

# 温度 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. 日最高气温的贝叶斯多模式集合平均

### 1.1 数据

该模型用到了 CMA, EC, UKMO, NCEP4 个中心 S2S 资料。由于贝叶斯多模式集合平均(BMA)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，该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 BMA 方法简介

令  $f = f_1, f_2, \dots, f_k$ ，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：

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

$w_k$  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  $\sum_{k=1}^K w_k = 1$ ，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，对于温度近似满足正态分布： $N(\mu, \sigma^2)$ ，其中均值  $\mu = a_k + b_k f_k$ 。

### 1.3 温度 BMA 训练步骤

- 1)、将各家模式进行集合平均得到  $f_k$   $k=1, 2, 3, 4$ ，对应 4 家模式
- 2)、用滑动训练的办法分别对每家模式进行训练，训练样本为前 T 个训练样本观测资料和集合平均预报  $f_k^T$
- 3)、首先进行回归分析训练，得到  $a_k$  和  $b_k$ ，概率分布的均值  $\mu_k = a_k + b_k * f_k$ ， $\mu_k$  同时也是每家模式的确定性预报
- 4)、其次通过极大似然估计得到概率分布的标准差  $\sigma$  和  $w_k$ ，极大似然估计是通过 EM 迭代法实现的。

5)、将以上参数代入(1)得到BMA预报模型 $p[y|(f_1, \dots, f_k, y^T)] = \sum_{k=1}^K w_k p_k[y|(f_k, y^T)]$

6)、训练样本数T的确定。BMA采用滑动训练的办法,对1999年到2000年的6到9月模式最高气温回报产品进行回报实验,分别选取不同的样本T,重复以上BMA的计算步骤,并对不同样本数的BMA概率预报结果进行MAE评分和CRPS评分,在滑动训练样本数达到15个以后,无论是MAE还是CRPS评分都趋于收敛达到最小(图略),因此日最优滑动训练样本数为15个。

7)实时预报时,采用15个训练样本进行滑动训练,根据预报值 $f_k$ 得到各个参数,确定BMA概率模型,最终做出日最高气温的概率预报。图1给出了未来三十天全国范围内出现35度高温的概率。

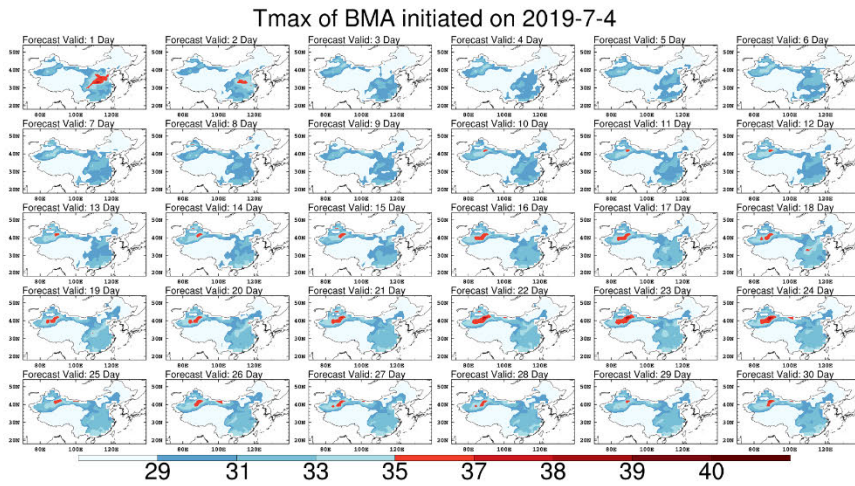


图1 2019.7.4起报BMA预报的35度高温出现概率

## 2 基于区域持续性极端高温指标的贝叶斯多模式集合平均

### 2.1 区域持续性极端高温指标的定义

对逐日最高气温进行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高温指标(RPEHI),通过判定是否大于一定阈值来确定持续性极端过程。RPEHI的阈值则是通过百分位的方法进行确定。对于普通人,当环境温度高于35°C的临界温度时,人体会出现不适。因此,指标的百分位并不是取固定的90百分位或者95百分位,而是取与35°C有关的百分位。具体算法如下:从1981-2010年每个格点的逐日日最高气温中计算得到35°C日最高气温在气候态中所处的百分位(图2),对该值进行四个区域的平均得到四个区域的极端百分位阈值:82%,80%,93%,97%。四个区域的RPEHI指标从低到高排序,取上述百分位对应的值33.77,33.85,33.65,31.86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高温事件。这些事件90%以上都包含林等(2020)的持续性事件中。

图 3 对比给出了本项目（红色线）和林等（2020）（蓝色线）的华南区域 2000-2010 年持续性极端高温出现日期，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 REPHI 指标作为判别区域持续性极端高温过程的标准。

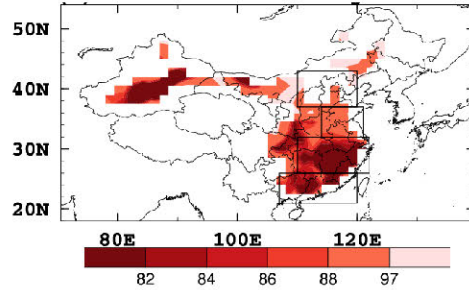


图 2 逐日 35°C 高温在气候态(1999 至 2010 年 6 到 9 月)中所处的百分位。图 a 中的 4 个方框自下而上分别代表华南，长江，黄淮和华北 4 个区域

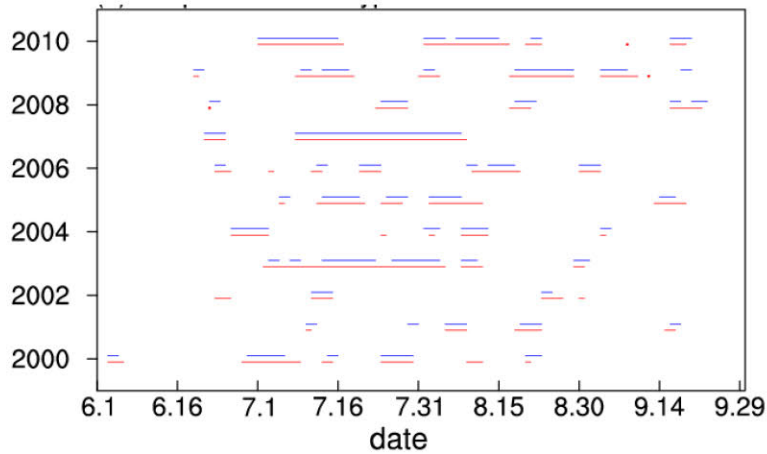


图 3 RPEHI 和林等（2020）（蓝色线）持续性极端高温事件在华南区域 2000-2010 年出现日期。

## 2.2 基于 RPEHI 的贝叶斯多模式集合概率预报

将 1.3 中 BMA 的温度替换为 RPEHI，就得到基于 RPEHI 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 82%，80%，93%，97% 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最高气温很容易计算出 RPEHI，将该值输入 BMA 模型，得到 BMA 预报的关于 RPEHI 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高温过程进行实时做出概率预报。图 4 给出了华南 2010 年 7.1 至 7.16 日出现的一次持续性极端高温过程概率预报产品。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 贝叶斯模型的区域持续性极端高温端阈值

预报时效	华南	长江	黄淮	华北
第 3 天	33.8428	33.89753	32.96951	32.32766
第 4 天	33.74672	33.91259	32.9925	32.62354
第 5 天	33.66722	33.78209	32.86483	32.15408
第 6 天	33.64685	33.6498	32.69702	32.18572
第 7 天	33.62174	33.60661	32.70766	32.2768
第 8 天	33.59122	33.66167	32.77243	31.96672
第 9 天	33.51101	33.68399	32.7721	31.54713
第 10 天	33.52547	33.70086	32.78706	31.63548
第 11 天	33.54495	33.71121	32.88475	31.41626
第 12 天	33.51729	33.68283	32.66032	31.52952
第 13 天	33.47613	33.60579	32.61955	31.60884
第 14 天	33.48001	33.63245	32.69774	31.62118
第 15 天	33.50592	33.54123	32.64324	31.47614
第 16 天	33.51959	33.55412	32.66224	31.9398
第 17 天	33.52579	33.53339	32.7433	31.89494
第 18 天	33.47377	33.54587	32.74575	31.49578
第 19 天	33.53307	33.56335	32.71981	31.57491
第 20 天	33.64635	33.62646	32.62955	31.4764
第 21 天	33.67904	33.62511	32.59129	31.60654
第 22 天	33.65416	33.6572	32.4396	31.36636
第 23 天	33.57472	33.73192	32.36902	31.36436
第 24 天	33.60168	33.83299	32.58228	31.54082
第 25 天	33.61905	33.84483	32.62895	31.53996
第 26 天	33.70481	33.85358	32.40081	31.40599
第 27 天	33.72478	33.92703	32.40697	31.1932
第 28 天	33.8032	33.81464	32.36282	31.57044

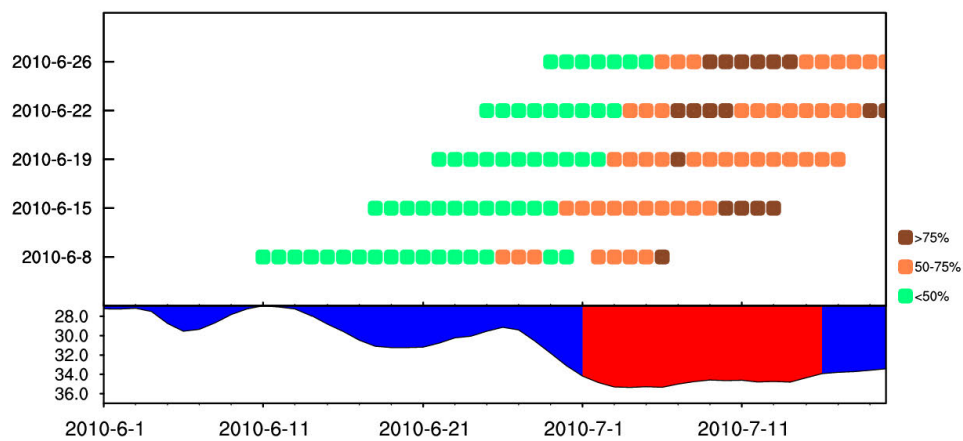


图 4 上半部为 BMA 在不同起报时间下对于 2010 年 7 月 1 日至 7 月 16 日华南区域持续性极端高温过程的概率预报（方块填色）。下半部为持续性极端高温过程的实时监测（基于 RPEH）。