NCEP 集合预报成员对南京站 24、72 和 120 h 的降水预报值建立亚高斯似然模型,根据式(10)得到各集合预报成员的  $IS$  值,其中 NCEP 集合成员的结果如图 6a 所示。

由图 6a 可见,基于 NCEP 集合成员 24 h 降水预报得到的  $IS$  值明显高于 72 和 120 h,其中预报能力最强的为集合成员 7,预报能力最弱的为集合成员 9,  $IS$  值分别为 0.91 和 0.70。

$IS$  值可以较好地体现集合成员为预报量提供有效预报信息的能力,因此本文采用正比于  $IS^3$  的值作为权重系数(陈法敬, 2011)来体现不同预报能力的集合成员其预报信息在所有集合成员预报信息融合中的地位,见式(13)。

$$r_i(IS_i) = \frac{IS_i^3 - \min(IS^3)}{\sum_{i=1}^n IS_i^3 - n \cdot \min(IS^3)} \quad (13)$$

式中,  $\min$  运算为取各集合成员组成的  $IS^3$  向量中最小的元素,  $n$  为集合成员总数。

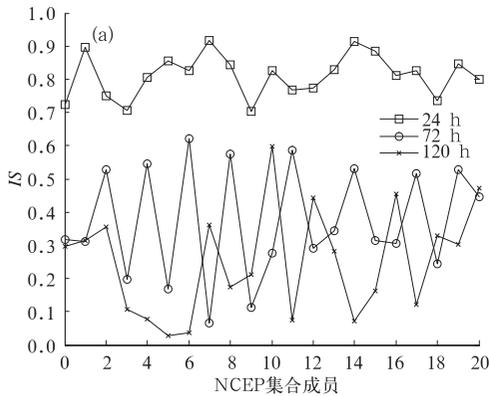
以 NCEP 对南京站 24 h 降水预报为例,获得与图 6a 中 24 h 预报时效的集合成员  $IS$  值对应的  $r_i(IS_i)$  值,如图 6b 所示。

由图 6b 可见,权重系数的变化趋势与  $IS$  值一致,对于 24 h 预报能力最强的集合成员 7,其预报信息的权重系数  $r_i(IS_i)$  值也最大,预报能力最弱的集合成员 9,其预报信息的权重系数  $r_i(IS_i)$  值最小,  $r_i(IS_i)$  值分别为 0.11 和 0。

另外,分别基于 CMA 和 NCEP 集合预报对广

州、武汉、成都和北京站 24、72 和 120 h 降水预报得到集合成员  $IS$  值以及 24 h 降水预报信息的权重系数  $r_i(IS_i)$ 。结果表明,权重系数可以很好地体现不同预报能力的集合成员其预报信息在融合中的地位。同时,各站的  $IS$  值变化特征与南京相似,两个集合预报系统对 24 h 降水预报能力均明显高于 72 和 120 h 预报。

基于权重系数  $r_i(IS_i)$  根据式(14)对一组集合成员概率密度预报信息加权平均且归一化,得到融



合后的集成贝叶斯降水概率密度预报。由式(15)对概率密度函数积分,得到概率分布预报。

$$p(y | \mathbf{X}) = \frac{\sum_{i=1}^n p(y | x_i) \cdot r_i(IS_i)}{\int_0^{\infty} \left[ \sum_{i=1}^n p(\xi | x_i) \cdot r_i(IS_i) \right] d\xi} \quad (14)$$

$$P(Y \leq y | \mathbf{X}) = \int_0^y p(\xi | \mathbf{X}) d\xi \quad (15)$$

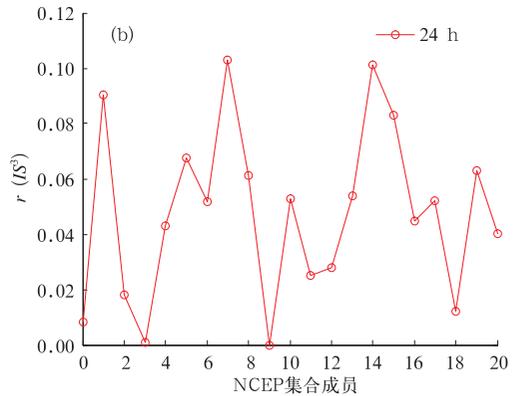


图 6 基于 NCEP 集合预报对南京站降水预报得到的  $IS$  值(a)和权重系数  $r_i(IS_i)$ (b)

Fig. 6 The  $IS$  (a) and  $r_i(IS_i)$  (b) obtained based on NCEP ensemble forecast for Nanjing precipitation

### 3.3 集成贝叶斯降水概率预报结果分析及检验

以 NCEP 对广州 2008 年 6 月 8 日的 24 h 降水预报为例,分别对各集合预报成员建立 BPO 降水概率预报模型,获得一组集合成员贝叶斯降水概率预报,并根据式(14)和(15)对预报信息融合,得到集成贝叶斯降水概率预报,如图 7 所示。

由图 7 可以看出,个例的降水观测值为 24.8 mm, NCEP 集合预报给出的累积概率在此处已达到极值 1, 概率密度为 0。集合模式直接概率预报的概率密度峰值主要位于 9、13 和 15 mm, 对应的降水等级范围则为小到中雨。修订得到的集成概率分布和概率密度预报在预报范围内是连续变化的曲线,且降水观测值处的概率密度有所提高,增加了预报的可靠性。

分别基于 CMA 和 NCEP 集合预报对广州、南京、北京、成都和武汉的 2008 年 6 月逐日 24 h 降水预报,建立 BPO 降水概率预报模型,得到站点逐日集成贝叶斯降水概率预报。采用连续等级概率评分

(Continuous Ranked Probability Score, CRPS)对各站点的概率预报结果进行检验,CRPS 值越小说明概率预报可靠性越高,预报结果越接近真实情况。

图 8 给出了 5 个站点集成贝叶斯降水概率预报和集合直接概率预报的 6 月 CRPS 评分均值,其中图 8a 和 8b 是分别为基于 CMA 和 NCEP 集合预报得到的检验结果。

由图 8 可以看出,基于 CMA 和 NCEP 集合预报得到的广州、武汉、南京、成都和北京站 6 月集成贝叶斯降水概率预报的 CRPS 评分均值明显小于集合直接概率预报,即集成贝叶斯降水概率预报的准确性高于集合直接概率预报。同时,各站点的集成预报与集合预报的 CRPS 评分变化趋势一致,说明得到的集成贝叶斯降水概率预报的预报效果仍基于集合预报结果。其中,基于 CMA 和 NCEP 集合预报获得的集成贝叶斯降水概率预报均对广州站的预报准确性改进最大。基于 CMA 集合预报得到的集成预报结果中,武汉站的 CRPS 值最小,预报准确性最高,基于 NCEP 集合预报得到的集成预报结果

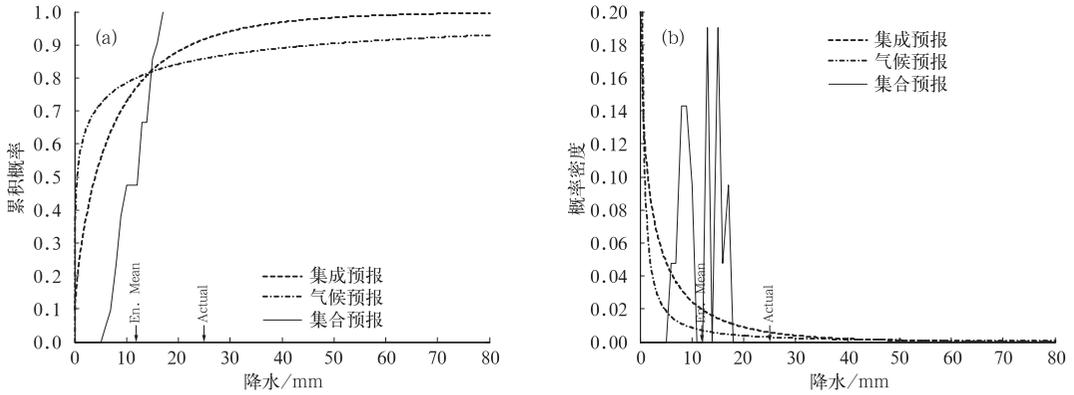


图 7 NCEP 对广州 2008 年 6 月 8 日 24 h 降水的集合模式直接概率预报(集合预报), 气候概率预报(气候预报)和集成贝叶斯降水概率预报(集成预报) (a) 概率分布形式, (b) 概率密度形式

(Actual: 降水观测值 24.8 mm, En. Mean: 集合平均值 11.28 mm)

Fig. 7 NCEP 24 h ensemble forecasts at Guangzhou on 7 June 2008 from the ensemble direct probability forecast (Ensemble Forecast), climate probability forecast (Climate Forecast), and integration Bayesian precipitation probability forecast (Integrated Forecast);

(a) probability distribution, and (b) probability density

(Actual observation: 24.8 mm, ensemble mean: 11.28 mm)

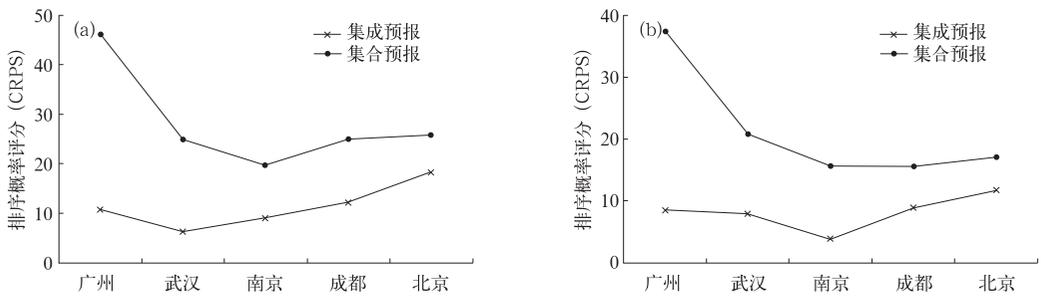


图 8 站点集成贝叶斯降水概率预报(集成预报)和集合直接概率预报(集合预报)的 6 月 CRPS 评分均值

(a) 基于 CMA 集合预报得到的结果, (b) 基于 NCEP 集合预报得到的结果

Fig. 8 Mean value of CRPS in June of integration Bayesian precipitation probability forecast (Integrated Forecast) and ensemble direct probability forecast (Ensemble Forecast) at the stations (a) based on CMA ensemble forecast, and (b) based on NCEP ensemble forecast

中,南京站的 CRPS 值最小,预报准确性最高。

综上所述,分别基于 CMA 和 NCEP 集合预报对广州、武汉、南京、成都和北京站的 2008 年 6 月逐日 24 h 降水建立降水概率预报模型并采用 CRPS 评分检验。结果表明,各站点的集成贝叶斯降水概率预报的可靠性均高于集合直接概率预报,并对广州站的概率预报准确性有较大提高,这种基于贝叶斯理论的集合降水概率预报方法经试验证明是科学合理的。

### 4 结 论

采用 1952—2007 年历史观测资料、2008 年集合预报资料对基于贝叶斯理论的集合降水概率预报方法进行初步应用试验,得到以下结论:

(1) 基于 BPO 方法对广州、武汉、南京、成都和北京站 6 月逐日有降水的先验概率  $g$  修订得到的后验概率  $\pi$  逐日变化明显,预报值提供的预报因子有

效信息发挥了修订作用。

(2)得到的单一集合成员贝叶斯降水概率预报与确定性预报相比,给出了预报范围内更多的预报信息,并以概率的形式定量化、连续地表达了降水预报的不确定性。

(3)每个集合成员的预报性能存在差异,以  $IS$  值表征集合成员的预报能力,结果表明集合成员预报能力随着预报时效增加而降低。用正比于  $IS^3$  的值作为权重系数可以有效地体现不同预报能力的集合成员其预报信息在融合集成中的地位。

(4)采用 CRPS 方法对 5 个研究站点的集成贝叶斯降水概率预报和集合直接概率预报检验,结果表明,基于 BPO 方法得到的集成预报可靠性高于集合预报,并对广州站的概率预报准确性有较大提高,该降水概率预报方法具有合理性。

文中采用双线性插值的方法由格点集合预报资料得到站点降水预报值,存在一定误差,对试验结果略有影响;试验样本长度仅为 1 个月,较小的样本量可能导致存在较大的预报模型偏差,其结果的代表性不是很强。另外值得注意的是,本文根据历史观测资料,采用威布尔分布估计得到的气候分布作为先验信息,其方法及结果存在一定局限性。由于先验信息的获取有多种途径,因此还可以尝试其他先验信息的估计方法,并与本文所得结果做进一步对比。下一步的工作将针对上述缺陷进行改进,对基于贝叶斯理论的集合降水概率预报方法进一步深入研究。

## 参考文献

陈法敬. 2011. 亚高斯贝叶斯预报处理器及其初步试验. 气象学报, 69(5): 872-882.  
陈法敬, 矫梅燕, 陈静. 2011. 一种温度集合预报产品释用方法的初步研究. 气象, 37(1): 14-20.

陈静, 陈德辉, 颜宏. 2002. 集合数值预报发展与研究进展. 应用气象学报, 13(4): 497-507.  
杜钧, 陈静. 2010. 单一值预报向概率预报转变的基础: 谈谈集合预报及其带来的变革. 气象, 36(11): 1-11.  
矫梅燕. 2010. 天气业务的现代化发展. 气象, 36(7): 1-4.  
茆诗松. 1999. 贝叶斯统计. 北京: 中国统计出版社, 12-30.  
王东海, 杜钧, 柳崇健. 2011. 正确认识和对待天气气候预报的不确定性. 气象, 37(4): 385-391.  
吴喜之. 2000. 现代贝叶斯统计学. 北京: 中国统计出版社, 10-35.  
赵琳娜, 梁莉, 王成鑫, 等. 2011. 基于贝叶斯模型平均的集合降水预报偏差订正. 第 28 届中国气象学会年会——S3 天气预报灾害天气研究与预报.  
Bishop C. 2008. Bayesian model averaging's problematic treatment of extreme weather and a paradigm shift that fixes it. Mon Wea Rev, 136(12): 4641-4652.  
Kelly K, Krzysztofowicz R. 1997. A bivariate meta-Gaussian density for use in hydrology. Stoch Hydrol Hydraul, 11(1): 17-31.  
Krzysztofowicz R. 1983. Why should a forecaster and a decision maker use Bayes theorem. Water Resour Res, 19(2): 327-336.  
Krzysztofowicz R. 1992. Bayesian correlation score: A utilitarian measure of forecast skill. Mon Wea Rev, 120(1): 208-220.  
Krzysztofowicz R. 1997. Transformation and normalization of variates with specified distributions. J Hydrol, 197(1-4): 286-292.  
Krzysztofowicz R. 1999a. Point-to-area rescaling of probabilistic quantitative precipitation forecasts. J App Meteorol, 38(6): 786-796.  
Krzysztofowicz R. 1999b. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. Water Resour Res, 35(9): 2739-2750.  
Krzysztofowicz R, H D Herr. 2001. Hydrologic uncertainty processor for probabilistic river stage forecasting: Precipitation-dependent model. J Hydrol, 249(1-4): 46-68.  
Lorenz E N. 1963. Deterministic nonperiodic flow. J Atmos Sci, 20(3): 130-141.  
Lorenz E N. 1965. A study of the predictability of a 28-variable atmospheric models. Tellus, 17(3): 321-333.  
Raftery A. 2005. Using Bayesian model averaging to calibrate forecast ensembles. Mon Wea Rev, 133(5): 1155-1174.