

张宇彤, 矫梅燕, 陈静. 2013. 基于模式先验信息的贝叶斯集合降水概率预报试验. 气象, 39(10): 1233-1246.

# 基于模式先验信息的贝叶斯集合 降水概率预报试验<sup>\* 1</sup>

张宇彤<sup>1,2</sup> 矫梅燕<sup>3</sup> 陈 静<sup>4</sup>

1 南京信息工程大学, 南京 210044

2 大连市气象局, 大连 116001

3 中国气象局, 北京 100081

4 中国气象局数值预报中心, 北京 100081

**提 要:** 为了更好地利用降水预报历史先验概率分布函数信息修订集合概率预报效果, 基于贝叶斯原理和贝叶斯降水概率预报模型, 分别使用 1952—2007 年历史观测资料和 2009—2011 年 6 月 24~120 h 中国 T213 全球集合预报历史资料作为先验信息, 对中国不同气候区代表站(广州、南京、武汉和成都)建立贝叶斯降水概率预报模型, 对比不同先验信息下集合成员与集成贝叶斯降水概率预报拟合结果差异, 分析先验信息对贝叶斯降水概率预报模型的影响, 在此基础上, 采用模式先验信息的贝叶斯降水概率预报模型, 进行 2008 年 6 月降水概率预报试验。试验结果表明, 由 T213 集合预报产生的先验信息较历史观测资料产生的先验信息更优, 当先验信息的降水概率分布函数曲率最大处偏向降水大值区时, 贝叶斯模型的降水预报结果也偏向降水大值区, 反之亦然。结果还显示, 先验信息对贝叶斯降水概率预报模型有重要影响, 若先验信息偏向更多更大降水量时, 贝叶斯降水概率预报对有降水的预报更优, 若降水先验信息偏向更少更小降水量时, 对无雨或微量降水预报效果越好。

**关键词:** 贝叶斯方法, 先验信息, 模式气候, 降水概率预报

**中图分类号:** P457

**文献标志码:** A

**doi:** 10.7519/j.issn.1000-0526.2013.10.001

## Bayesian Ensemble Probabilistic Forecasting Model Experiment of Precipitation Based on Model Priori Information

ZHANG Yutong<sup>1,2</sup> JIAO Meiyuan<sup>3</sup> CHEN Jing<sup>4</sup>

1 Nanjing University of Information Science and Technology, Nanjing 210044

2 Dalian Meteorological Service, Dalian 116001

3 China Meteorological Administration, Beijing 100081

4 CMA Numerical Prediction Centre, Beijing 100081

**Abstract:** In the rainfall probabilistic forecasting, information from the end of the probability distribution function should be applied. Then better forecast about rain and the ability of probabilistic forecasting will be developed. Based on the Bayesian theory and probabilistic forecasting model of precipitation, observational and T213 ensemble prediction data are used as the different priori information sources and experiments about precipitation in Guangzhou, Nanjing, Wuhan and Chengdu are carried out. Differences on the forecasts that are dependent on different priori informations are compared. Also influence of priori information on the Bayesian precipitation probabilistic forecast model is analyzed. And then the model is devel-

\* 国家自然科学基金面上项目(41075035)、公益性行业(气象)科研专项(GYHY200906007)、国家科技支撑计划项目(2009BAC51B00)和国家重点基础研究发展计划项目(2012CB417204)共同资助

2013 年 1 月 24 日收稿; 2013 年 5 月 2 日收修定稿

第一作者: 张宇彤, 主要从事集合预报, 贝叶斯理论方面的研究. Email: 13732531@qq.com

通信作者: 矫梅燕, 主要从事集合预报, 中小尺度天气方面的研究. Email: jiaomy@ema.gov.cn

oped as well. Furthermore, experiment based on Bayesian precipitation probabilistic forecasting which is dependent on model priori information is conducted. Results show that the priori information got from the ensemble prediction data is better. While the priori precipitation amount is more, the precipitation from the forecasting of model is also more and vice versa. Meanwhile, priori information has a strong impact on the Bayesian precipitation probabilistic forecast model. If the precipitation amount is much more, Bayesian ensemble probabilistic forecasting model can produce much more accurate prediction of rainy days. If the precipitation amount is little, prediction of no rain or light rain from the model is better.

**Key words:** Bayesian theory, prior information, model climate, precipitation probabilistic forecast

## 引 言

由于大气是一个高度非线性的混沌系统(Lorenz, 1965), 而模式的初始场只是大气真实状态的近似, 数值模式所描述的大气过程也并非真实的大气过程, 所以单一的确定性预报水平的提高已经变得越来越困难(邓国等, 2006; 王东海等, 2011), 概率预报成为天气预报发展的必然趋势。

随着天气预报技术发展, 集合预报方法的成功应用被认为是大气科学领域的一项重要成就(Sicililo et al, 1997)。集合预报是比单一确定性预报更加合理的、可以表达预报不确定性的天气预报形式(杜钧等, 2010; 矫梅燕, 2010; 穆穆等, 2011)。集合预报系统会产生大量数据, 为做出有气象意义的解释, 已提出很多的集合预报产品释用方法(陈静等, 2002; 毛恒青等, 2001), 如集合平均图、集合预报离散度、面条图、概率烟羽图、各种聚类方法等。而将贝叶斯理论应用于集合预报产品的解释中, 是近年来集合预报应用的一个新趋势。贝叶斯理论可将历史资料和集合预报系统产生的数据进行融合, 并给出连续的概率预报曲线作为预报结果。

贝叶斯方法的原理是将预报因子历史样本获得的先验信息与预报样本信息综合, 根据贝叶斯公式得到预报因子的后验信息。贝叶斯方法是与经典统计方法相对应的, 贝叶斯学派对于概率的理解上具有主观性, 在总体信息和样本信息的基础上还利用了先验信息, 而先验信息既可以是客观的, 也可以是主观的, 经典统计学派对于先验信息的忽略会造成信息利用的浪费(茆诗松, 1999)。贝叶斯方法作为一个新的学派, 在很多领域得到了广泛的应用(吴喜之, 2000)。国内外很多学者已经应用贝叶斯方法对水文和气象领域的预报结果进行订正。Adrian 等(2005)将统计模型贝叶斯模式平均(Bayesian

Model Averaging, BMA)运用到集合预报的修订中, 阐述了一种基于贝叶斯方法的集合预报后处理方法。Bishop 等(2008)通过研究发现 BMA 方法对于极端事件的预报并不是很准确, 因此他提出一种新的基于贝叶斯理论的集合预报后处理方法。他将估计得到的气候分布与贝叶斯理论结合, 从而将似然函数转换成后验分布。事实证明这种处理方法既可以提供可靠的非极端天气气候事件概率预报, 又可以提供可靠的极端天气气候事件概率预报。Krzysztofowicz 等(2006a)提出了针对二分类预报量即判断降水是否发生的贝叶斯产品处理(Bayesian processor of output, BPO)理论框架, 完善了二分类型预报量的 BPO 理论, 并做出检验(Krzysztofowicz et al, 2006b)。同年他提出了将连续型变量的贝叶斯处理器, 应用于温度和降水两个不同的预报量, 该方法是将单一数值预报模式扩展到集合概率预报方法——贝叶斯集合处理(Bayesian processor of ensemble, BPE)的理论来源。Krzysztofowicz 等(2008)在其先前研究贝叶斯理论的基础上, 提出了一种基于贝叶斯理论的预报方法——贝叶斯预报处理器(Bayesian processor of forecast, BPF)。

在国内, 贝叶斯方法被应用于不同的领域。张洪刚等(2005)基于贝叶斯方法, 提出并建立了实时洪水校正模型, 综合考虑水文资料和模型参数的不确定性, 得到预报流量的不确定性。殷志远等(2012)基于 ARME 模式, 引入贝叶斯统计模型对降水值进行修订, 使得洪水的预报精度在一定程度上有所提高。在气象领域, 马培迎(1999)较早应用贝叶斯原理修正了降水概率预报, 将天气分为有、无降水, 分别求得其期望概率作为先验概率, 引用贝叶斯原理, 提高了降水的预报精度。陈朝平等(2010)利用贝叶斯公式求得暴雨的后验分布, 并建立了暴雨预警模型。陈法敬等(2011)选取长沙和武汉站地面气温作为预报量, 利用 BPF 方法, 建立集合预报成

员的似然模型,并将单一集合成员预报转化为概率预报,得到预报能力高于单个成员的集成贝叶斯概率预报。

韩焱红等(2013)使用的历史观测资料作为先验信息的来源初步建立了贝叶斯降水概率预报模型。贝叶斯方法中的先验信息,既可以来自于主观经验,也可以来自于客观资料。但是,在气象研究中,不同先验信息对于贝叶斯模型的影响还没有较深入的研究。从韩焱红等(2013)研究中看到,先验信息对于集成贝叶斯概率预报的结果影响很大,在有降水的个例中,降水的预报结果普遍偏小,集成贝叶斯概率预报更多的是向气候概率预报结果靠近,不能充分体现概率预报在小概率事件上的预报优势。因此,为了更好修订降水的集合预报结果,分别使用历史观测资料和中国 T213 全球集合预报历史资料作为先验信息,对中国不同气候区代表站(广州、南京、武汉和成都)建立贝叶斯降水概率预报模型,分析先验信息对贝叶斯降水概率预报模型的影响,对贝叶斯降水预报模型进行优化,提高概率预报的修订能力。

### 1 方法和资料

#### 1.1 基于贝叶斯方法的降水集合产品概率化方法

##### 1.1.1 贝叶斯方法

由连续型预报量  $w$  和预报因子  $x$  组成的二维随机变量,在贝叶斯统计理论中,它们的概率密度函数存在如下关系:

$$\varphi(w | x) = \frac{h(x, w)}{k(x)} = \frac{f(x | w)g(w)}{k(x)} \quad (1)$$

式(1)即贝叶斯公式的密度函数形式,其中,  $h(x, w)$  是  $(x, w)$  的联合概率密度函数,  $g(w)$  是  $w$  的边缘概率分布函数,包含了预报量  $w$  的先验信息;  $f(x | w)$  表示的是在事件  $w$  发生的条件下  $x$  发生的概率密度函数;  $k(x)$  是  $x$  的边缘密度函数,且  $k(x) = \int_{-\infty}^{\infty} f(x | w)g(w)dw$ ; 后验概率密度函数  $\varphi$  表示的是使用预报因子对预报量的先验信息  $g(w)$  做出调整的结果。

##### 1.1.2 降水的亚高斯贝叶斯处理器

对于降水这一气象要素,降水有无的判断是一个二分类预报量的判断;而在降水发生的条件下,降水量则成为了一个连续型的气象要素。

设  $x$  为预报因子,  $y$  为预报量,  $\pi$  表示通过贝叶斯公式修订后的降水后验概率,  $\Phi(y)$  表示了条件降水量的概率分布的修订, Krzysztofowicz 等(2006a) 提出降水贝叶斯产品处理技术 BPO, 在 BPO 方法基础上, 发展了定量降水的概率预报 (Probabilistic Quantity Precipitation Forecast, PQPF), 图 1 是 PQPF 示意图, 运用式 (2a) 和式 (2b) 将后验概率与后验分布结合起来, 可以得到最终的降水概率化预报, 其中  $\delta$  为狄拉克函数。

概率分布形式

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

概率密度形式

$$p(y | X = x) = (1 - \pi)\delta(y) + \pi\phi(y), \quad y \geq 0 \quad (2b)$$

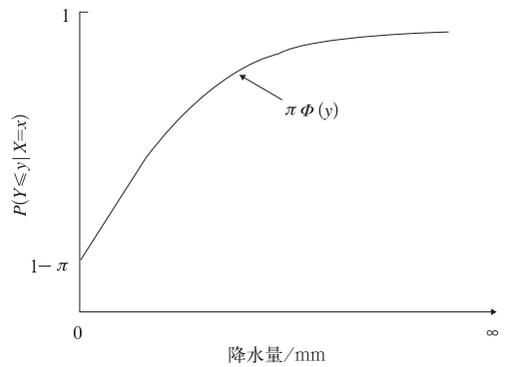


图 1 由降水发生概率和条件降水量分布函数组成的定量降水概率预报示意图  
Fig.1 The precipitation probabilistic quantity forecast diagram consisting of posterior probability and posterior distribution function

首先介绍后验概率  $\pi$  计算方法。根据 Krzysztofowicz 等(2006a) 提出的 BPO 原理, 首先对降水有无进行二分类判断, 设  $V$  为有无降水的气候样本,  $V=0$ , 代表降水不发生,  $V=1$ , 代表降水发生, 降水发生概率为先验概率, 设为  $g, g = P(V=1)$ 。  $x$  为预报因子(本文  $x$  取为集合预报成员降水预报量), 预报因子的概率密度函数  $f_v(x) = p(x | V)$ , 包含了两个条件概率密度函数  $f_0$  和  $f_1, f_0$  和  $f_1$  可从联合样本  $\{(x, V)\}$  中获得, 代表预报因子  $x$  和预报量  $V$  之间的随机相关性。利用先验概率  $g$  和条件概率密度函数  $f_0$  及  $f_1$ , 可得预报因子  $x$  的概率密度函数  $k(x)$ , 如式(3)所示:

$$k(x) = f_0(x)(1 - g) + f_1(x)g \quad (3)$$

根据贝叶斯定理,降水概率预报如式(4):

$$\pi = \frac{f_1(x)g}{k(x)} \quad (4)$$

将式(3)代入式(4),得到降水后验概率  $\pi$ :

$$\pi = \left[ 1 + \frac{1-g}{g} \frac{f_0(x)}{f_1(x)} \right]^{-1} \quad (5)$$

其次介绍条件概率分布  $\Phi(y)$  的计算方法(Krzysztofowicz et al,2006b)。设  $y$  为有降水条件下( $V=1$ )的降水气候样本,其概率密度函数即为先验概率密度分布, $x$  为预报因子,代入式(1)中,可获得连续预报量的贝叶斯概率预报结果,即  $\varphi(y|x) = \frac{f(x|y)g(y)}{k(x)}$ ,其中  $g(y)$  为降水的条件概率密度函数, $k(x)$  为  $x$  的概率密度函数。

由于降水这一气象要素并不符合二元高斯分布随机依赖结构,但符合亚高斯结构,需要将  $G(y)$  (条件累积概率分布函数)和  $K(x)$  ( $x$  的累积概率函数)进行正态分位数转换(normal quantile transform, NQT)(Kelly et al,1997;Krzysztofowicz,1997),转换后, $U=Q^{-1}G(y)$ , $Z=Q^{-1}K(x)$ ,其中  $Q^{-1}$  为正态分布函数的反函数, $U$  和  $Z$  均服从正态分布(如图 2 所示),以广州站 2008 年 6 月降水为例,图 2a 表示了降水量的控制预报与观测值的随机依赖结构,散

点分布图为联合样本  $x$  与  $y$  的分布形式,对于预报因子  $x$  和预报量  $y$ ,其边缘密度函数是偏态分布函数的。图 2b 是经过 NQT 转换后的变量,点线为  $y$  经 NQT 转换后的变量  $U$  的概率密度曲线,虚线为  $x$  经 NQT 转换后变量  $Z$  的概率密度曲线,可见,转换后的变量均符合正态分布,因此, $U$  与  $Z$  在转换空间内基本符合线性回归,具有线性同方差形式的随机依赖结构。

在转换空间内, $\mu_1$  和  $\mu_0$  分别为预报量和预报因子转换后  $U$  和  $Z$  的均值, $\sigma_1$ 、 $\sigma_0$  分别为  $U$  和  $Z$  的均方差, $\sigma_{10}$  为  $U$  和  $Z$  的协方差。且满足  $E(Z|U=u) = au + b$ ,  $\text{Var}(E|U=u) = \sigma^2$ ,似然参数  $a$ 、 $b$  和  $\sigma^2$  以及后验参数  $c_1$ 、 $c_0$  和  $T$  的结果由式(6a)和式(6b)给出:

$$a = \frac{\sigma_{10}}{\sigma_0^2}, \quad b = \mu_1 - \frac{\sigma_{10}}{\sigma_0^2} \mu_0, \quad \sigma^2 = \sigma_1^2 - \frac{\sigma_{10}^2}{\sigma_0^2} \quad (6a)$$

$$c_1 = \frac{a}{a + \sigma^2}, \quad c_0 = \frac{-ab}{a^2 + \sigma^2}, \quad T = \left( \frac{\sigma^2}{a^2 + \sigma^2} \right) \quad (6b)$$

转换后的变量再通过 NQT 逆转换和雅克比式将转换空间中的概率预报公式转换回原始空间,得到原始预报量的条件概率分布式(7a)和式(7b)。

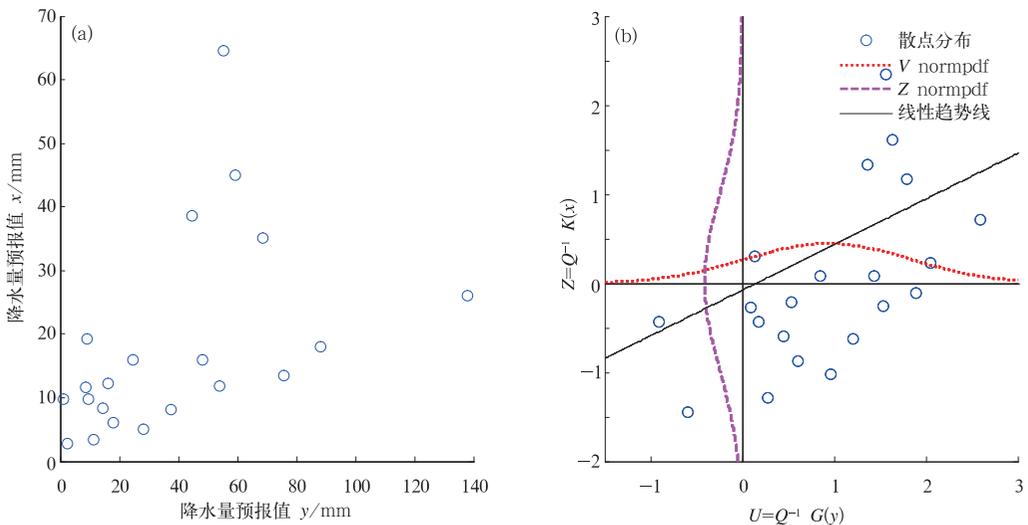


图 2 广州站 2008 年 6 月降水值的 NQT 转换示意图

(a) 转换前降水量的预报值和观测值的散点分布, (b) NQT 转换后的预报值和观测值的散点分布 (实线为线性趋势线,  $Q^{-1}$  为正态分布函数的反函数)

Fig. 2 Normal quantile transform (NQT) diagram of precipitation at Guangzhou Station in June 2008 (a) scatter diagram of observation and forecasting precipitation before NQT, (b) scatter diagram of observation and forecasting precipitation after NQT (Solid line is linear trend,  $Q^{-1}$  is the inverse function of normal distribution)

后验概率累积分布形式

$$\Phi(y) = Q\left\{\frac{1}{T}[\mathbf{Q}^{-1}G(y) - C_1\mathbf{Q}^{-1}K(x) - C_0]\right\} \quad (7a)$$

后验概率密度形式

$$\phi(y) = \frac{1}{T}\exp\left(\frac{1}{2}\{[\mathbf{Q}^{-1}G(y)]^2 - [\mathbf{Q}^{-1}\Phi(y)]^2\}\right)g(y) \quad (7b)$$

将上文中式(5)和式(7a)的结果代入式(2a)中,就可以得到贝叶斯降水概率预报的分布形式,通过这种方法,可以将确定的降水预报结果修订为一个连续的概率预报,为降水的预报提供更多的信息。

## 1.2 有效信息评分 (Informativeness Score, IS)

在亚高斯模型中,选取有效信息评分来代表预报因子包含的有效信息(Krzysztofowicz,1992)。预报因子  $x$  的有效信息主要通过预报因子  $x$  和预报量  $y$  的依赖程度来判断。通过正态分位数转换(NQT)后, $x$  和  $y$  的随机依赖性分别由它们的转换变量  $Z$  和  $U$  的皮尔逊相关系数  $\gamma$  来表征。由此定义有效信息评分  $IS$  的值:

$$IS = |\gamma| = \left[\left(\frac{\alpha}{\sigma}\right)^{-2} + 1\right]^{-\frac{1}{2}} \quad (8)$$

式中,参数  $\alpha$  和  $\sigma$  的含义与式(6a)中的相同。这里, $IS \in (0, 1)$ ,当  $IS=1$  时,说明预报因子的预报效果最佳。

## 1.3 资料

本文选取了广州、南京、武汉和成都 4 个代表测站进行讨论,降水的历史观测资料使用国家气象中

心提供的全国基准站 1952—2007 年的 6 月逐日 20—20 时的 24 h 观测降水量,降水的模式历史资料采用中国气象局(CMA)数值预报中心 T213 的 2009—2011 年 6 月逐日 24~120 h 预报时效的降水集合预报资料。以站点的 24 h 降水量作为预报量  $y$ ,以 2008 年各测站 6 月 24~120 h 的 CMA 集合成员作为预报因子  $x$ 。

## 2 两种先验信息下的贝叶斯降水概率预报模型

### 2.1 两种先验信息的选取

贝叶斯方法中的关键一步是确定先验分布(茆诗松,1999)。当对降水进行二分类事件进行判断时,先验概率即是降水发生概率;而当降水发生时,就需要构造一个连续的先验概率密度函数。韩焱红等(2013)认为威布尔分布(Weibull Distribution)是对降水量拟合的最佳分布函数,所以我们选取该分布函数对条件降水量进行拟合,再确定威布尔分布中的参数。

为了分析不同先验信息对贝叶斯降水概率预报模型的影响,本文设计两种先验信息选取方案(如表 1 所示),先验信息方案 1 选取历史观测资料作为的先验信息来源,代表站来自中国不同气候区域(广州、南京、武汉和成都),降水阈值为 0.1 mm,以下简称观测先验信息;先验信息方案 2 选取中国气象局数值预报中心 T213 集合预报资料,计算所选站点周围的 4 个格点平均降水量,代表该站点的降水量,降水阈值仍为 0.1 mm,以下简称模式先验信息。

表 1 观测先验信息与模式先验信息方案设计

Table 1 Design of observed and model prior information plan

	观测先验信息方案	模式先验信息方案
站点	广州、武汉、南京和成都	广州、武汉、南京和成都
阈值	0.1 mm ( $\geq 0.1$ mm 视为有降水)	0.1 mm
先验概率 $g$	近 50 年观测资料中,6 月各测站发生降水的概率	2009—2011 年 6 月 CMA T213 集合预报各测站发生降水的概率
先验分布 $G(w)$	近 50 年观测资料中,6 月各测站降水发生情况下的条件概率分布	2009—2011 年 6 月 CMA T213 集合预报各测站降水发生情况下的条件概率分布

由于不同的气候区观测和模式预报的降水有很大差异,所以先验信息也不同,下面通过 2009—2011 年 6 月 24 h 降水的先验概率和先验分布讨论先验信息在不同测站的差异特征。

#### 2.1.1 先验概率 $g$

先验概率  $g$  表示降水发生的概率。对于观测先验信息,由 1952—2007 年 6 月历史资料,可获得 56 个降水样本,产生站点的观测先验概率。对于模

式先验信息,每日模式产生 15 个集合预报成员,2009—2011 年 6 月总计有 45 个模式降水量样本,由此获得站点的模式先验概率。

图 3 是 6 月广州、南京、武汉和成都的逐日观测先验概率和模式先验概率对比图。由图 3a 可见,广州的观测先验概率在 50%~70%左右;南京站观测先验概率在 20%~50%左右,且上半月概率小于下半个月的,这主要是由于 6 月下旬江淮流域开始进入梅雨期,降水量增加。再看模式先验概率,广州的

模式先验概率明显大于观测的,这是由于模式对广州降水预报量级偏大造成的,降水大于 0.1 mm 时,即被认为有降水发生,表明模式预报存在降水偏多的趋势;从图 3b 可见,南京的模式先验概率逐日变化较大,且下半月的降水概率总体大于上半月的。武汉和成都站的模式先验概率逐日变化特征与广州和南京相似,起伏较大(图 3c 和 3d),其他预报时效的模式先验概率亦有类似特征(图略),不再赘述。总体来看,模式先验概率略大于观测先验概率。

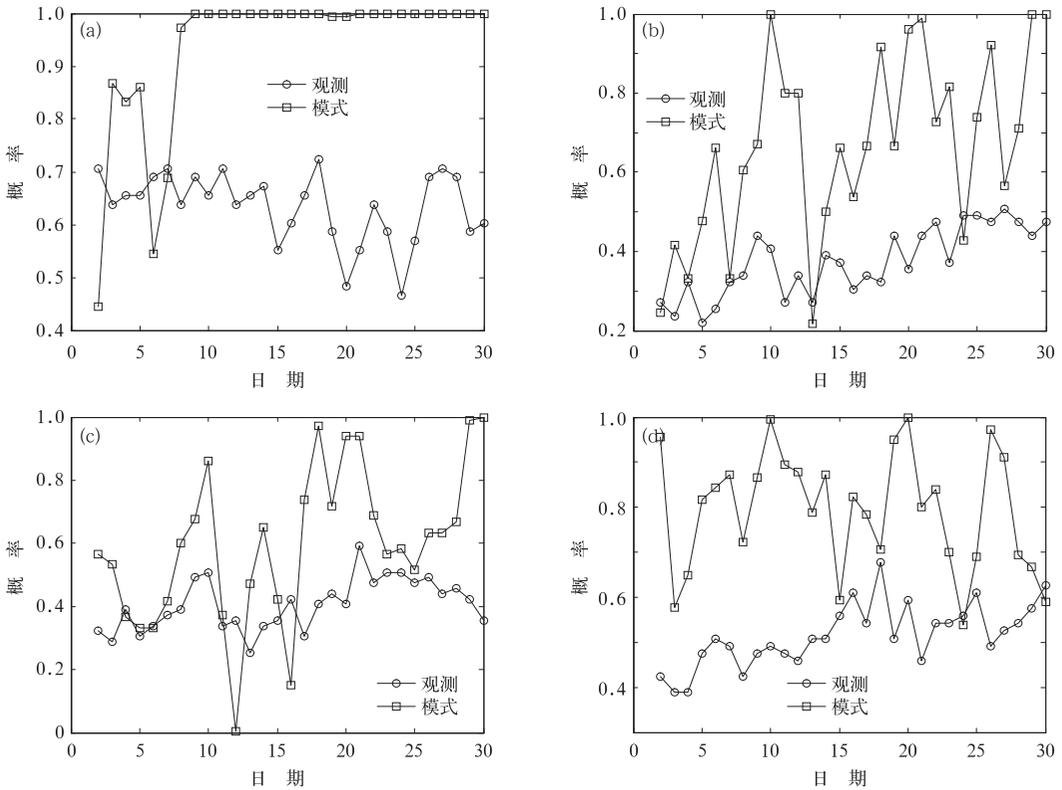


图 3 6 月观测先验概率和模式先验概率的对比图

(a)广州,(b)南京,(c)武汉,(d)成都

(圆点:观测先验概率;方块:模式先验概率,24 h 预报时效)

Fig. 3 Comparison of observation and forecasting priori informations in June

(a) Guangzhou, (b) Nanjing, (c) Wuhan, (d) Chengdu

(circle: observation priori information; box: model priori information, 24 h forecast)

### 2.1.2 先验概率分布 G

先验概率分布 G 表示了降水发生条件下的降水累积分布特征,将 24~120 h 预报时效的模式条件降水和观测条件降水用威布尔分布[如式(9)所示]进行拟合,再确定威布尔分布中的参数,其中,α 是形状参数,β 是尺度参数,表 2 为各测站观测和模式条件降水量的威布尔分布拟合参数。通过对比条

件降水量的观测与模式的累积概率先验分布,来探讨上述 4 个代表站在两种先验信息下的条件概率分布特征。

$$f(x | \alpha, \beta) = e^{-\left(\frac{x}{\beta}\right)^\alpha} x^{\alpha-1} \alpha \beta^{-\alpha} \quad (9)$$

图 4 是条件降水量的观测与模式的累积概率先验分布对比图。由图 4 可见,就广州而言,当降水量较小(大)时,观测条件降水量累积概率分布曲线位

于模式各预报时效分布曲线之上(下),即观测降水多为小雨,而模式降水多为中雨和大雨,表明模式高估降水强度;再看南京和武汉(见图 4b 和 4c),6 月降水量不多,观测条件降水量与模式条件降水量的先验分布相似,表明模式预报与观测量级相当;成都(图 4d)的观测条件降水量曲线略高于模式条件降

水量曲线,表明模式低估降水强度。

从上可知,观测和模式的先验信息具有不同的特点。在下一节中,分别将这两种先验概率和先验分布输入贝叶斯降水概率预报模型中,分析后验参数和后验分布的变化,讨论集合成员的贝叶斯降水概率预报结果。

表 2 各测站观测和模式条件降水量的威布尔分布拟合参数

Table 2 Weibull parameters of observed and model conditional precipitation at different stations

	广州		南京		武汉		成都	
	$G(w)$	24 h	$G(w)$	24 h	$G(w)$	24 h	$G(w)$	24 h
$\alpha$	12.29	13.72	10.22	11.46	11.63	10.19	4.88	6.83
$\beta$	0.69	1.20	0.61	0.62	0.59	0.72	0.63	0.82

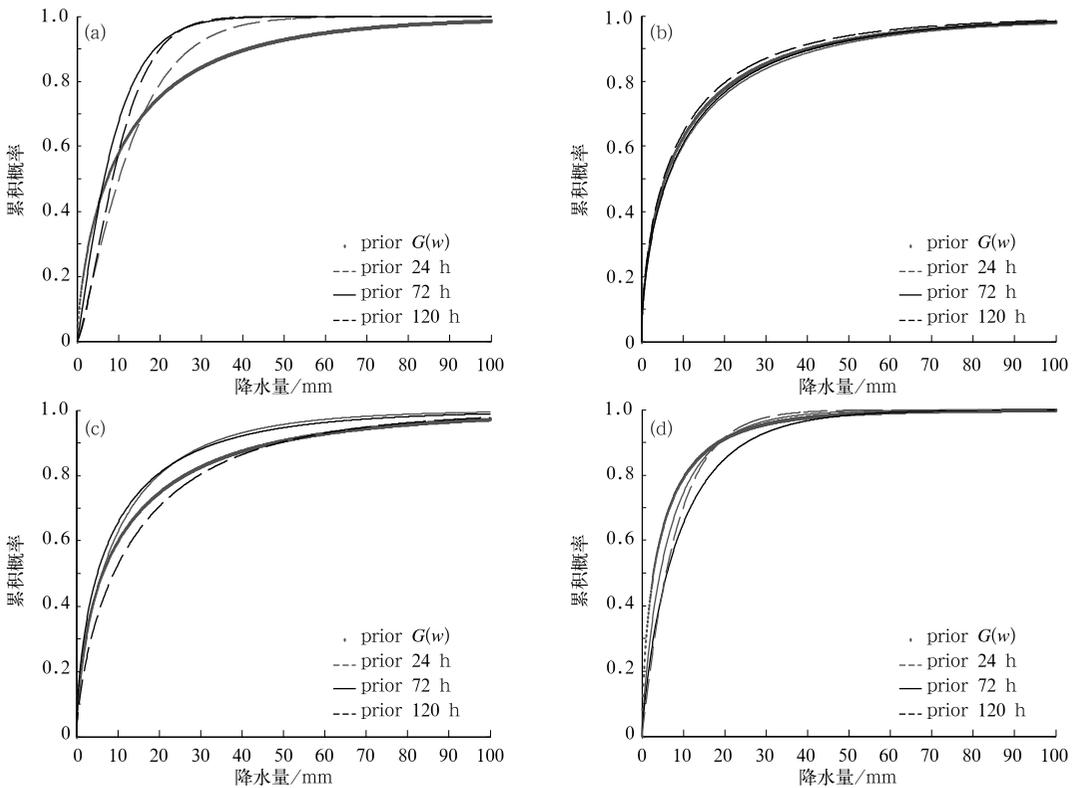


图 4 条件降水量的观测先验累积概率分布与  $G(w)$  与 24 h(实线)、72 h(虚线)

和 120 h(点划线)预报时效的模式先验累积概率分布对比图

(a)广州, (b)南京, (c)武汉, (d)成都

Fig. 4 Cumulative probabilistic distribution of observation priori probability (thick solid line) and model priori probability based on 24 h (solid line), 72 h (dashed line), and 120 h (dash-dotted lines) forecast

(a) Guangzhou, (b) Nanjing, (c) Wuhan, (d) Chengdu

## 2.2 后验分布函数

以 T213 集合预报 24 h 累积降水的控制预报为例,利用式(7a)得到后验分布函数,为了简洁起见,设式(7a)中

$A = [C_1 Q^{-1} K(x) + C_0]$ ,  $f(y) = Q^{-1} G(y)$   
则式(7a)简化表达为

$$\Phi(y) = Q \frac{1}{T} [f(y) - A] \quad (10)$$

式中, $Q$  为正态分布函数, $T$  由式(6b)求得。 $A$  中的

$C_0$  和  $C_1$  与式(7a)中参数值相同,  $T$  表示该函数的方差,  $T$  越大(小), 降水的后验概率分布函数越平缓(陡峭);  $A$  表示该函数的均值,  $A$  值越大(小), 降水概率分布函数的平均值越大(小), 由式(7a)中的  $C_0$  和  $C_1$  确定。表 3 为各测站 2008 年 6 月观测先验信息方案与模式先验信息方案中后验参数值, 其中  $C_0$  和  $C_1$  决定了降水后验概率分布函数的平均状况。由比较可知, 模式先验信息方案的  $T$  值大于观测先验信息方案中的, 因此模式先验信息方案中降水概率分布函数曲线更陡峭, 观测先验信息方案的则更平缓; 模式先验信息方案中  $C_0$  小于 0,  $C_1$  大于 1, 观

测先验信息方案中  $C_0$  大于 0,  $C_1$  小于 1,  $A$  值无法直接估算, 但具体特征可从图形更直观展现, 以广州 2008 年 6 月 8 日的降水情况为例(如图 5 所示), 图 5 中曲线由式(10)确定, 由观测与模式后验分布的累积概率函数曲线可知, 由模式先验信息确定的后验分布函数(图 5b)较观测先验信息的(图 5a)更为陡峭, 且整体略偏向于右侧, 说明模式先验信息确定的后验概率分布函数均值更大。对于其他测站及其他集合成员后验分布函数也有类似的特征, 这里不再赘述。

表 3 观测先验信息与模式先验信息方案下的后验参数值比较

Table 3 Parameters comparison of observed and model prior informations

站点	$C_0$		$C_1$		$T$	
	观测先验信息方案	模式先验信息方案	观测先验信息方案	模式先验信息方案	观测先验信息方案	模式先验信息方案
广州	0.3062	-0.1220	0.5915	1.2106	0.8110	0.7913
南京	0.1114	-1.5701	0.6107	1.5033	0.6887	0.4112
武汉	0.0293	-2.0962	0.3855	1.7716	0.8918	0.4138
成都	0.2330	-1.9264	0.6369	1.8419	0.6788	0.5074

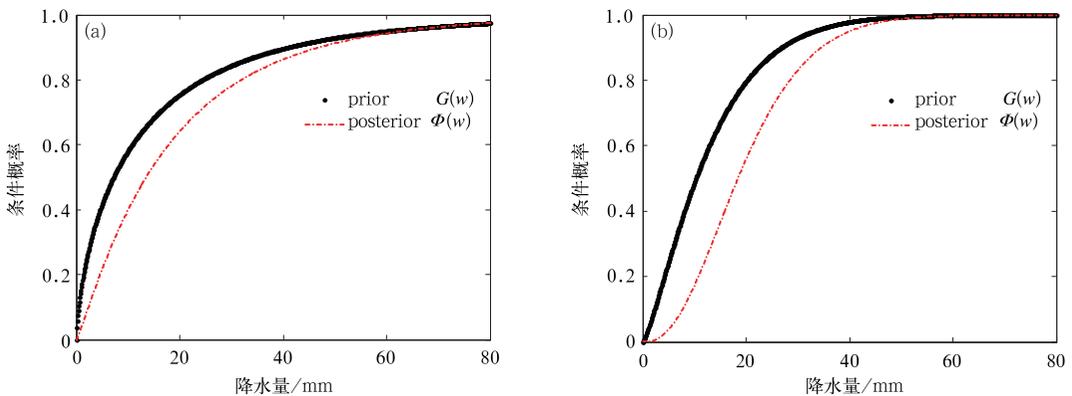


图 5 广州站 2008 年 6 月 8 日后验分布累积概率函数  $\Phi(w)$  为后验分布函数,  $G(w)$  为先验分布函数  
(a) 观测先验信息方案, (b) 模式先验信息方案

Fig. 5 Cumulative function of posterior distribution in Guangzhou on 8 June 2008

$\Phi(w)$  is posterior distribution function,  $G(w)$  is priori distribution function

(a) observation priori information scheme, (b) model priori information scheme

### 2.3 不同先验信息下的集合成员贝叶斯降水概率拟合预报

前文主要介绍了两种先验信息来源以及不同先验信息对于后验分布函数的影响。在这一部分中, 将模式和观测先验信息应用到贝叶斯降水概率化模型中, 分析先验信息对于该模型的影响。将每一个集合成员的预报值, 应用 BPO 模型, 获得一组降水

概率拟合预报。为了更清晰地展示集合成员的概率拟合预报结果, 我们随机选取第 1、3、9、13 个集合预报成员的概率拟合预报结果进行讨论(图 6 所示), 对于所选的集合成员的贝叶斯概率预报结果, 由文中式(2a)确定, 式中  $\pi$  与  $\Phi(y)$  分别在 1.1.2 和 2.2 中给出。

图 6 是 2008 年 6 月 8 日广州、南京、武汉和成都四站基于两种先验信息的贝叶斯集合降水累积概

率分布对比图。观测数据显示,南京和成都无降水,广州和武汉的降水量观测值分别为 24.8 和 17 mm。对比两种先验信息方案可知,在模式先验信息方案的预报试验中,无论有雨还是无雨日,降水概率均大于观测先验信息方案的,这主要是受模式先验概率较大的影响所致,还可看到,广州和武汉各集合成员的累积概率曲线曲率最大处,对应的降水量

正接近观测降水值,对于南京和成都,集合成员的累积概率曲线曲率最大处对应的降水量为小雨区。在观测先验信息方案的预报试验中,各站的集合成员累积概率曲线曲率最大值对应的降水量为无雨和小雨区。可见,模式先验信息获得的概率预报,对有降水的预报更优。

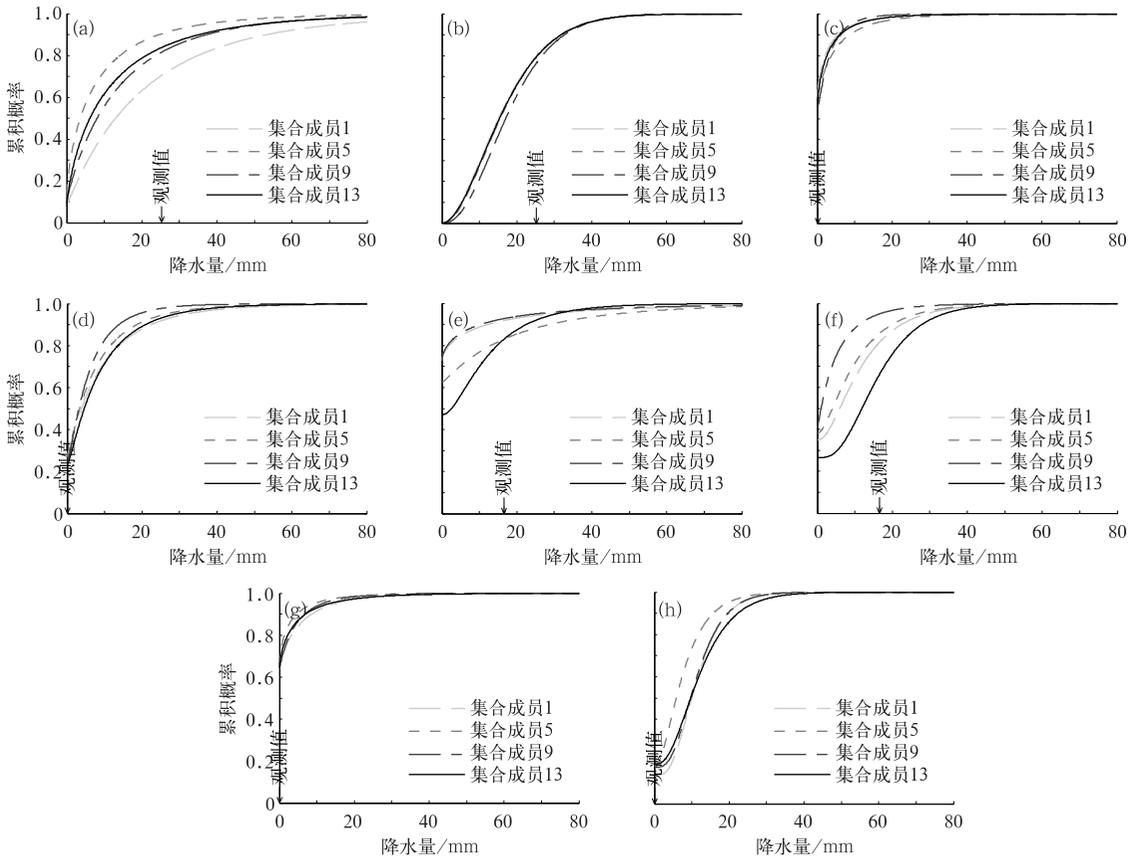


图 6 2008 年 6 月 8 日广州、南京、武汉和成都四站基于两种先验信息的贝叶斯集合降水累积概率分布(1,5,9,13 集合成员)  
(a,c,e,g)观测先验信息方案,(b,d,f,g)模式先验信息方案

Fig. 6 Cumulative probability distribution of Bayesian precipitation probabilistic forecast based on ensemble member (the 1st,5th,9th and 13th) in Guangzhou, Nanjing, Wuhan and Chengdu on 8 June 2008

(a,c,e,g) observation priori information scheme, (b,d,f,g) model prior information scheme

## 2.4 两种先验信息下的集成贝叶斯降水概率拟合预报

### 2.4.1 集合成员的集成

在 2.3 中,应用贝叶斯降水模型得到了各集合成员的贝叶斯降水概率拟合预报,根据 1.2 中有效信息评分的定义以及各测站集合成员概率预报结

果,由式(6)计算各测站在两种先验信息下各集合成员的有效信息评分  $IS$  值。

图 7 以 24、72 和 120 h 预报时效为例,给出了南京在观测和模式先验信息方案下集合成员  $IS$  值对比图,由图 7 可见,在观测先验信息方案下,24 h 预报时效的  $IS$  值在 0.5~0.9 区间内,而在模式先验信息下,24 h 预报时效的  $IS$  值在 0.8~0.9 区间

内变化,随预报时效的延长,各成员  $IS$  值呈减小趋势。对于广州、武汉和成都(图略),模式先验信息方案较观测先验信息方案的  $IS$  值均有一定程度的提

高,其值更接近于 1,而  $IS$  值越接近于 1 预报效果越好,说明在模式先验信息方案下,集合成员包含的有效预报信息更多。

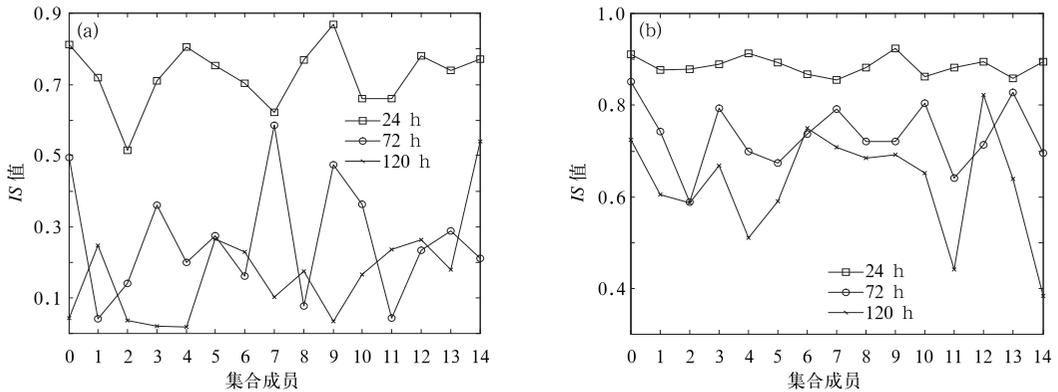


图 7 南京站的 24、72 和 120 h 集合降水概率预报  $IS$  值对比图

(a) 观测先验信息方案, (b) 模式先验信息方案

Fig. 7  $IS$  value of ensemble members based on two priori information of Nanjing

(a) observation priori information scheme, (b) model priori information scheme

陈法敬等(2011)根据不同集合成员的有效信息评分,以  $IS^3$  为权重,定义了  $r_i(IS_i)$  作为各集合成员的权重系数,如式(10)所示,对于  $IS$  值越大的成员,其权重系数也越大。对于  $i(i=15)$  个集合成员的贝叶斯降水概率预报  $p(y|x_i)$ ,根据  $r_i$  进行信息融合,可以得到集成贝叶斯降水概率预报的结果,公式(11)给出了集成贝叶斯降水概率预报的概率密度形式。

$$r_i(IS_i) = \frac{IS_i^3 - \min(IS^3)}{\sum_{i=1}^n IS_i^3 - n \cdot \min(IS^3)} \quad (11)$$

$$p(y|\mathbf{X}) = \frac{\sum_{i=1}^n p(y|x_i) \cdot r_i(IS_i)}{\int_0^{\infty} \left[ \sum_{i=1}^n p(\xi|x_i) \cdot r_i(IS_i) \right] d\xi} \quad (12)$$

#### 2.4.2 两种先验信息的集成贝叶斯降水概率拟合预报比较

根据式(11)确定的集合成员的贝叶斯降水概率预报的权重系数,代入式(12)中,对 2.3 中各测站集合成员贝叶斯概率预报结果进行集成,得到两种先验信息下的集成贝叶斯降水概率预报,如图 8 所示。广州(图 8a 和 8b)和武汉(图 8e 和 8f)降水观测值均为中雨,分别为 24.7 和 17 mm,模式先验信息方案降水发生概率分别为 100% 和 70%,明显大于观测先验信息方案的。模式先验信息方案(图 8b 和

8f)的降水累积概率预报曲线的曲率最大值,对应的降水量分别为 12.3 和 11.7 mm,位于中雨区,更接近于观测值;而观测先验信息方案(图 8a 和 8e)降水累积概率预报曲线的曲率最大值处,对应的降水量分别为 2.0 和 4.7 mm,位于小雨区。南京(图 8c 和 8d)和成都(图 8g 和 8h)的降水量观测值均为 0,模式先验信息方案的降水发生概率仍然大于观测先验信息方案,两种先验信息方案的累积概率预报曲线的曲率最大值处,对应的降水量均位于无雨和小雨区。由此可见,模式先验信息方案对于有降水个例的预报效果更佳,对于无降水的个例,模式先验信息方案修订后的降水发生概率略偏大。

对于其他个例来说,两种方案下的集成贝叶斯降水概率预报结果与以上个例相似,不再赘述。下面将利用模式先验信息对 2008 年 6 月 4 个代表站的降水量进行预报试验并检验。

### 3 模式先验信息贝叶斯降水概率预报模型的预报试验

#### 3.1 2008 年 6 月集成贝叶斯降水概率预报结果的检验

利用上节建立的模式先验信息贝叶斯降水概率预报模型,进行 2008 年 6 月的降水概率预报试验。

采用排序概率评分 (continuous ranked probability score, CRPS) 方法, 对 2008 年 6 月贝叶斯降水概率预报和集合预报进行检验。对于预报量  $y$ , CRPS 的值由式(13)确定, 其值越小, 预报结果与观测更接近。

$$CRPS = \int_{-\infty}^{\infty} [F(y) - F_0(y)]^2 dy \quad (13)$$

表 4 给出了 2008 年 6 月 4 个代表站 24~72 h 贝叶斯降水概率预报 CRPS 平均值, 可以看出, 南京 24、48 和 72 h 预报时效的贝叶斯降水概率预报的 CRPS 检验值分别为 4.14、4.49 和 5.23, 而集合预报的 CRPS 检验值相应为 7.83、5.40 和 6.56, 后者均大于前者, 表明南京的贝叶斯降水概率预报效果更好; 广州和武汉 24 h 预报时效贝叶斯降水概率

预报的 CRPS 检验值分别为 17.27 和 3.65, 而集合预报的 CRPS 检验值分别为 17.60 和 3.69, 广州和武汉的贝叶斯降水概率预报对 24 h 降水预报效果更好, 集合预报对其他预报时效预报效果更好。成都站的贝叶斯模型 24 h 预报时效改进效果最不明显, 集合预报的 CRPS 检验值略小, 但 48 和 72 h 预报时效的 CRPS 值均大于集合预报的。可见, 广州、南京和武汉 24 h 集成贝叶斯降水概率预报, 以及成都的 48 和 72 h 集成贝叶斯降水概率预报的技巧均有一定的提高。

### 3.2 集成贝叶斯降水概率预报结果

下面以广州站为例, 对典型降水时段进行详细分析, 对比贝叶斯降水概率预报与集合预报结果。

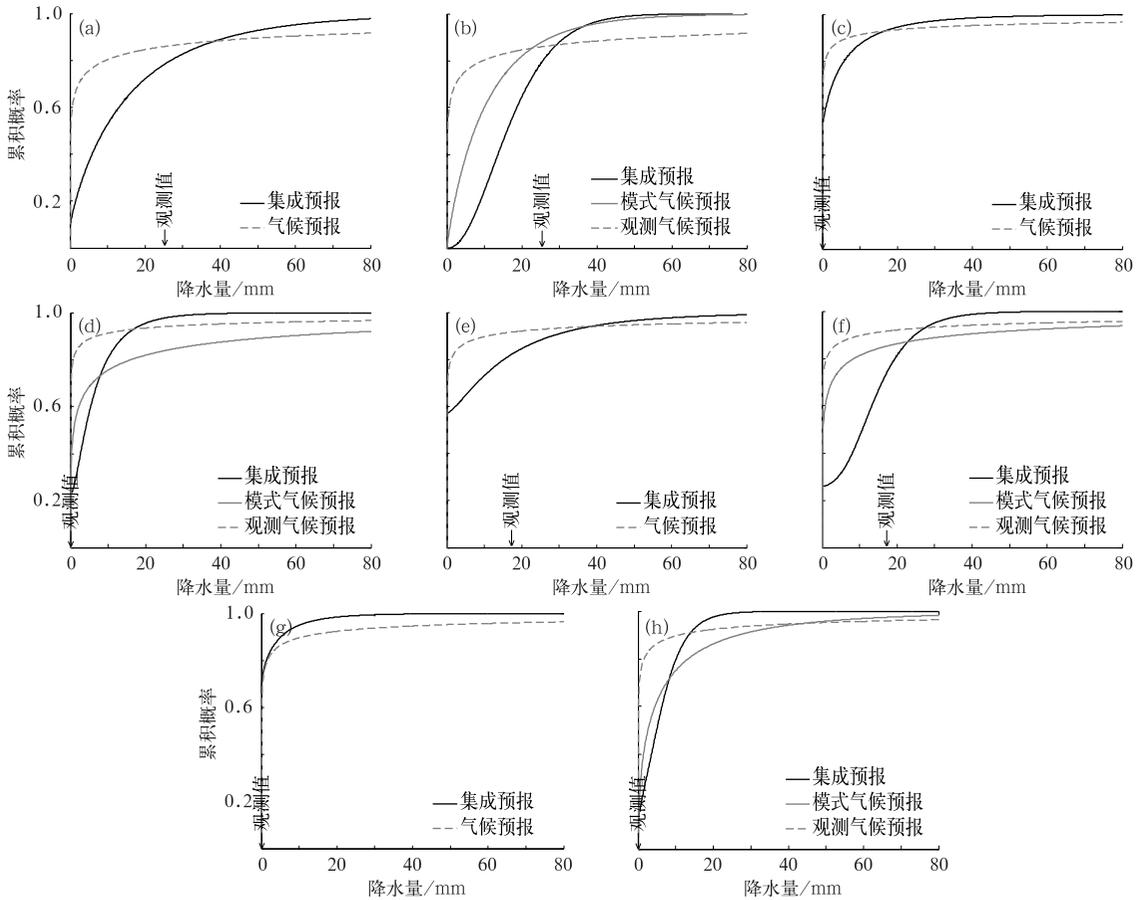


图 8 2008 年 6 月 8 日广州 (a,b)、南京 (c,d)、武汉 (e,f) 和成都 (g,h) 4 站基于 2 种先验信息的集成贝叶斯降水累积概率分布

(a,c,e,g) 观测先验信息方案, (b,d,f,h) 模式先验信息方案

Fig. 8 Cumulative probability distribution of integrated Bayesian precipitation probabilistic forecast based on ensemble member (the 1st, 5th, 9th and 13th) in Guangzhou (a, b), Nanjing (c, d), Wuhan (e, f) and Chengdu (g, h) on 8 June 2008

(a,c,e,g) observation priori information scheme, (b,d,f,h) model prior information scheme

表 4 2008 年 6 月 4 个代表站 24~72 h 贝叶斯模型与集合预报的降水概率预报 CRPS 平均值

Table 4 Average CRPS of Bayesian and ensemble rainfall prediction in June 2008, 24 h to 72 h forecasting

		广州	南京	武汉	成都
24 h 预报时效	Bayesian CRPS	17.272	4.138	3.649	3.898
	Ensemble CRPS	17.597	7.831	3.688	2.888
48 h 预报时效	Bayesian CRPS	18.355	4.494	4.700	4.394
	Ensemble CRPS	16.113	5.396	4.168	4.546
72 h 预报时效	Bayesian CRPS	18.979	5.230	4.479	4.735
	Ensemble CRPS	17.622	6.562	4.335	6.087

3.2.1 24 h 概率预报结果分析

图 9 和图 10 分别是 2008 年 6 月 12 日(观测为中雨的个例预报)和 2008 年 6 月 2 日(观测为大雨的个例预报)贝叶斯降水概率预报结果。首先看中雨的个例,6 月 12 日降水观测值为 11.3 mm,从图 9a 和 9b 可见,集合成员预报降水均为小雨,概率密度曲线上在对应观测值处概率密度值为 0,而贝

叶斯降水概率预报的概率密度曲线有明显的波峰,峰值为 0.0529,波峰所对应降水量为 8.7 mm,接近观测值,这说明贝叶斯降水概率预报曲线对降水预报起到明显的修订作用。广州 6 月 2 日的观测值为 28.3 mm,达大雨量级,与中雨的预报类似,集合成员的预报结果主要集中在小雨和中雨量级,在观测值处,集合成员的概率密度值为 0,而修订后的预报

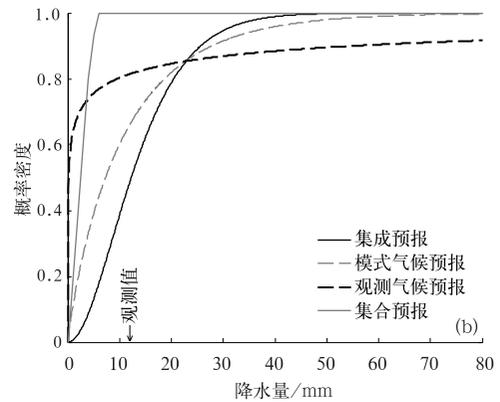
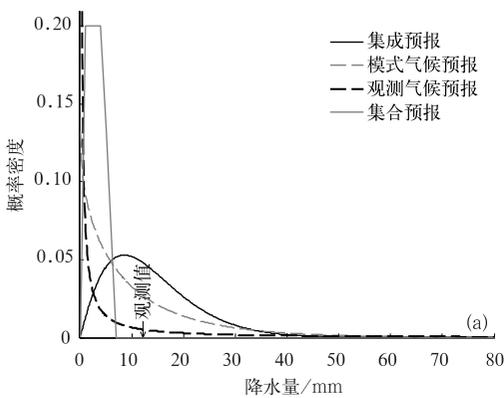


图 9 2008 年 6 月 12 日广州站贝叶斯降水概率预报结果

(a) 概率密度分布, (b) 累积概率分布

Fig. 9 Bayesian precipitation probabilistic forecast in Guangzhou on 12 June 2008

(a) probabilistic density, (b) cumulative distribution

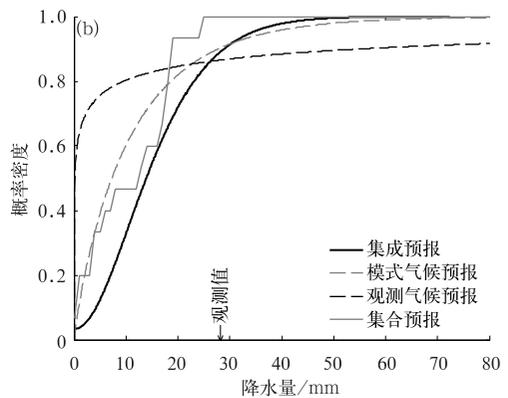
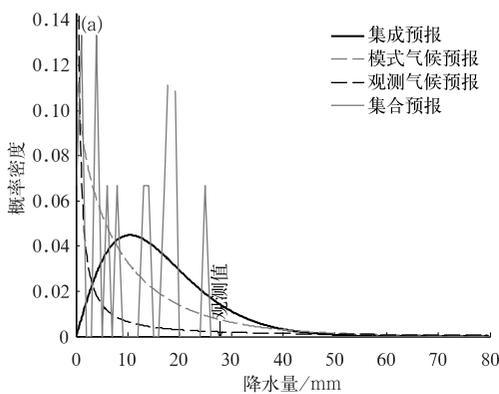


图 10 同图 9, 但为 2008 年 6 月 2 日预报

Fig. 10 Same as Fig. 9, but on 2 June 2008

在观测值处有相应的概率密度值,为 0.0137。总体来说,贝叶斯降水概率预报在一定程度上修订了集合预报对中雨及大雨预报量偏低的结果,提高了降水的预报能力。

由 2.4.1 中集合成员各预报时效的有效信息评分可知,预报时效为 24 h 时,预报因子包含的有效信息最大,但其他预报时效也提供了一定的预报依据,下文将对不同预报时效贝叶斯降水概率模型预报结果进行讨论。

### 3.2.2 不同预报时效的概率预报结果分析

以 2008 年 6 月 12 日广州站降水预报为例,图 11 是广州站 2008 年 6 月 12 日贝叶斯降水概率密度预报曲线图,广州降水观测值为 11.3 mm,而

各集合成员在各预报时效预报有雨,但预报多为小雨,24 和 48 h 预报时效的集合平均值分别为 3.6 和 6.37 mm,通过贝叶斯降水模型修订后,24 和 48 h 概率密度预报曲线峰值对应的降水量分别为 8.7 和 6.9 mm,与观测值更接近;而 72 和 96 h 集合成员预报结果出现了多个峰值,表明预报不确定性较大,而贝叶斯降水概率密度预报曲线则相对平缓。对比各时效修订后的结果可见,预报时效 24 h 的贝叶斯降水概率预报结果好于其他时效,表明基于模式先验信息的贝叶斯降水概率预报结果在很大程度上依赖于集合预报的准确性,集合预报结果越准确,修订后结果也越准确;随着预报时效延长,预报不确定性增加,预报准确率也有所下降。

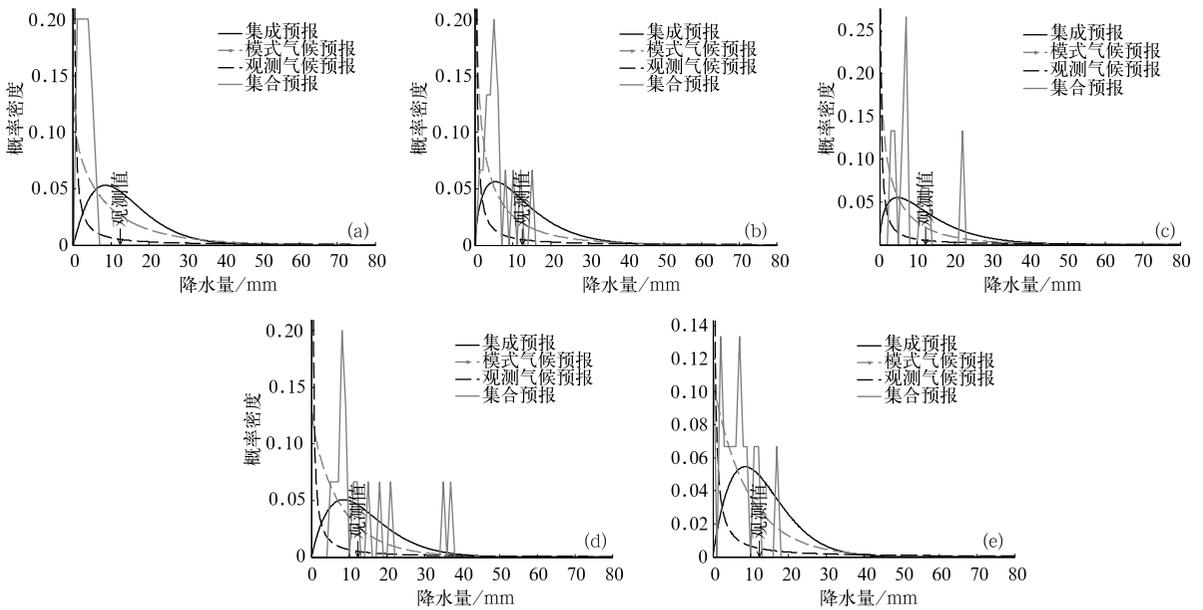


图 11 广州站 2008 年 6 月 12 日贝叶斯降水概率密度预报曲线  
(a~e)图分别为 24、48、72、96 和 120 h 预报时效

Fig. 11 Bayesian precipitation probabilistic forecast in Guangzhou on 12 June 2008  
(a-e) for 24 h, 48 h, 72 h, 96 h and 120 h forecast, respectively

## 4 结论与讨论

为了提高概率预报的修订能力和降水天气的预报效果,采用 1952—2007 年历史观测资料和 2009—2011 年 6 月 24~120 h 中国 T213 全球集合预报历史资料,分别作为贝叶斯降水概率预报模型作为先验信息,对中国不同气候区代表站(广州、南京、武汉和成都)建立贝叶斯降水概率预报模型,对比不同先验信息下集合成员与集成贝叶斯降水概率

预报拟合结果差异,分析先验信息对贝叶斯降水概率预报模型的影响,在此基础上,采用模式先验信息的贝叶斯降水概率预报模型,进行 2008 年 6 月降水概率预报试验。得出如下结论:

(1) 各测站的模式先验分布与观测先验分布特征不同。模式先验概率较观测先验概率偏大,逐日变化也大;各测站的条件降水量先验分布也不同,广州站模式预报高估降水强度,武汉和南京模式预报降水强度与观测相当,而成都的模式预报则低估降水强度。

(2) 模式后验分布函数较观测后验分布函数的方差偏小,均值略大。观测先验信息方案产生的集合成员贝叶斯降水概率预报曲线,其曲率最大处所对应的降水量多位于小雨及无雨区;模式先验信息方案产生的集合成员的贝叶斯降水概率预报曲线,其曲率最大处所对应的降水量更偏向于观测值区,且有效信息评分值也更大。

(3) 贝叶斯降水概率预报检验的结果表明,模式先验信息下的贝叶斯降水概率预报较集合预报结果,有一定程度的提高,其中南京站的提高效果最显著。

(4) 基于模式先验信息的贝叶斯降水概率预报试验表明,对于降水量较大的测站,贝叶斯降水概率预报效果更好,对降水量较小的测站,贝叶斯降水概率预报对无雨或小雨的预报效果更好。

在应用贝叶斯降水概率预报方法时,需要指出的是,本文的研究资料是中国 T213 集合预报模式资料,正如前面的研究所知,尽管中国 T213 集合预报系统的结果具有一定的代表性,可以获得较好的预报结果,但如将贝叶斯降水概率模型应用于其他集合预报系统时,由于模式特征不同,先验信息概率分布会有所改变,从贝叶斯降水概率预报模型中获得的计算参数(如条件概率分布曲线的参数和后验参数等)也会发生相应的变化,由贝叶斯降水概率预报模型预报的概率分布函数也会发生相应的改进,因此将该方法应用于其他集合预报系统时,需要分析参数变化特征及对降水概率分布的影响。

## 参考文献

陈法敬, 矫梅燕, 陈静. 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.

陈静, 陈德辉, 彦宏. 2002. 集合数值预报发展与研究进展. 应用气象学报, 13(4): 497-507.

陈朝平, 马汉中, 陈静. 2010. 基于贝叶斯方法的四川暴雨集合概率预报产品释用. 气象, 35(5): 32-39.

邓国, 陈怀亮, 周玉淑. 2006. 集合预报技术在暴雨灾害风险分析中的应. 自然灾害学报, 15(1): 116.

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.

韩焱红, 矫梅燕, 陈静, 等. 2013. 基于贝叶斯理论的降水集合产品概率化方法及初步试验. 气象, 39(1): 1-10.

矫梅燕. 2010. 天气业务的现代化发展. 气象, 36(7): 1-4.

马培迎. 1999. 应用贝叶斯原理修正降水概率预报. 气象科技, (1): 45-48.

茆诗松. 1999. 贝叶斯统计. 北京: 中国统计出版社.

毛恒青, 陈谊, 陈德辉. 2001. 基于神威中期集合预报系统的产品开发. 应用气象学报, 13(1): 47-55.

穆穆, 陈博宇, 周非凡, 等. 2011. 气象预报的方法与不确定性. 气象, 37(1): 1-13.

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. 北京: 中国统计出版社.

殷志远, 彭涛. 2012. 基于 AREM 模式的贝叶斯洪水概率预报试验. 暴雨灾害, 31(1): 59-65.

张洪刚, 郭生练. 2005. 基于贝叶斯方法的实时洪水校正模型. 武汉大学学报, 38(1): 59-64.

Adrian E R, Tilmann G, Fadoua B. 2005. Using Bayesian model averaging to calibrate forecast ensembles. Mon Wea Rev, 133(5): 1155-1174.

Bishop C, Shanley K. 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. Stochastic Hydrology and Hydraulics, 11(1): 17-31.

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 varieties with specified distributions. J Hydrol, 197(1-4): 286-292.

Krzysztofowicz R, Coire J M. 2006a. Bayesian processor of output: Probability of precipitation occurrence. <http://faculty.virginia.edu/rk/BPO.htm>.

Krzysztofowicz R, Coire J Maranzano. 2006b. Bayesian processor of output: Probabilistic quantitative precipitation forecast. <http://faculty.virginia.edu/rk/BPO.htm>.

Krzysztofowicz R, Evans W. 2008. Probabilistic forecasts from the national digital forecast database. Wea and Forecasting, 23(2): 270-289.

Lorenz E N. 1965. A study of the predictability of a 28-variable atmospheric model. Tellus, 17: 321-333.

Sicillo J K, Ahkqyust J E, Toth Z. 1997. An ensemble forecasting primer. Wea and Forecasting, 12(12): 809-817.