

# 基于雨量分级回归分析的站点日降水量预报订正

王姝苏<sup>1</sup> 周红梅<sup>2</sup> 朱寿鹏<sup>3</sup>

(1 江苏省海门市气象局,南通 226100; 2 江苏省射阳县气象局,盐城 224300;

3 南京信息工程大学气象灾害教育部重点实验室/气象灾害预报预警与评估协同创新中心,南京 210044)

**摘要** 利用 ECMWF 24 h 累计降水量预报资料,以全国范围内 2403 个国家地面气象观测站 24 h 累计降水量作为观测资料,对站点预报结果进行雨量分级回归订正,并与直接双线性插值的预报结果进行对比,利用频率匹配法进一步对不同量级的降水预报结果进行二次订正。结果表明,雨量分级回归相比双线性插值,可以减小预报误差,提高预报结果与观测值之间的相关系数以及降水预报的 ETS 评分,使站点预报值更接近实况。频率匹配法能改善各降水量级的预报效果,降水面积偏差减小,小雨空报率和大雨漏报率减小。对于不同起报时间、不同降水量级和不同预报时效,雨量分级回归和频率匹配法的改进程度各不相同。雨量分级回归对于 20:00 起报的降水预报改进幅度大于 08:00,对 0.1 mm 和 50 mm 量级的降水预报改进较为有限,对 5~15 mm 量级的降水预报改进明显,且随预报时效的延长,对降水预报的改进幅度呈增大趋势。此外,频率匹配法对于起报时次效果较差的降水预报改进幅度较大。

**关键词** 降水;订正;雨量分级回归;频率匹配法

中图分类号: P456 DOI: 10.19517/j.1671-6345.20190159 文献标识码: A

## 引言

我国位于东亚季风区,旱涝灾害时有发生,各地汛期气象工作的重点多是降水预报及相关的灾害性天气预报<sup>[1-4]</sup>,降水预报方法的研究<sup>[5-9]</sup>是气象科研的重点,其中应用最广泛的是数值预报模式产品。但是模式预报存在一定的随机性和系统性误差,利用统计方法对模式预报结果进行后处理和订正,从而减小模式预报的误差<sup>[10-11]</sup>。尤其是降水具有不连续和非正态分布等特点,预报不确定性较大,需要建立合适的订正模型<sup>[12-13]</sup>。Piani 等<sup>[14]</sup>基于经验累计概率分布订正欧洲的日降水量场,但这种方法计算成本较大。台站日常业务中更倾向于计算量小且有效的预报订正方法<sup>[15-16]</sup>。王亚男和智协飞<sup>[17]</sup>的研究表明,将观测资料与预报资料建立线性回归模型,能对日降水量进行有效订正。相比雨量未分级的回归模型,针对不同等级雨量分别建立观测资

料和预报资料之间的回归方程,能更显著地提高各降水量级的预报技巧<sup>[18-20]</sup>。线性回归对降水量有一定的平滑作用,Ebert<sup>[21]</sup>提出“频率匹配”的思路来研究小降水区偏大和大降水区偏小的问题。国内外大量研究表明,频率匹配法能有效减小模式预报雨量和雨区面积的偏差<sup>[22-25]</sup>。本文对 ECMWF 24 h 累计降水量双线性插值到站点的预报结果进行雨量分级回归订正,利用频率匹配法进行二次订正,从而得到更加准确的站点预报结果。

## 1 资料和方法

### 1.1 资料

本文采用的预报资料是 ECMWF 24 h 累计降水量预报,预报时段取 2017 年 5 月 1 日至 10 月 31 日,起报时间为 08:00 和 20:00(北京时),预报时效为 24~96 h,预报间隔 24 h,空间范围 70°~140°E,15°~55°N,空间分辨率为 0.125°×0.125°。

<http://www.qxkj.net.cn> 气象科技

国家重点研发计划重点专项(2017YFC1502002)、江苏省研究生科研与实践创新计划项目(KYCX17\_0875)、中国民用航空华东地区管理局委托项目(2019378)、盐城市气象局科技项目(YQK2017002)资助

作者简介:王姝苏,女,1992 年生,硕士,主要从事气象预报服务工作,Email:wangshusu\_nt@126.com

收稿日期:2019 年 4 月 10 日;定稿日期:2020 年 1 月 9 日

观测资料是全国范围内 2403 个国家地面气象观测站(图 1)24 h 累计降水量,观测时段取 2017 年 5 月 2 日至 11 月 4 日的 08:00 和 20:00 与预报对应的 24 h 累计观测降水。

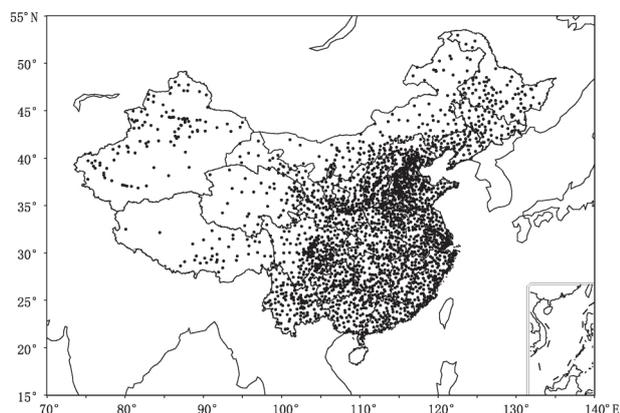


图 1 中国 2403 个国家地面气象观测站点分布

## 1.2 方法

### 1.2.1 雨量分级回归

首先利用双线性插值的方法,将降水量的模式预报值插值到全国范围内 2403 个站点上。

根据日降水预报值进行分级,分为小于 10 mm(小雨)、10~25 mm(中雨)和大于 25 mm(大雨)3 个量级,挑出各量级的降水样本,分别建立各自的回归方程<sup>[26]</sup>。采用一元线性回归,选取一定长度的训练期,建立模式预报值和观测值之间的统计关系式:

$$y_i = ax_i + b \quad (1)$$

式(1)中,  $a$ 、 $b$  为回归系数,  $x_i$  为模式预报结果,  $y_i$  为回归订正结果,对于每个站点,在训练期确定回归系数  $a$ 、 $b$  后,通过该公式,对预报期的模式预报值进行回归订正。

为了充分运用所有的样本资料,增加不同降水量级的样本个数,采用交叉检验的方法建立回归方程,交叉检验即从资料序列的第一个样本开始,每次轮流留出一个样本作为预报检验,余下样本都作为训练期样本进行模拟<sup>[27]</sup>。

### 1.2.2 频率匹配法

某一阈值在预报中出现的频率应与在实况中出现的频率一致(即图 2 中纵坐标应保持不变),例如预报中的 25 mm 降水应被订正到和实况频率一致的 15 mm 降水,35 mm 降水预报应被订正到 45 mm 降水。这种保持预报值和观测值频率一致的方法叫作“频率匹配法”。从空间分布来理解,频率的大小

实际就是空间范围的大小(站点数的多少)。根据 Zhu 和 Luo<sup>[23]</sup> 的研究思路,本文统计预报和观测频率采用的是递减平均法。

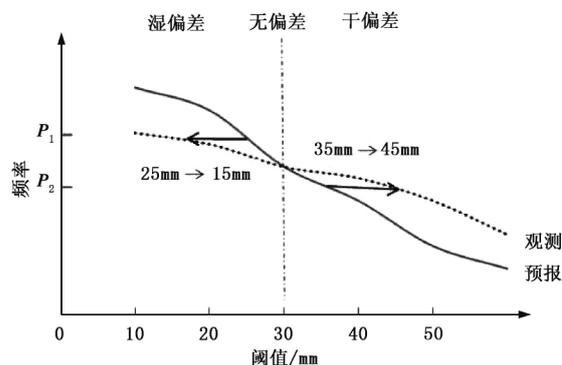


图 2 频率匹配法示意

$$\overline{f_{k,t}} = (1 - \omega) \overline{f_{k,t-1}} + \omega f_{k,t} \quad (2)$$

$$\omega = \frac{1}{n} \quad (3)$$

式(2)~(3)中,  $\overline{f_{k,t}}$  代表第  $t$  天阈值为  $k$  的降水频率平均值,  $\overline{f_{k,t-1}}$  代表第  $t-1$  天阈值为  $k$  的降水频率平均值,  $f_{k,t}$  代表第  $t$  天阈值为  $k$  的降水频率,  $\omega$  是权重系数,  $n$  是递减平均天数,本文  $n$  取 30 d,用迭代 30 d 后的降水频率平均值作为初值启动。

找出与预报频率相等的观测频率对应的降水量,除以预报的降水量即为该降水预报值的订正系数,通过线性插值,得到所有降水预报值的订正系数,最后用模式预报值乘以对应的订正系数即可得频率匹配法订正后的降水预报值。

### 1.2.3 检验方法

采用相关系数、均方根误差、空报率、漏报率、降水偏差和 ETS 评分对降水预报进行检验。

$$r = \frac{\frac{1}{n} \sum_{i=1}^n (f_i - \bar{f})(o_i - \bar{o})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - \bar{f})^2 \frac{1}{n} \sum_{i=1}^n (o_i - \bar{o})^2}} \quad (4)$$

$$\text{RMSE} = \left[ \frac{1}{n} \sum_{i=1}^n (f_i - o_i)^2 \right]^{\frac{1}{2}} \quad (5)$$

式(4)~(5)中,  $r$  是相关系数, RMSE 是均方根误差,  $n$  为站点样本数,  $f_i$  为第  $i$  个样本的预报值,  $o_i$  为第  $i$  个样本的观测值,  $\bar{f}$  为  $n$  个样本预报值的平均值,  $\bar{o}$  为  $n$  个样本观测值的平均值。相关系数越大,则模式预报效果越好。均方根误差越小,表示预

报误差越小。

$$FAR = \frac{c}{a+c} \tag{6}$$

$$MR = \frac{b}{a+b} \tag{7}$$

$$BIAS = \frac{a+c}{a+b} \tag{8}$$

$$ETS = \frac{a-e}{a+b+c-e} \tag{9}$$

$$e = \frac{(a+b)(a+c)}{a+b+c+d} \tag{10}$$

式(6)~(10)中,FAR 是空报率,MR 是漏报率,BIAS 是降水偏差,ETS 是 ETS 评分,  $a$  是预报正确的,  $b$  是漏报的,  $c$  是空报的,  $d$  是实况和预报都没有降水的,  $e$  是随机预报正确的。ETS 评分针对某个量级以上的降水进行评分。ETS ≤ 0 为无技巧预报, ETS > 0 为有技巧预报, ETS = 1 为最佳预报<sup>[28]</sup>。降水偏差是降水范围的偏差,即降水预报面积与实况面积的比,本文为了更直观地比较,面积偏差都减去了 1,降水偏差等于 0,两者面积相等;降水偏差大于 0,预报面积偏大(湿偏差);降水偏差小于 0,预报面积偏小(干偏差)<sup>[29]</sup>。空报率是降水空报面积与预报面积的比。漏报率是降水漏报面积与实况面积的比。空报率和漏报率越小,预报效果越好。

### 2 雨量分级回归预报

将 ECMWF 模式预报格点资料双线性插值到站点上,根据降水量进行分级,分别建立不同降水量级各自的回归方程,用站点观测资料对预报值进行订正,从而使降水预报更接近实况。

将直接双线性插值的预报值和经过雨量分级回

归订正的预报值分别与实况资料进行对比,计算 ECMWF 08:00 和 20:00 起报的 1~4 d 预报时效日降水量的双线性插值和雨量分级回归的相关系数和均方根误差,如图 3。随着预报时效的延长,不同起报时间的预报结果和观测资料的相关系数在减小,均方根误差在增大,08:00 起报的直接双线性插值的相关系数与 20:00 的相比差距在增大,20:00 起报的均方根误差在各个预报时效均小于 08:00 起报的。通过雨量分级回归订正,相关系数相比直接双线性插值在不同预报时效都有明显地提高,基本都提高了 0.2 左右,对于 20:00 起报的提高幅度稍大些。雨量分级回归订正后的均方根误差相比直接插值明显减小,减小幅度在 1.8~2.4 mm/d,对于 20:00 起报的减小幅度更大。雨量分级回归订正可以提高模式预报值与观测值之间的相关性,同时减小模式预报值与观测值之间的误差,对降水预报的改进幅度随预报时效的延长略微增大。

图 4 是 08:00 和 20:00 起报的 24 h 预报时效不同雨量等级以上的 ETS 评分,不同起报时间的日降水量 ETS 评分较接近。0.1 mm 降水量级 ETS 评分较低,经过雨量分级回归订正后,ETS 评分略有提高,但仍然较低,这与严重的小雨空报有关。50 mm 量级经过订正,ETS 评分提高也较少。其他降水量级经雨量分级回归后的 ETS 评分相比双线性插值都有较明显的提高,尤其 5~15 mm 量级的降水预报改进程度相比其他量级更明显,可能是样本较多,建立的回归方程较稳定。随着雨量的增大,小雨以上量级的 ETS 评分在减小,雨量分级回归的 ETS 评分提高程度也在减小。

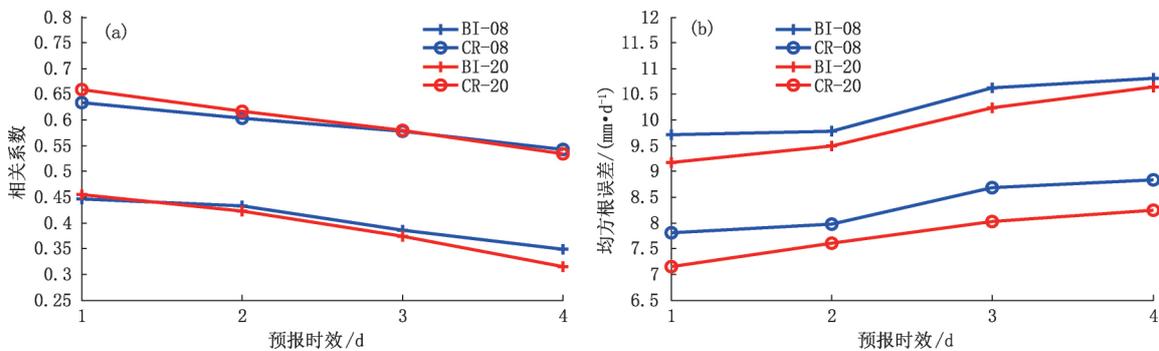


图 3 08:00(蓝线)和 20:00(红线)起报的 1~4 d 预报时效的日降水量双线性插值(十字)、雨量分级回归(圆圈)的相关系数(a)和均方根误差(b)

(图例中 BI 表示双线性插值法,CR 表示分级回归法,下同)

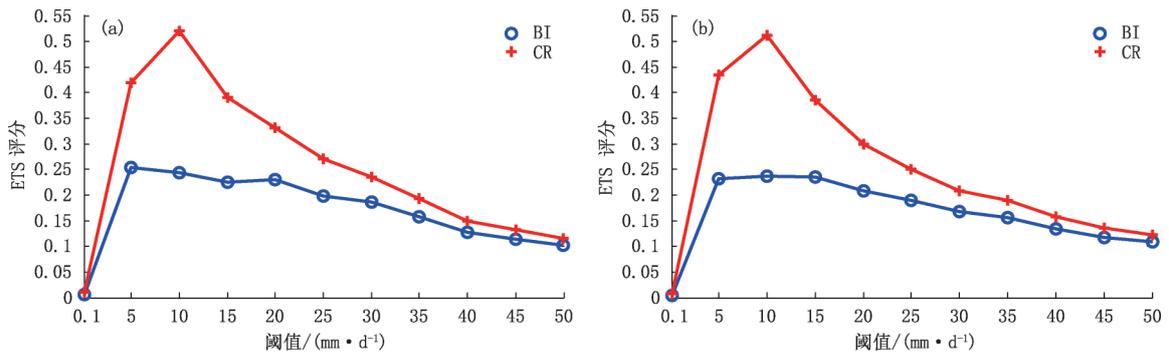


图 4 08:00(a)和 20:00(b)起报的 24 h 预报时效的日降水量双线性插值(蓝线)、雨量分级回归(红线)的不同雨量等级以上 ETS 评分

为了更直观地看出雨量分级回归对预报效果的具体改进,计算 1~4 d 预报时效日降水量 10 mm 和 25 mm 以上的 ETS 评分,如图 5。随着预报时效的延长,08:00 和 20:00 起报日降水量的 ETS 评分都呈下降趋势。20:00 起报的中雨和大雨量级的 ETS 评分都不如 08:00 起报的。相比直接双线性插值,中雨量级的预报经雨量分级回归后的 ETS 评

分在各个预报时效都提高了 1 倍以上,在第 4 d 预报时效的 ETS 评分比双线性插值在第 1 d 预报时效的 ETS 评分还提高了 0.5 倍。由于大雨样本数较少,大雨量级的预报 ETS 评分相对较低,经雨量分级回归后的 ETS 评分在各个预报时效相比双线性插值都提高了 0.05 以上。

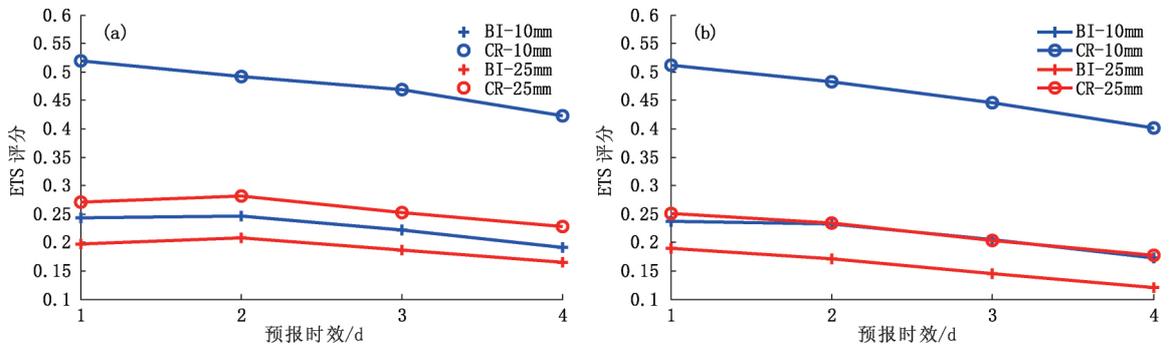


图 5 08:00(a)和 20:00(b)起报的 1~4 d 预报时效日降水量双线性插值(十字)、雨量分级回归(圆圈)的 10 mm(蓝线)、25 mm(红线)以上的 ETS 评分

随机挑取一个站点对比双线性插值、雨量分级回归的预报与实况降水量,图 6 是盐城站 08:00 起报的去除了缺测资料的 24 h 预报时效的降水量和实况,直观地显示出雨量分级回归对降水预报结果的改进。预报结果相比降水量实况存在一定的小雨空报和大雨漏报的问题。对于实况降水量为 0 的情况,雨量分级回归减小了直接双线性插值的预报值,使预报雨量与实况相差较小;对于降水量较大的实况,雨量分级回归相比双线性插值更接近观测值。

以上研究表明,不同起报时间的降水预报在不同的检验方法下效果不尽相同。相比直接双线性插值,雨量分级回归可以提高模式预报值与观测值之

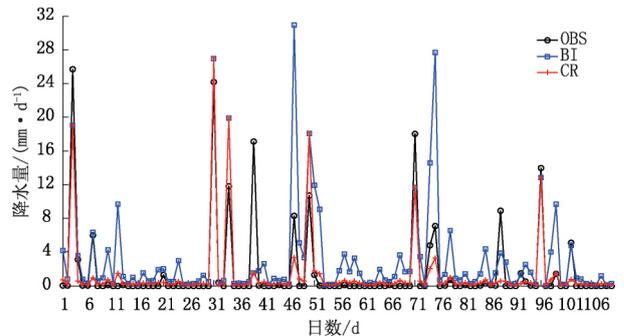


图 6 08:00 起报的 24 h 预报时效日降水量实况(黑线)、双线性插值(蓝线)与雨量分级回归(红线)的预报

间的相关系数,减小模式的预报误差,提高不同降水量级的 ETS 评分。但由于线性回归的平滑作用,还存在较严重的小雨空报和大雨漏报问题。

### 3 频率匹配法对降水量级的订正

利用频率匹配法对雨量分级回归后的降水预报进行处理,为了更直观地看出频率匹配法对小雨空报和大雨漏报问题的改善,计算了 1~4 d 预报时效频率匹配法处理前后日降水量 0.1 mm 的空报率和

25 mm 的漏报率,如图 7。20:00 起报的小雨空报率和大雨漏报率在各预报时效均大于 08:00 起报的。经频率匹配法的处理,各预报时效的小雨空报率都在减小,其中 20:00 起报的 24 h 预报时效的小雨空报率减小了 5%左右。随预报时效的延长,大雨漏报率在增大,频率匹配使各预报时效的大雨漏报率都明显地减小,其中 20:00 起报的大雨漏报率在各预报时效减小幅度都更大。

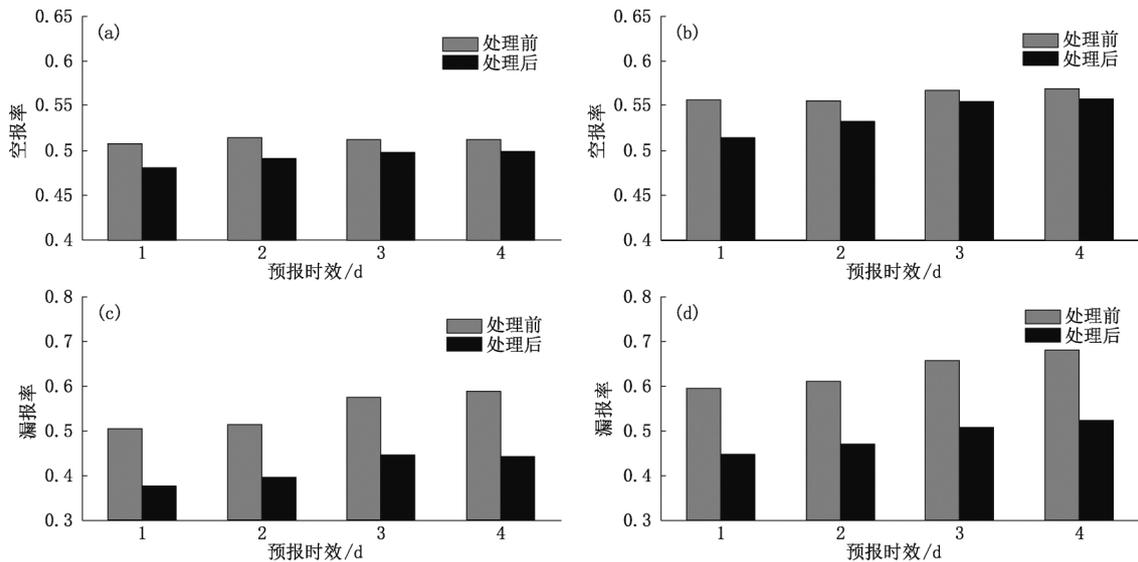


图 7 08:00(a,c)、20:00(b,d)起报的 1~4 d 预报时效频率匹配法处理前(灰柱)、后(黑柱)日降水量 0.1 mm 的空报率(a,b) 和 25 mm 的漏报率(c,d)

图 8 计算了 24 h 预报时效频率匹配法处理前后大雨及以上量级降水的漏报率。研究时段内大雨样本数量不多,但大雨漏报的问题比较严重,尤其

20:00 起报的大雨漏报率更高,经频率匹配后的大雨及以上量级的漏报率都在明显减小,20:00 起报的减小幅度更大。

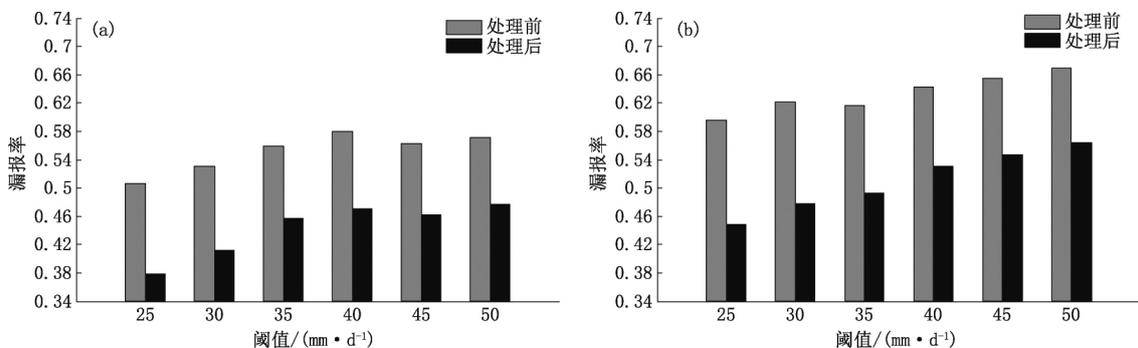


图 8 08:00(a)和 20:00(b)起报的 24 h 预报时效频率匹配法处理前(灰柱)、后(黑柱)大雨及以上量级降水的漏报率

图 9 是频率匹配法处理前后 24 h 预报时效 0.1、10、25、50、75 mm 这 5 个降水量级的面积偏

差。从图中可以看出,0.1 mm 降水量级的湿偏差经频率匹配后有所减小,10 mm 和 25 mm 降水量级

的干偏差经频率匹配后也都有明显的减小,50 mm 和 75 mm 降水量级的干偏差经频率匹配后变成了

湿偏差,但总的来说,偏差的面积也有明显地减小。

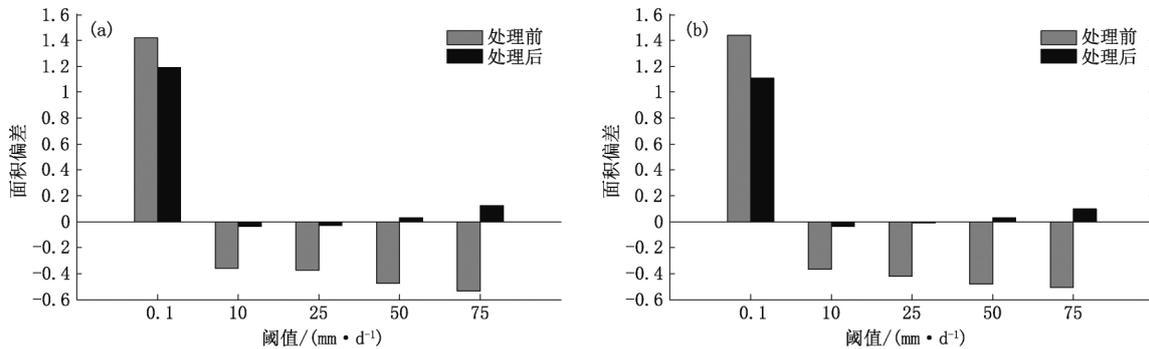


图9 08:00(a)和20:00(b)起报的24 h预报时效频率匹配法处理前(灰柱)、后(黑柱)不同降水量级的面积偏差

综合以上分析,08:00起报的小雨空报率和大雨漏报率都小于20:00的,利用频率匹配法对预报结果进行二次订正,可以进一步改善各个降水量级的预报效果,降低小雨空报率和大雨漏报率,减小降水面积偏差,得到更接近实况的站点预报结果,对于20:00起报的降水预报改进幅度较大。

#### 4 结论

本文利用ECMWF 24 h累计降水量预报资料,以全国范围内2403个国家地面气象观测站24 h累计降水量作为观测资料,进行雨量分级回归订正,并利用频率匹配法进一步处理降水预报结果,得到以下几点结论:

(1)将雨量分级回归引入观测资料订正,相比直接使用双线性插值的预报结果,能减小预报误差,提高预报值与观测值之间的相关性,同时提高各降水量级的ETS评分。对于不同起报时间、不同降水量级和不同预报时效,雨量分级回归订正的改进程度各不相同,20:00起报的降水预报改进幅度大于08:00,对0.1 mm和50 mm量级的降水预报改进幅度较小,对5~15 mm量级的降水预报改进明显,且随预报时效的延长,对降水预报的改进幅度略微增大。

(2)通过频率匹配法将预报资料和实况资料的频率相匹配,降低了小雨空报率和大雨漏报率,减小了降水面积偏差。频率匹配法对于起报时效效果较差的降水预报改进幅度较大。

(3)本文使用自动站实况作为观测资料,通过雨量分级回归和频率匹配法得到针对站点的降水预报

结果,计算量小,可操作性强,订正效果显著,各台站可用将来模式预报的格点降水量转化为所需站点(例如区域站)的降水量,以期得到更精细化的预报结果。

#### 参考文献

- [1] 王雨. 2004年主汛期各数值预报模式定量降水预报评估[J]. 应用气象学报, 2006, 17(3):316-324.
- [2] 张冰, 魏建芬, 王文兰, 等. 转折性天气降水预报检验方法及应用[J]. 气象科技, 2012, 40(3):411-416.
- [3] 吴进, 付宗钰, 吴剑坤, 等. 北京地区夏季多模式降水预报的降尺度集成应用[J]. 气象科技, 2016, 44(1):55-60.
- [4] 周北平, 杜爱军, 苟尚, 等. 短时强降水监测和预警技术及其在山区的应用[J]. 气象科技, 2018, 46(3):490-496.
- [5] 曹萍萍, 肖递祥, 徐栋夫, 等. 基于概率匹配的西南区域模式定量降水订正试验[J]. 气象科技, 2018, 46(1):102-111.
- [6] 宇如聪. 一个 $\eta$ 坐标有限区域数值预报模式对1993年中国汛期降水的实时预报试验[J]. 大气科学, 1994, 18(3):284-292.
- [7] 王晨稀, 端义宏. 短期集合预报技术在梅雨降水预报中的试验研究[J]. 应用气象学报, 2003, 14(1):69-78.
- [8] 彭九慧, 丁力, 杨庆红. 几种降水集成预报方法的对比分析[J]. 气象科技, 2008, 36(5):520-523.
- [9] 杜钧, 李俊. 集合预报方法在暴雨研究和预报中的应用[J]. 气象科技进展, 2014, 4(5):6-20.
- [10] 毕宝贵, 代刊, 王毅, 等. 定量降水预报技术进展[J]. 应用气象学报, 2016, 27(5):534-549.
- [11] 卞正奎, 朱寿鹏, 胡航菲, 等. 泰州市气温多模式集成预报系统的建立与评估[J]. 气象科技, 2016, 44(4):605-611.
- [12] 王海霞. 基于多模式的降水量统计降尺度预报研究[D]. 南京:南京信息工程大学, 2014.
- [13] 王海霞, 智协飞. 基于TIGGE多模式降水量的统计降尺度研究[J]. 气象科学, 2015, 35(4):430-437.
- [14] Piani C, Weeden G P, Best M, et al. Statistical bias correction of global simulated daily precipitation and temperature for

- the application of hydrological models [J]. *J Hydrol*, 2010, 395(3-4): 199-215.
- [15] 陈力强, 韩秀君, 张立祥. 基于 MM5 模式的站点降水预报释用方法研究[J]. *气象科技*, 2003, 31(5):268-272.
- [16] 王建鹏, 薛春芳, 潘留杰, 等. 陕西省精细化网格预报业务系统技术方法[J]. *气象科技*, 2018, 46(5):910-918.
- [17] 王亚男, 智协飞. 多模式降水集合预报的统计降尺度研究[J]. *暴雨灾害*, 2012, 31(1):1-7.
- [18] 智协飞, 王姝苏, 周红梅, 等. 我国地面降水的分级回归统计降尺度预报研究[J]. *大气科学学报*, 2016, 39(3):329-338.
- [19] 陈艳蝶, 智协飞, 王姝苏, 等. 一个台风降水预报的分级回归统计降尺度应用研究[J]. *科技通报*, 2018, 34(5):52-58.
- [20] 王姝苏, 智协飞, 俞剑蔚, 等. 降水的统计降尺度预报及其空间相关性和时间连续性重建[J]. *大气科学学报*, 2018, 41(1):36-45.
- [21] Ebert E E. Ability of a poor man's ensemble to predict the probability and distribution of precipitation [J]. *Monthly Weather Review*, 2001, 129(10):2461-2479.
- [22] 李莉, 朱跃进. T213 降水预报订正系统的建立与研究[J]. *应用气象学报*, 2006, 17(增刊 1):130-134.
- [23] Zhu Y, Luo Y. Precipitation calibration based on the frequency-matching method [J]. *Wea Forecasting*, 2013, 30(5): 1109-1124.
- [24] 李俊, 杜钧, 陈超君. 降水偏差订正的频率(或面积)匹配方法介绍和分析[J]. *气象*, 2014, 40(5):580-588.
- [25] 李俊, 杜钧, 陈超君. “频率匹配法”在集合降水预报中的应用研究[J]. *气象*, 2015, 41(6):674-684.
- [26] 王姝苏. 基于雨量分级回归的我国地面降水的统计降尺度预报研究[D]. 南京:南京信息工程大学, 2016.
- [27] 王亚男. 多模式降水集合预报资料的统计降尺度及误差订正研究[D]. 南京:南京信息工程大学, 2012.
- [28] 王雨, 闫之辉. 降水检验方案变化对降水检验评估效果的影响分析[J]. *气象*, 2007, 33(12):53-61.
- [29] 赵声蓉, 裴海瑛. 客观定量预报中降水的预处理[J]. *应用气象学报*, 2007, 18(1):21-28.

## Station Forecast Calibration of Daily Precipitation Using Categorized Rainfall Regression

WANG Shusu<sup>1</sup> ZHOU Hongmei<sup>2</sup> ZHU Shoupeng<sup>3</sup>

(1 Haimen Meteorological Service, Jiangsu, Nantong 226100; 2 Sheyang Meteorological Service, Jiangsu, Yancheng 224300; 3 Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Key Laboratory of Meteorological Disaster, Ministry of Education, Nanjing University of Information Science & Technology, Nanjing 210044)

**Abstract:** Based on the precipitation forecast dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF) and the daily accumulated precipitation of 2403 national surface meteorological observation stations across China, the calibration on daily precipitation by means of the categorized rainfall regression and the further calibration by means of the frequency-matching method are conducted. The results show that compared with the bilinear method, the categorized rainfall regression is more effective in decreasing the forecast biases, and improves the correlation coefficient with the observed data and the equitable threat score. The forecasts of different thresholds become more accurate after applying the frequency-matching method, with the smaller area deviation. The false alarm rate of light rain and the missing rate of heavy rain are also both reduced. Improvements of the forecasts by the categorized rainfall regression and the frequency-matching method differ in initialized times, rainfall thresholds and lead times. After the calibration of the categorized rainfall regression, the forecast initialized at 20:00 is improved with a larger magnitude than that at 08:00. The improvement of the forecast is relatively limited for rainfall thresholds of 0.1 mm and 50 mm, but significant for thresholds of 5 mm, 10 mm and 15 mm. Additionally, the amplitude of the improvement increases slightly over the lead time. The improvement induced by the frequency-matching method is greater for precipitation forecasts initialized at specific times that show worse performances.

**Keywords:** precipitation; calibration; categorized rainfall regression; frequency-matching method