降水 BMA 说明文档

完成人：张维，高建芸

完成单位：福建省气象科学研究所

1.降水的贝叶斯多模式集合平均

1.1 数据

本模型用到了 CMA，EC，UKMO 和 NCEP 的 4 个中心 S2S 资料。为了解决各家模式的起报时间不一致的问题，采用“妥协原则”即：以 EC 的起报时间为基准，假如其他家模式不存在该起报时间，则分别向前和向后找最接近 EC 的起报时间进行集合。假如前、后同时存在近似时间，则以向前的为准。

BMA 训练的时候还需要用到观测资料，本模型中选用全国两千多个站的站点资料作为观测资料。为统一观测和模式分辨率问题，将观测资料插值到模式网格中，即 1.5X1.5 网格，插值算法为自然邻域插值法（natural neighbor interpolation method)。

1.2 降水 BMA 方法简介

令 $f = f_1, f_2, \cdots\cdots, f_k$, 分别表示$k$ 个不同数值模式的预报结果，y 代表需要预报的变量，y^T代表培训数据。根据贝叶斯原理，BMA 预报模型（BMA 预报 PDF）可表示为如下的多模式概率预报加权平均的形式：

$$p[y|(f_1, \ldots, f_K, y^T)] = \sum_{k=1}^{K} w_k p_k[y|(f_k, y^T)] \quad (1)$$

w_k表示在模型训练阶段第 k 个成员预报为最佳预报的后验概率，非负且满足$\sum_{k=1}^{K} w_k = \mathbf{1}$，反应的是每个模型成员在模型训练阶段对预报技巧的相对贡献程度。$p_k[y|(f_k, y^T)]$为先验概率密度，且必须满足一种带参分布。

已有研究（Sloughter et al., 2007,）表明，只有当预报量$y \neq 0$时，降水的立方根才近似符合伽玛分布，因此对原始的降水资料进行立方根转换，$p_k[y|(f_k, y^T)]$则可以表达为如下的分段函数：

$$p_k(y|f_k) = P(y = 0|f_k)I[y = 0] + P(y > 0|f_k)g_k(y|f_k)I[y > 0]) \quad (2)$$

I［·］为指示因子，当大括号中的条件满足时为1，不满足时则为0。$g_k(y|f_k)$ 为伽玛分布。

将（2）式代入（1）可得到降水BMA预报的PDF：

$$p(y|f_1, \ldots, f_k) = \sum_{k=1}^{K} w_k [P(y = 0|f_k)I[y = 0] + P(y > 0|f_k)g_k(y|f_k)I[y > 0]] \quad (3)$$

模型训练的过程中不仅需要计算伽玛分布的参数，还需要建立非零降水概率（依赖于预报量f_k）模型。

- 对于$P(y=0|f_k)$和$P(y>0|f_k)$，有如下回归模型的关系式：

$$logit\,P(y=0|f_k)=log\frac{P(y=0|f_k)}{P(y>0|f_k)}=a_{0,k}+a_{1,k}{f_k}^{1/3}+a_{2,k}\delta_k$$

（4）

当f_k=0时，δ_k=1；反之，δ_k则为0。根据逻辑回归模型关系，可知：

$$P(y=0|f_k)=1-1/(1+\exp(a_{0,k}+a_{1,k}{f_k}^{1/3}+a_{2,k}\delta_k))$$

（5）

- 对于$g_k(y|f_k)$而言，其伽玛分布的表达式如下：

$$g_k(y|f_k)=\frac{1}{{\beta_k}^{\alpha_k}\Gamma(\alpha_k)}y^{\alpha_k-1}\exp(-y/\beta_k) \quad (6)$$

其中，$\Gamma(\cdot)$是伽玛函数，α_k和β_k分别为伽玛分布的形状参数和尺度参数。这两个因子均依赖于伽玛分布的期望μ_k和方差σ_k^2，其相互之间的关系为：$\alpha_k={\mu_k}^2/\sigma_k^2$ ，$\beta_k=\sigma_k^2/\mu_k$。而μ_k和σ_k^2则可以通过f_k进行估计，其公式为：

$$\mu_k=b_{0,k}+b_{1,k}f_k^{1/3} \quad (7)$$

$$\sigma_k^2=c_0+c_1f_k \quad (8)$$

将$P(y=0|f_k)$的表达式（5）和$g_k(y|f_k)$的表达式（6）代入公式（3），就能得到降水BMA的PDF预报模型。

1.3 降水 BMA 训练方法

根据以上公式推导可知，共有$a_{0,k}$，$a_{1,k}$，$a_{3,k}$，$b_{0,k}$，$b_{1,k}$，c_0，c_1，w_k等参数需要求解。通过滑动训练的办法，将某预报时次的前 T 个预报样本和对应的观测样本作为训练样本，通过以下方法训练得到上述参数。

（1）这里求解参数a0,a1,a2是用到梯度上升法求解极大似然函数，使得参数满足训练期内的样本出现概率达到最大。

$$P_{总}=P(y_1|x_1)P(y_2|x_2)P(y_3|x_3)\cdots P(y_n|x_n)=\prod_{n=1}^{N}p^{y_n}(1-p)^{1-y_n} \quad (9)$$

注意$P_{总}$是个函数，且未知的量只有ω（在P里面）

因为逻辑回归的损失函数$F(\omega)$是一个连续的凸函数（conveniently convex）。这样的函数的特征是它只会有一个全局优的点，不存在局部优，因此可以用梯度上升法来求解.

$$F(\omega) = \ln(P_{总}) = \ln\left(\prod_{n=1}^{N} p^{y_n}(1-p)^{1-y_n}\right) = \sum_{n=1}^{N} \ln(p^{y_n}(1-p)^{1-y_n})$$

$$= \sum_{n=1}^{N} (y_n \ln(p) + (1-y_n)^{y_n} \ln(1-p))$$

其中，$p = \frac{1}{1+e^{-\omega^T x}}$

（2）对于每家模式k，$b_{0,k}$，$b_{1,k}$则是通过对T个训练样本进行线性回归训练得到。

（3）对于上述模型中的$w_1 \cdots w_k$以及c_0和c_1则可以通过极大似然法则来进行求解，即：

$$l(w_1 \cdots, w_k, c_0, c_1) = \sum_{i=1}^{N} \log p(y|f_{1,i}, \cdots, f_{k,i}) \tag{11}$$

（4）针对集合平均时，极端降水容易削弱的问题，本模型还进一步采用了分层贝叶斯模型，其做法是根据培训期内某一集合成员的降水预报方差分为若干个区间，如[0, 10]，(10, 25]，(25, 50]，(50, 100]，…等。然后对落在不同区间内的降水预报进行上述的逻辑回归和 BMA 模型训练，再用分层训练出来的权重系数，对集合预报降水进行订正，最后再整合在一起，从而达到既能对降水预报进行订正，同时还能够保留极端降水的信息。

基于训练样本，依据以上步骤训练出 BMA 预报模型的参数，再根据各个模式的预报量则可以对各个量级的降水做出概率预报。图 1 给出了 2021.7.19 起报的 BMA 预报未来 30 天出现大于 25mm 概率

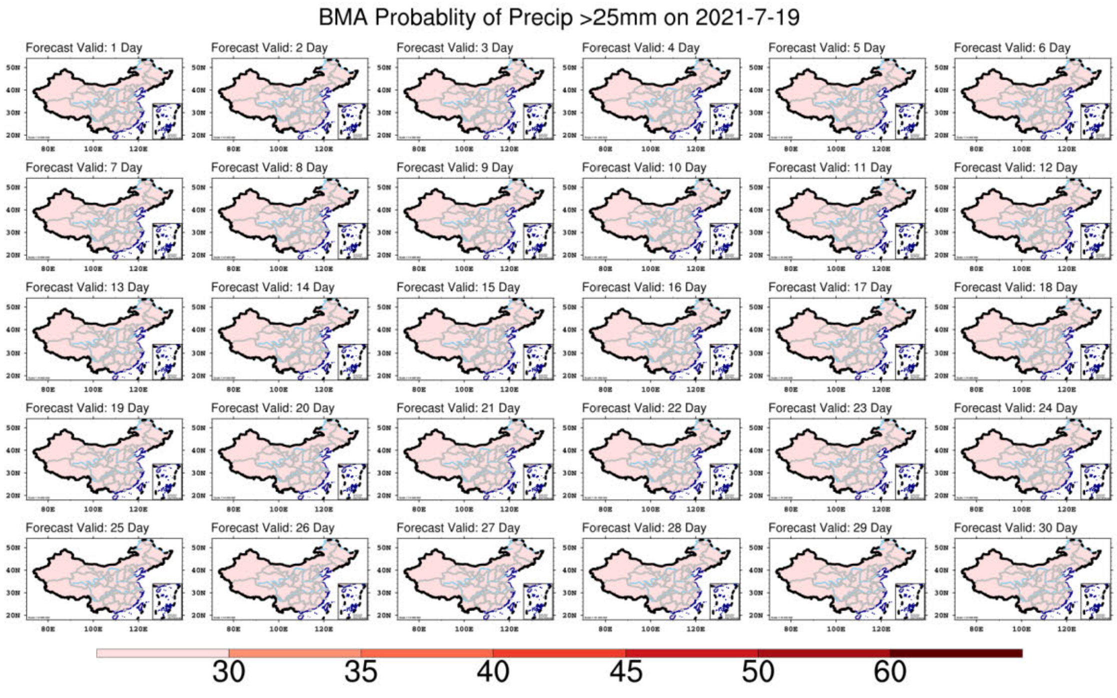

图 1 2021.7.19 起报 BMA 预报未来 30 天出现大于 25mm 的概率

2 基于区域持续性极端强降水指标的贝叶斯多模式集合平均

2.1 区域持续性极端强降水指标的定义

对逐日降水进行 5 天时间滑动，再进行 4 个区域的平均得到 4 个区域的持续性极端强降水指标(RPEPI)，通过判定是否大于一定阈值来确定持续性极端过程。RPEPI 的阈值则是通过 85 百分位的方法进行确定。四个区域的 RPEPI 指标从低到高排序，取上述百分位对应的值 12.0、13.6、12.5 和 8.0 即是四个区域的极端阈值。依据此阈值分别挑选出四个区域持续性极端强降水事件。这些事件 82%以上都包含林等（2020）的持续性事件中，图 2 对比给出了本项目（红色线）和林等（2020）（蓝色线）的长江流域 2000-2010 年持续性极端强降水出现日期，从图中可以看出，二者的重合度非常高。因此在实时监测和预报中都以 RPEPI 指标作为判别区域持续性极端强降水过程的标准。

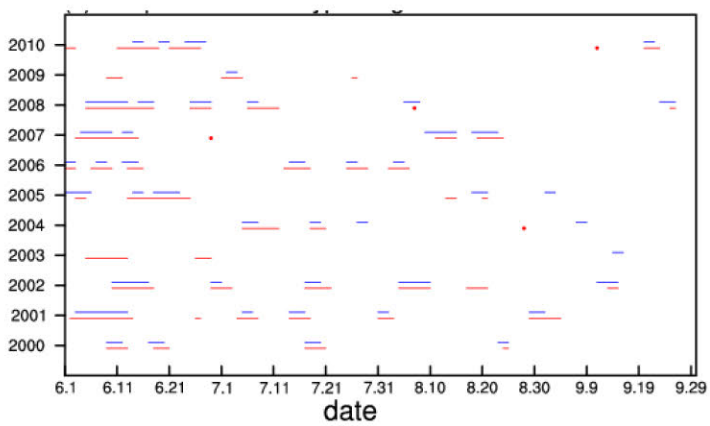

图 2　RPEPI 和林等（2020）（蓝色线）持续性极端强降水事件在华南区域 2000-2010 年出现日期。

2.2 基于 RPEPI 的贝叶斯多模式集合概率预报

由于 RPEPI 的三次方根同样满足伽马分布（图略），因此将（3）式中的降水量替换为 RPEPI，就得到基于 RPEPI 的 BMA 模型。基于该模型进行历史回报试验，得到 BMA 模型 359 个起报时次的 4 个区域未来 30 天预报均值，对该值进行从小到大排序，根据 85 百分位得到贝叶斯模型的 4 个区域未来 30 天极端阈值（表 1）。

在实时预报中根据四家模式预报的日降水量很容易计算出 RPEPI，将该值输入 BMA 模型，得到 BMA 预报的关于 RPEPI 的概率密度函数。基于以上得到的 BMA 阈值，可以对四个区域的持续性极端强降水过程进行实时做出概率预报。图 3 给出了华南 2021 年 6.3 至 6.17 日 5 个起报时次的持续性极端强降水概率预报产品。图中下半部分为实时监测，红色区域为持续性极端过程。上半部分为预报，方块的填色代表 BMA 预报概率的大小。对于每

次起报过程（上半部分纵坐标），将大于 50%概率的时间段作为预报的持续性极端过程，该过程出现的概率为这段时间概率的平均。

表 1 BMA 的区域持续性极端强降水端阈值

预报时效	华南	长江	黄淮	华北
第 3 天	10.85	7.45	6.01	3.56
第 4 天	10.26	7.36	5.47	3.31
第 5 天	10.07	6.81	5.37	3.36
第 6 天	9.79	6.74	5.50	3.12
第 7 天	9.38	6.75	5.07	3.08
第 8 天	8.66	6.45	4.82	3.07
第 9 天	8.89	6.43	4.73	3.22
第 10 天	8.56	6.19	4.97	2.99
第 11 天	8.97	6.09	4.61	2.93
第 12 天	8.86	5.98	4.69	3.02
第 13 天	8.43	5.94	4.90	3.05
第 14 天	8.73	6.03	4.57	2.89
第 15 天	8.68	5.67	4.63	2.78
第 16 天	8.22	5.46	4.81	2.95
第 17 天	7.60	5.62	4.68	2.85
第 18 天	8.13	5.65	4.37	2.79
第 19 天	8.09	5.67	4.61	3.04
第 20 天	7.76	5.63	4.67	2.95
第 21 天	7.70	5.67	4.43	2.97
第 22 天	8.08	5.52	4.50	2.98
第 23 天	7.99	5.36	4.71	3.13
第 24 天	7.78	5.30	4.65	2.92
第 25 天	8.08	5.31	4.60	2.83
第 26 天	7.95	5.36	5.01	2.99
第 27 天	7.74	5.34	4.89	2.90
第 28 天	7.58	5.27	4.66	2.85

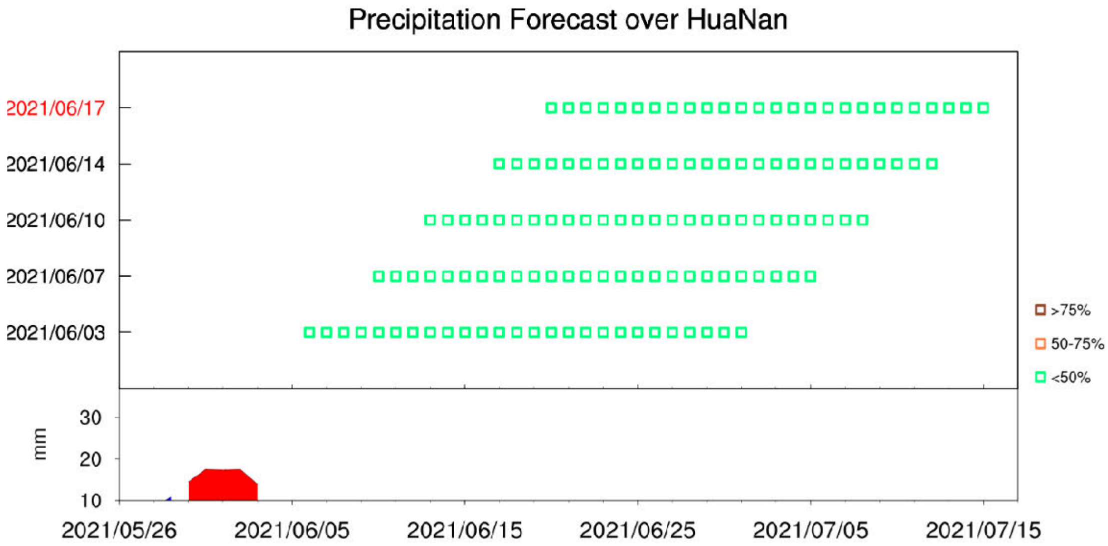

图 3 华南 2021 年 6.3 至 6.17 日 5 个起报时次的持续性极端强降水概率预报（方块填色）。下半部为持续性极端强降水过程的实时监测（基于 RPEH）。