

四种定量降水预报客观订正方法对比研究*

苏翔^{1,2} 袁慧玲³ 朱跃建⁴
SU Xiang^{1,2} YUAN Huiling³ ZHU Yuejian⁴

1. 江苏省气象台, 南京, 210008

2. 中国气象局交通气象重点开放实验室, 南京, 210009

3. 南京大学大气科学学院/中尺度灾害性天气教育部重点实验室, 南京, 210023

4. 美国国家环境预报中心/环境模拟中心, 马里兰, 20740

1. *Jiangsu Meteorological Observatory, Nanjing 210008, China*

2. *Key Laboratory of Transportation Meteorology, China Meteorological Administration, Nanjing 210009, China*

3. *Key Laboratory of Mesoscale Severe Weather Ministry of Education/School of Atmospheric Sciences, Nanjing University, Nanjing 210023, China*

4. *NCEP/Environmental Modeling Center, Maryland 20740, USA*

2020-06-22 收稿, 2020-08-25 改回.

苏翔, 袁慧玲, 朱跃建. 2021. 四种定量降水预报客观订正方法对比研究. 气象学报, 79(1): 132-149

Su Xiang, Yuan Huiling, Zhu Yuejian. 2021. A comparative study of four objective quantitative precipitation forecast calibration methods. *Acta Meteorologica Sinica*, 79(1):132-149

Abstract Four objective quantitative precipitation forecast (QPF) calibration methods, including frequency matching method (FMM), optimal threat score (OTS), optimal percentile (OP) and probability matching (PM), are comprehensively verified based on annual and seasonal ECMWF QPFs at different forecast lead times. An ideal model is proposed to study the performance of FMM and OTS under different scenarios of spatial displacement and dry/wet biases. A heavy rain case is used to demonstrate basic characteristics of the four different calibration methods. Results show that FMM and OTS can only adjust the magnitude of deterministic QPF, while OP and PM can change the pattern of QPF to some extent by using ensemble forecast information. Aiming at optimizing the frequency bias, FMM can well eliminate the dry/wet bias of QPFs, but it can only improve the threat score (TS) of original QPFs when the displacement error is small and the dry/wet bias is large. OTS has limited skills in improving the TS of moderate rain with weak wet bias. By contrast, OP can improve the TS of all precipitation thresholds, benefiting from using ensemble forecast information, especially for longer forecast lead times. However, OP shows large wet biases during spring and summer seasons, while PM suffers from large dry biases for torrential rain events due to the lack of historical observation information. The evaluation of economic value model shows that OP has relatively higher reference value for torrential rain in risk decision making.

Key words Frequency matching method, Optimal threat score, Optimal percentile, Probability matching, Quantitative precipitation forecast, Bias correction

摘要 基于2019年全年、不同季节、不同预报时效的欧洲中期天气预报中心模式的定量降水预报, 检验评估了频率匹配、最优TS评分、最优百分位、概率匹配4种定量降水预报客观订正法的综合性能。利用理想模型研究了不同雨带位移偏差和干湿偏差情形下频率匹配法与最优TS评分的表现, 并通过个例订正展示了4种定量降水预报订正法的基本特征。结果表明: 频率匹配与最优TS评分仅能对确定性预报的降水量级进行调整, 最优百分位和概率匹配法通过引入集合预报信息可在一定程度上改

* 资助课题: 国家自然科学基金项目(41805078、41675109)、中国气象局预报员专项(CMAYBY2020-046)、国防预研项目(305090417)。

作者简介: 苏翔, 主要从事集合预报、定量降水预报研究。E-mail: suxiang01@163.com

通信作者: 袁慧玲, 主要从事中尺度数值模拟和水文气象研究。E-mail: yuanhl@nju.edu.cn

变预报的降水落区。频率匹配法以频率偏差最优为目标,可以很好地消除模式的干湿偏差,但仅在位移偏差较小且存在较大干湿偏差时提升原始预报的TS评分。最优TS评分法难以改进存在弱湿偏差的中雨预报的TS评分,而最优百分位法利用集合预报信息可以显著提升所有降水等级的TS评分,在较长预报时效下优势尤其明显,但也存在春、夏两季湿偏差较大的问题。概率匹配法由于没有使用历史实况信息,在暴雨订正中干偏差较大。经济价值模型检验评估表明,最优百分位法在暴雨量级的风险决策中具有较高的参考价值。

关键词 频率匹配法, 最优TS评分, 最优百分位, 概率匹配, 定量降水预报, 偏差订正

中图法分类号 P456.8

1 引言

定量降水预报是最重要的气象预报业务之一(宗志平等, 2012; 毕宝贵等, 2016)。提升定量降水预报质量,对于防灾减灾、工农业生产和公众生活等方面具有重要意义。数值模式是定量降水预报的基础,然而由于初值和模式本身的不确定性,模式定量降水预报存在一定的误差。

对数值模式直接输出的定量降水预报产品进行客观订正是提升定量降水预报质量的有效手段。美国国家环境预报中心(NCEP)在业务上使用频率匹配法(FMM)对全球模式降水预报产品进行客观订正,显著减小了模式定量降水预报的系统偏差(Zhu, et al, 2015),该方法在中国进行了一定的推广和应用(李俊等, 2014, 2015; 周迪等, 2015; 徐姝等, 2018a)。袁慧玲等于2013年从NCEP引进频率匹配法开发了欧洲中期天气预报中心(ECMWF)确定性降水预报频率订正产品(栾菲, 2017),并于2015年在国家气象中心实现业务运行(代刊等, 2016)。江苏省气象台苏翔基于频率匹配法开发了用于订正中国GRAPES全球预报系统(GFS)确定性降水预报的业务产品,并于2018年在中国气象局数值预报中心实现业务运行。吴启树等(2017)基于频率匹配法的思路,将频率偏差(Frequency Bias, 简称FB)最优改为TS评分最优,发展了最优TS评分(OTS)算法,显著改进了ECMWF、日本气象厅(JMA)、NCEP和T639模式定量降水预报的TS评分。此外,由于集合预报产品相比于确定性预报产品可以提供更加丰富的预报不确定性信息,在降水预报中也发挥了越来越重要的作用(Zhu, et al, 2002; Buizza, 2008; 杜钧等, 2014)。国家气象中心基于集合预报开发了集合多统计量融合技术(FUSE),对不同降水量级分别参考特定的集合分位数(例如集合最大值、90%集合分位数、75%集合分位数、中位数、10%集合分位数等)进行订正(陈博宇等,

2015, 2016; 徐姝等, 2018b)。若将集合多统计量融合技术方法中固定的集合分位数改为根据TS评分最优滑动建模得到的集合百分位,则转变为更加灵活的最优百分位(OP)技术(代刊等, 2018)。最优百分位技术于2015年在国家气象中心投入业务应用,2015年夏季24h预报时效的暴雨主客观预报评分结果显示,最优百分位法的TS评分已略超过预报员的预报结果(代刊等, 2016)。概率匹配(PM)结合了集合成员的量级和集合平均的空间分布进行模式定量降水预报订正,无需历史训练数据,常用于中小尺度降水集合预报(Chien, et al, 2004; Clark, et al, 2009; Gallus, 2010; Berenguer, et al, 2012; Surcel, et al, 2014; 陈博宇等, 2015)。其他定量降水预报客观订正方法还包括分位值映射法(Voisin, et al, 2010)、相似法(Hamill, et al, 2006)、多模式集成(李芳, 2012; 智协飞等, 2013)等,具体可参考代刊等(2018)的综述文章,这里不再赘述。

文中重点研究上述的频率匹配、最优TS评分、最优百分位、概率匹配这4种常见的定量降水预报客观订正方法,主要目的有以下3点:(1)厘清上述4种订正方法的基本概念。目前中国部分科研和业务工作者在使用上述定量降水预报客观订正方法时,由于概念上缺乏清晰的认识以及翻译文献的不准确,出现了将名称相近的方法混淆的情况。例如标注“频率匹配法”的名称但实际使用的是“概率匹配”订正,或是不清楚“集合多统计量融合技术”法和“最优百分位”法的区别;(2)通过理想模型和个例订正加深对不同订正方法的直观理解。TS评分等传统检验指标包含的信息量有限,预报员难以根据评分大小对不同客观订正方法产生直观的理解。通过理想模型研究原始预报雨区在不同位移和干湿偏差情况下频率匹配和最优TS评分订正方法的表现,并对4种不同方法进行个例订正展示,有助于预报员直观地理解上述方法的基本特

性,并将其合理应用于主观分析中;(3)检验评估上述4种定量降水预报订正方法的总体性能。对比不同定量降水预报客观订正方法在全年、不同季节、不同预报时效和不同降水量级下的表现,研究集合预报信息的引入是否有助于预报性能的提升,检验不同预报订正产品在风险决策中的经济价值(Zhu, et al, 2002),为气象业务工作提供科学参考。

2 资料和方法

2.1 资料

气象实况资料为2017年10月1日—2019年12月31日中国2411个气象站的逐日降水观测(08—08时,北京时)。数值预报资料为ECMWF确定性预报资料和集合预报资料,分辨率分别为 0.125° 和 0.5° ,其中集合预报包含50个扰动成员。选取起报时间12时(世界时),预报时效36—228 h(1.5—9.5 d),间隔24 h,并采用双线性插值将预报插值到2411个气象站点上。

2.2 检验方法

选取2019年全年为检验期,检验4个标准降水等级(小雨(0.1 mm/d)、中雨(10 mm/d)、大雨(25 mm/d)、暴雨(50 mm/d))的TS评分、频率偏差(FB)、漏报率(POFD)、虚警率(FAR),计算方法如下(Jolliffe, et al, 2003)

$$TS = \frac{h}{h+m+f} \quad (1)$$

$$FB = \frac{h+f}{h+m} \quad (2)$$

$$POFD = \frac{m}{h+m} \quad (3)$$

$$FAR = \frac{f}{h+f} \quad (4)$$

式中, h 为预报和实况同时达到降水等级的样本比例, m 为实况达到降水等级但预报未达到的样本比例, f 为预报达到降水等级但实况未达到的样本比例。

由式(1)—(4)可推知TS评分、频率偏差、漏报率和虚警率之间存在如下关系

$$FB = \frac{1-POFD}{1-FAR} \quad (5)$$

$$TS = \frac{1}{\frac{1}{1-POFD} + \frac{1}{1-FAR} - 1} \quad (6)$$

在研究不同季节的预报性能时进一步引入性能图(Performance Diagram)(Roebber, 2009; Guan, et al, 2017)将不同定量降水预报客观订正方法的TS评分、频率偏差、虚警率和漏报率在一张图上综合展示,并标注样本基础观测覆盖率(Base Rate,简称BR)。

表1 花费和损失的预报与观测二元事件联立表
Table 1 Contingency table of costs and losses depending on dichotomous forecasts and observed events

	实况发生	实况未发生
预报发生 (采取保护措施)	样本比例 h (花费 $C+L_u$)	样本比例 f (花费 C)
预报不发生 (不采取保护措施)	样本比例 m (花费 L_p+L_u)	样本比例 r (花费 0)

经济价值模型(Zhu, et al, 2002)常用来分析不同定量降水预报客观订正方法在风险决策中的表现情况。如表1所示,假定降水事件发生时:若采取保护措施,花费为 C ,但灾害仍会产生不可避免的经济损失为 L_u ;若未采取保护措施,造成的额外经济损失为 L_p ,总损失为 L_u+L_p 。记降水事件的观测频率为 s , $s=h+m$,则完美预报的平均经济损失为

$$E_{\text{perfect}} = (C+L_u) \cdot s \quad (7)$$

若按预报降水事件是否发生来决策是否采取保护措施,产生的平均经济损失为

$$E_{\text{fst}} = (C+L_u) \cdot h + (L_p+L_u) \cdot m + C \cdot f \quad (8)$$

若不参考预报信息,总是采取或总不采取保护措施的平均经济损失分别为

$$E_{\text{always}} = C + L_u \cdot s \quad (9)$$

$$E_{\text{never}} = (L_p+L_u) \cdot s \quad (10)$$

记花费/损失为 $\alpha=C/L_p$,选取式(9)和(10)中较小的平均经济损失作为参考值

$$E_{\text{ref}} = L_u \cdot s + L_p \cdot \min\{\alpha + s\} \quad (11)$$

通过式(7)、(8)、(11)可计算通过预报进行决策在减少经济损失方面产生的技巧,即经济价值评分

$$V = \frac{E_{\text{fst}} - E_{\text{ref}}}{E_{\text{perfect}} - E_{\text{ref}}} \quad (12)$$

对于特定的预报,经济价值(V)是 α 的函数。 V 越大代表预报对决策的贡献越大, $V>0$ 代表预报信息对决策减少经济损失有正效果, $V<0$ 代表预报

信息对决策减少经济损失有负效果, $V=0$ 代表无效果。

2.3 训练取样方法

针对不同预报时效分别订正, 采用长度为 61 d 的组合对称滑动窗口进行训练样本取样, 即距离预报日最近可用的 30 d 和上一年同期及其之后的 30 d 组合而成的滑动对称训练样本。以订正 2019 年 8 月 22 日 20 时(北京时)起报的 84 h 预报为例, 预报的降水时段为 2019 年 8 月 25 日 08 时—8 月 26 日 08 时, 选取 2019 年 7 月 24 日—8 月 22 日(30 d)和 2018 年 8 月 26 日—9 月 25 日(31 d)共 61 d 的历史实况和对应的 84 h 预报样本进行训练。

此外, 还测试了持续滑动窗口取样方法和历史对称窗口取样方法。持续滑动窗口选取距离预报日最近可用的 61 d 的滑动训练样本, 优点是能够适应模式由于升级更新产生的系统偏差的变化, 缺点在于季节更替时前期样本的季节代表性较差。历史对称窗口选取上一年同期及前后各 30 d 组成的 61 d 的滑动对称训练样本, 优点是季节代表性较好, 缺点是无法及时适应模式由于升级更新产生的系统偏差的变化。

从理论上讲, 组合对称滑动窗口取样方法结合了持续滑动窗口和历史对称窗口的优点。测试结果表明, 3 种取样方法和不同窗口长度对检验结果的影响不大。因此, 以下选取窗口长度为 61 d 组合对称滑动窗口取样方案应用于频率匹配、最优 TS 评分和最优百分位订正方法中(概率匹配订正不需要训练样本)。

2.4 频率匹配法

频率匹配法使用训练期内的确定性预报和对应的实况资料建立原始预报和预报订正值之间的一一映射关系, 目的是消除不同降水等级上的频率偏差, 使得预报与观测雨区的总面积保持一致。使用组合对称滑动窗口取样代替 Zhu 等(2015)中的卡尔曼滤波算法, 并将对订正系数插值改为对预报订正值直接插值, 具体算法是: 首先, 统计一组降水阈值 t_1, t_2, \dots, t_N 对应的预报累积频率 p_1, p_2, \dots, p_N 和观测累积频率 q_1, q_2, \dots, q_N , 通过坐标 $(p_1, t_1), (p_2, t_2), \dots, (p_N, t_N)$ 构建预报累积频率曲线; 其次, 通过插值计算累积频率 q_1, q_2, \dots, q_N 在预报累积频率曲线上对应的降水阈值 g_1, g_2, \dots, g_N ; 然后, 通过坐标 $(g_1, t_1), (g_2, t_2), \dots, (g_N, t_N)$ 构建原始预报值和预报订正值对应的订正曲

线; 最后, 对于任意的原始预报值, 可通过线性插值计算其在订正曲线上对应的预报订正值。选取 $N=10$ 个降水阈值 t_1, t_2, \dots, t_N 分别为 0.1、1、5、10、25、35、50、75、100 和 150 mm/d, 与吴启树等(2017)的最优 TS 评分方法保持一致。设定订正的上、下界分别为 250 和 0.1 mm/d, 分别用于防止极端降水外插导致过度订正和小雨消空。当原始预报大于订正上界时, 保留原值不订正; 当订正后的预报小于订正下界时, 设为 0 降水。

2.5 最优 TS 评分法

最优 TS 评分法与频率匹配法有相似之处, 也使用训练期内的确定性预报和对应的实况资料建立原始预报和预报订正值的对应关系, 但目的是通过对雨区面积进行缩放使得不同降水等级上的 TS 评分最大。参考吴启树等(2017)的最优 TS 评分算法

$$y = \begin{cases} 0 & x < F_1 \\ O_k + (O_{k+1} - O_k) \cdot \frac{x - F_k}{F_{k+1} - F_k} & F_k \leq x < F_{k+1}, \\ x \cdot \frac{O_N}{F_N} & x \geq F_N \end{cases} \quad k = 1, \dots, N-1 \quad (13)$$

式中, x 为原始降水预报值, y 为降水预报订正值, O_k 为第 k 等级降水阈值(分别取 0.1、1、5、10、25、35、50、75、100、150 mm/d), F_k 为原始预报订正到 O_k 时, 该降水等级对应的原始预报降水值。此外, 吴启树等(2017)还将 35 mm/d 以上降水的订正系数限制在 0.8 到 1.6 范围内以避免由于高值降水样本偏少导致的订正系数不稳定问题。

2.6 最优百分位法

最优百分位法使用训练期内的集合预报和对应的实况资料, 在不同等级降水上寻找 TS 评分最大的集合分位数(代刊等, 2018)。这里选取的降水等级阈值 t_k 与 2.4 节中的 10 个降水阈值相同。对每个 t_k , 计算训练期内所有样本不同集合百分位 0%, 2%, 4%, \dots , 98%, 100% 的总体 TS 评分, 找到评分最优的集合百分位 X_k , 然后按照 t_k 从大到小的顺序依次进行判断: 若原始集合预报的 X_k 百分位值 $P_{X_k} \geq t_k$, 则订正值为 P_{X_k} , 否则对下一个 t_k 进行判断; 若所有判断都不满足, 则订正值为 0。

2.7 概率匹配法

概率匹配法(陈博宇等, 2015)将集合成员的量

级和集合平均的空间分布相结合,将实时的集合预报数据直接处理成单一预报订正产品,订正过程不包含历史观测信息。假定预报区域内包含 N 个格点(或站点),集合成员数为 M ,概率匹配算法具体如下:首先,计算区域内 N 个格点(或站点)的集合平均,并对集合平均进行排序;其次,对区域内 N 个格点(或站点)共 $N \times M$ 个集合成员进行排序,再划分长度为 M 的 N 个模块;然后,计算每个模块的平均值,并按模块序号替换相同序号的集合平均值,生成概率匹配订正场。

3 理想模型与个例订正

3.1 理想模型

为了研究不同干湿偏差和雨区位移偏差情形下,频率匹配和最优TS评分法对原始预报(RAW)的订正效果,构建了如图1所示的理想模型,其中实况观测雨区是半径为1的单位圆,预报雨区是半径为 r 的圆,假设雨区的质心和圆中心重合,预报和观测雨区的中心距离为 d ,通过调节 r 值和 d 值的大小可分别设定不同干湿偏差和雨区位移偏差的预报情形。这里的干(湿)偏差是指降水预报的面积小于(大于)降水观测的面积,对应频率偏差小于(大于)1。设置两种雨区位移偏差情形(图2):(1) $d=0.5$,代表小位移偏差;(2) $d=1.5$,代表大位移偏差。此外,设定4种干湿偏差情形(图2):(1) $FB=0.5$,代表强干偏差;(2) $FB=0.8$,代表弱干偏差;(3) $FB=2.0$,代表强湿偏差;(4) $FB=1.2$,代表弱湿偏差。由于 FB 等于预报雨区和观测雨区的面积比,即

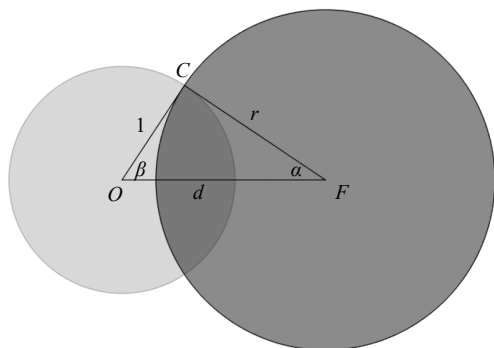


图1 理想模型(其中实况雨区是圆心为 O 的单位圆,预报雨区是圆心为 F 半径为 r 的圆,预报和观测雨区的圆心距为 d)

Fig. 1 An ideal model (the unit circle O is the observed rain area and circle F with a radius of r is the forecast rain area, the central distance between forecast and observed rain areas is d)

$FB=r^2$,故设定 FB 值与设定半径 r 值等效。

图2给出了不同位移和干湿偏差情形下,原始降水预报、频率匹配与最优TS评分法对应的理想模型以及理想模型的TS、频率偏差评分。由于是理想模型分析,这里的最优TS评分法不设置订正系数的限制。从图2可以归纳出以下几点:(1)频率匹配法订正消除了干湿偏差($FB=1$),而最优TS评分法订正总是伴随着湿偏差($FB>1$,严格地讲最优TS评分法仅当 $d=0$ 时 $FB=1$),且原始预报的位移偏差越大,最优TS评分法订正的湿偏差越大;(2)当原始预报位移误差较小时,频率匹配法与最优TS评分法订正的TS评分差异不大,都能显著改进除弱湿偏差以外的原始预报的TS评分;(3)当原始预报的位移偏差较大时,频率匹配法仅能改进干偏差情形的TS评分,而最优TS评分法可以改进所有干湿偏差情况的TS评分。

图3a给出了不同位移偏差 d 情形下,理想模型的TS评分随预报雨区半径(r)的变化。可以看出,TS评分峰值对应的预报雨区半径 r 总是大于1(仅当 $d=0$ 时TS峰值对应的 $r=1$),与图2中最优TS评分法总是伴随湿偏差一致。对于较小的位移偏差 $d=0.5$ (图3a黑色实线):频率匹配法订正($r=1$)的TS评分接近TS峰值,与最优TS评分法订正的TS评分差异不大;而弱的湿偏差(r 略大于1)对应的TS评分也接近TS峰值,因此频率匹配和最优TS评分法难以改进包含弱湿偏差的原始预报的TS评分。对于较大的位移误差 $d=1.5$ (图3a点划线):TS评分在干偏差区间(0,1)上单调递增,因此频率匹配订正($r=1$)能改进干偏差情形的TS评分;而在湿偏差区间(1,3.58)上的TS评分均不低于 $r=1$ 时的TS评分,因此频率匹配法难以在原始预报包含湿偏差时提升TS评分。图3b给出了最优TS评分(TS_{max})及相应的预报雨区半径(r_{OTS})、频率偏差(FB_{OTS})随位移偏差 d 的变化。可见,随着 d 的增大,最优TS评分法订正的湿偏差(FB_{OTS})也随之增大,且订正后的最优TS评分(TS_{max})也迅速下降。

3.2 订正个例

为了进一步直观理解4种订正方法的基本特性,本节采用2019年第11号台风“白鹿”的强降水(8月25日08时—26日08时)个例进行订正展示。图4给出了2019年8月22日20时(北京时)起报

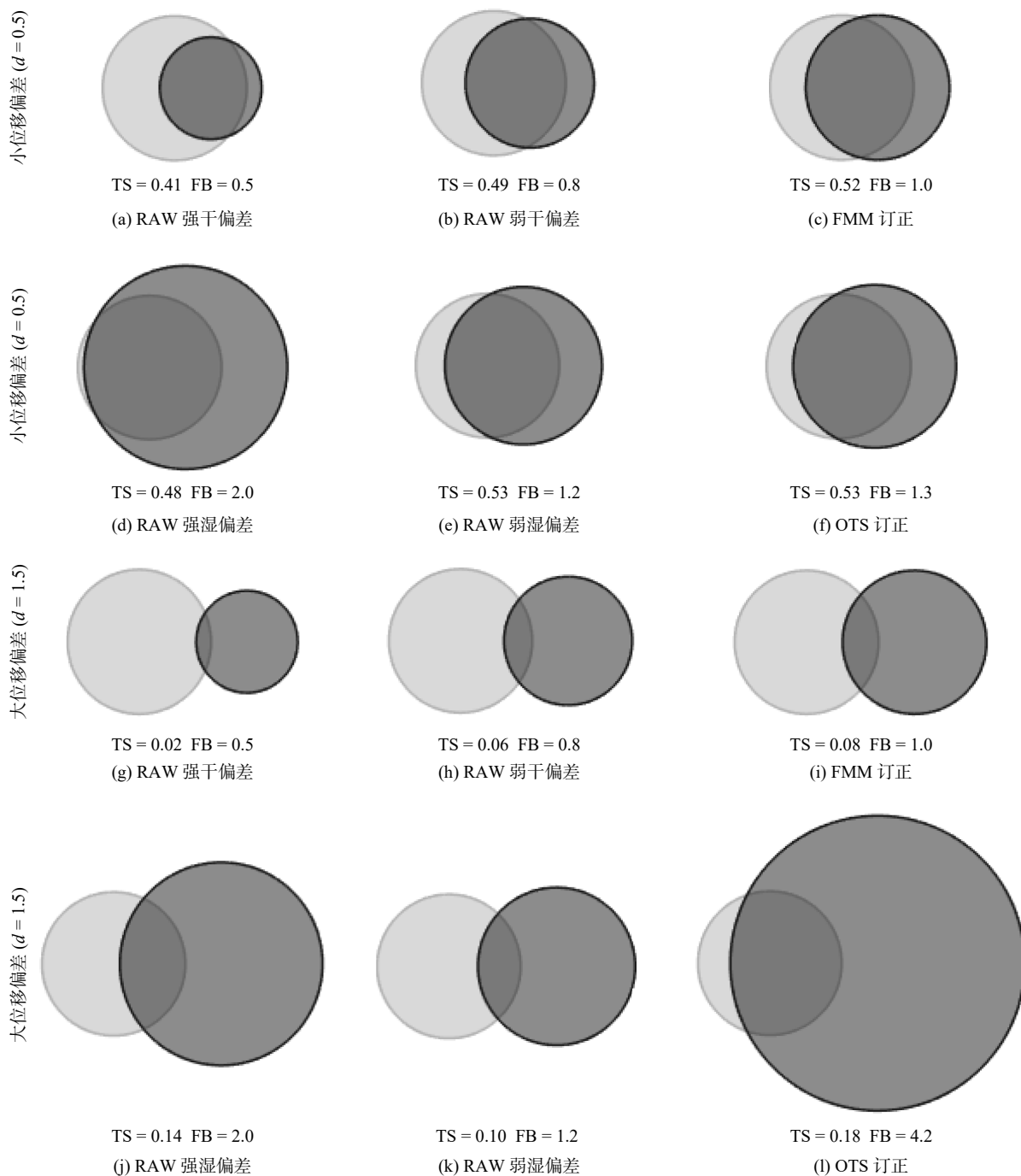


图 2 不同位移偏差和干湿偏差情形下,原始预报、频率匹配、最优 TS 评分法对应的理想模型(观测:浅色阴影,预报:深色阴影),以及理想模型的 TS 评分和频率偏差评分(其中的最优 TS 评分不设置订正系数的限制)

Fig. 2 Ideal model (observation: light shaded, forecast: dark shaded) and its TS, FB of RAW, FMM and OTS under different displacement bias d and dry/wet biases situations (the OTS has no limitations for calibration coefficient)

的 60—84 h 原始降水预报、4 种降水预报订正产品和相应的实况,由站点降水通过反距离权重插值到格点上。表 2 给出了该个例中原始降水预报和 4 种降水预报订正产品在图示区域内站点的暴雨(50 mm/d)

检验评分。可以看出,原始预报(图 4a)相比于实况(图 4f)存在偏北的位移误差,且在暴雨量级上存在弱的干偏差,频率偏差为 0.81(表 2)。经频率匹配和最优 TS 评分订正后,暴雨雨区有所扩大而小雨

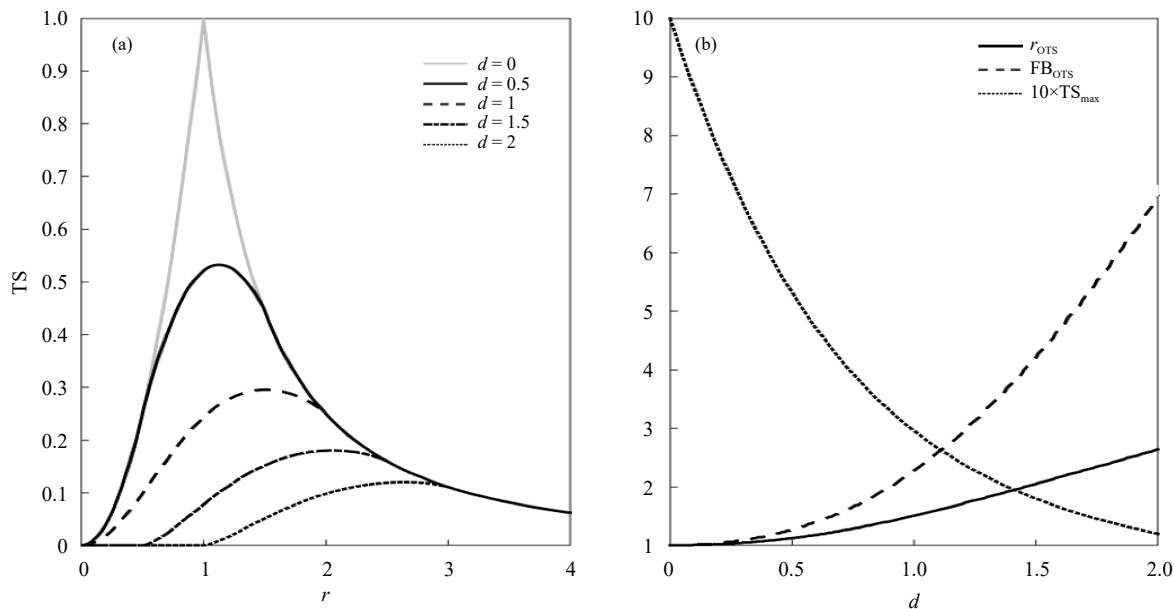


图3 (a) 不同位移偏差 d 情形下, 理想模型的 TS 评分随预报雨区半径(r)的变化, (b) 最优 TS 评分 (TS_{max}) 及相应的预报雨区半径 (r_{OTS})、频率偏差 (FB_{OTS}) 随位移偏差 d 的变化

Fig. 3 (a) Under different displacement bias d , TS as a function of the radius r of forecast rain area in the ideal model; (b) the optimal TS (TS_{max}), the corresponding radius of forecast rain area (r_{OTS}), and frequency bias (FB_{OTS}) as a function of the displacement bias d

雨区有所减小(图 4b、c), 雨区面积更接近实况, 频率偏差更接近 1(表 2), TS 评分也有明显提升。这一点与 3.1 节中理想模型分析的“频率匹配和最优 TS 评分法均能提升干偏差情形的 TS 评分”一致。需要注意的是, 频率匹配与最优 TS 评分法仅能订正雨带的量级, 无法订正雨带的位移误差, 因此在提升 TS 评分和降低漏报率的同时, 也存在虚警率略偏大的问题(表 2)。最优百分位法则基于历史集合预报与实况资料计算出该个例对应的历史训练期 TS 评分最大的集合百分位为 90%, 故最优百分位法对该个例订正的结果为原始集合预报的 90% 集合分位数(图 4d)。从表 2 中可以看出, 虽然最优百分位法的 TS 评分最高、漏报率最低, 但由于雨区面积过大(图 4d)也导致虚警率最高。简单的集合平均由于平滑作用常常导致强降水预报的降水量级偏低, 而概率匹配法既保留了集合平均的空间分布, 又参考了集合成员的降水量级。图 4e 为概率匹配法订正的降水预报, 具有强降水落区向集合平均中心集中的特征。由于概率匹配法不使用历史资料训练, 在本个例中无法根据历史系统偏差对集合预报进行订正, 因此对暴雨的量级预报偏弱, 订正效果不佳(表 2)。综合对比 4 种订正法的降水落

区, 可以看出频率匹配和最优 TS 评分法主要改变原始预报的降水量级而并不改变落区, 最优百分位与概率匹配法由于使用了集合预报信息可同时改变原始预报的降水量级和落区。

4 总体检验评估

4.1 全年预报性能

小雨量级(图 5), 原始模式预报存在较强的湿偏差, 不同预报时效的频率偏差都在 1.5 以上。频率匹配法订正显著消除了小雨的强湿偏差, 不同预报时效的频率偏差都接近 1。虽然频率匹配法订正后总体的雨区面积与实况较一致, 但对小雨的漏报率很高, 可以推断这是由于雨区的位移偏差较大造成的。最优 TS 评分和最优百分位法在一定程度上减小了小雨的湿偏差, 而概率匹配法反而略增大了原始模式的湿偏差。这是由于概率匹配法未引入历史实况信息对原始预报的湿偏差进行修正, 而是利用集合预报直接生成定量降水预报产品。从频率匹配法的频率偏差和漏报率随预报时效的变化分析可知, 雨区的总体面积保持不变的情况下漏报率不断增大, 说明雨区的位移偏差随预报时效的延长不断增大。因此, 在较长预报时效下, 原始预报

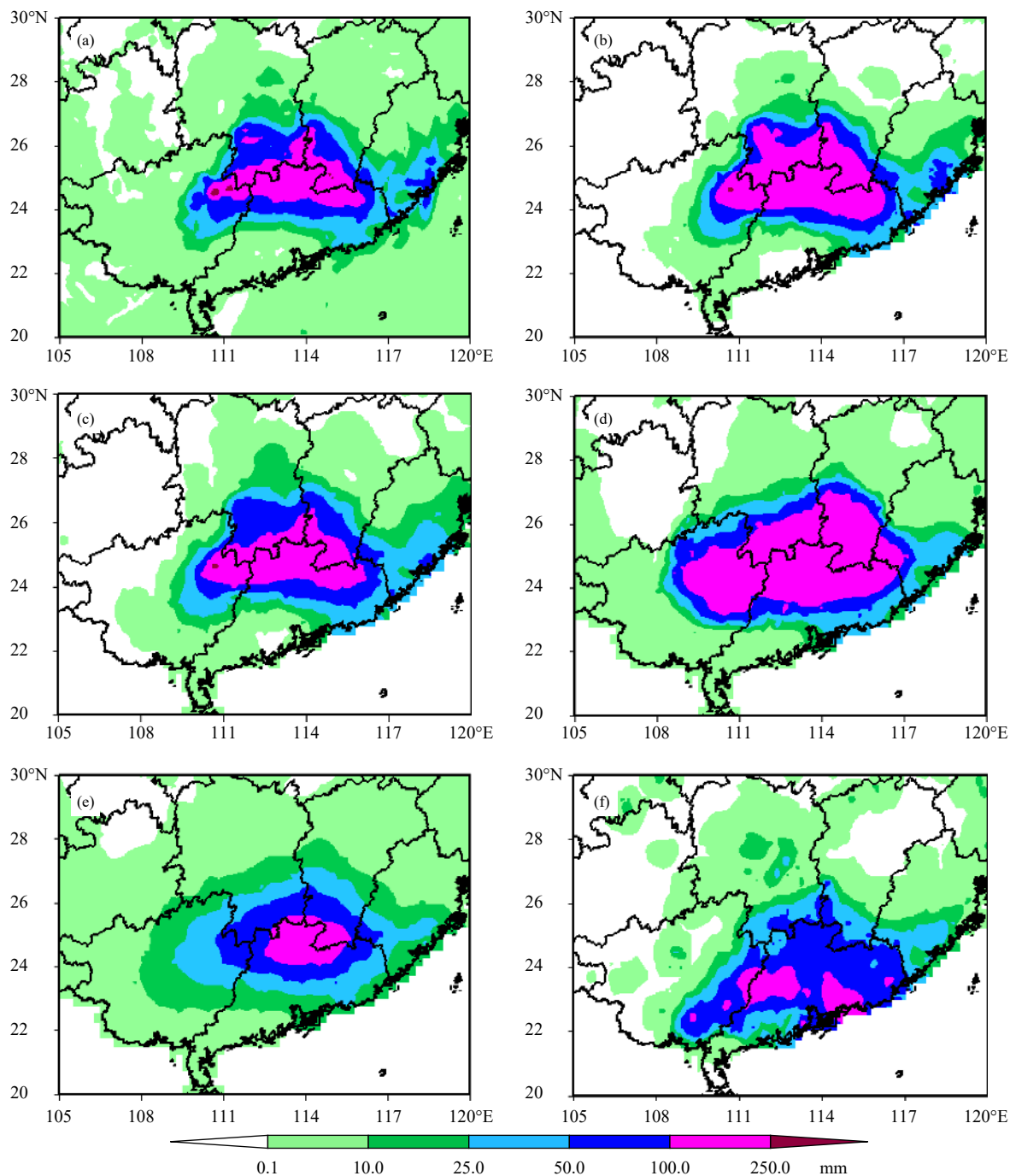


图4 2019年8月22日20时起报的60—84 h原始降水预报(a)、频率匹配法订正(b)、最优TS评分法订正(c)、最优百分位法订正(d)和概率匹配法订正(e)的结果以及2019年第11号台风“白鹿”的强降水(8月25日08时—26日08时)实况(f)

Fig. 4 Observation (f), the 60—84 h RAW (a) precipitation forecasts initialized at 20:00 UTC 22 August 2019, using FMM (b), OTS (c), OP (d) and PM (e) calibration during the heavy rain (08:00 BT 25—08:00 BT 26 August) of Typhoon Bailu (1911)

除了湿偏差还伴随着较大的位移误差,结合理想模型的分析结论“当原始预报的位移偏差较大时,频率匹配法仅能改进干偏差情形的TS评分”可知,频

率匹配法无法在较长预报时效下改进小雨的TS评分,这一点也从TS评分随预报时效的变化图(图5b)上得到了印证。最优TS评分法的TS评分在1.5—

表 2 2019 年第 11 号台风“白鹿”的强降水个例的原始预报、频率匹配、最优 TS 评分、最优百分位和概率匹配法的暴雨 (50 mm/d) 检验评分

Table 2 Verification scores of torrential rain (50 mm/d) of RAW, FMM, OTS, OP, and PM in the heavy precipitation event of Typhoon Bailu

	RAW	频率匹配	最优TS评分	最优百分位	概率匹配
频率偏差	0.81	0.98	0.93	1.47	0.67
TS评分	0.27	0.30	0.29	0.32	0.29
漏报率	0.62	0.55	0.57	0.40	0.62
虚警率	0.53	0.54	0.54	0.59	0.44

注: 加粗的数值为评分最优值。

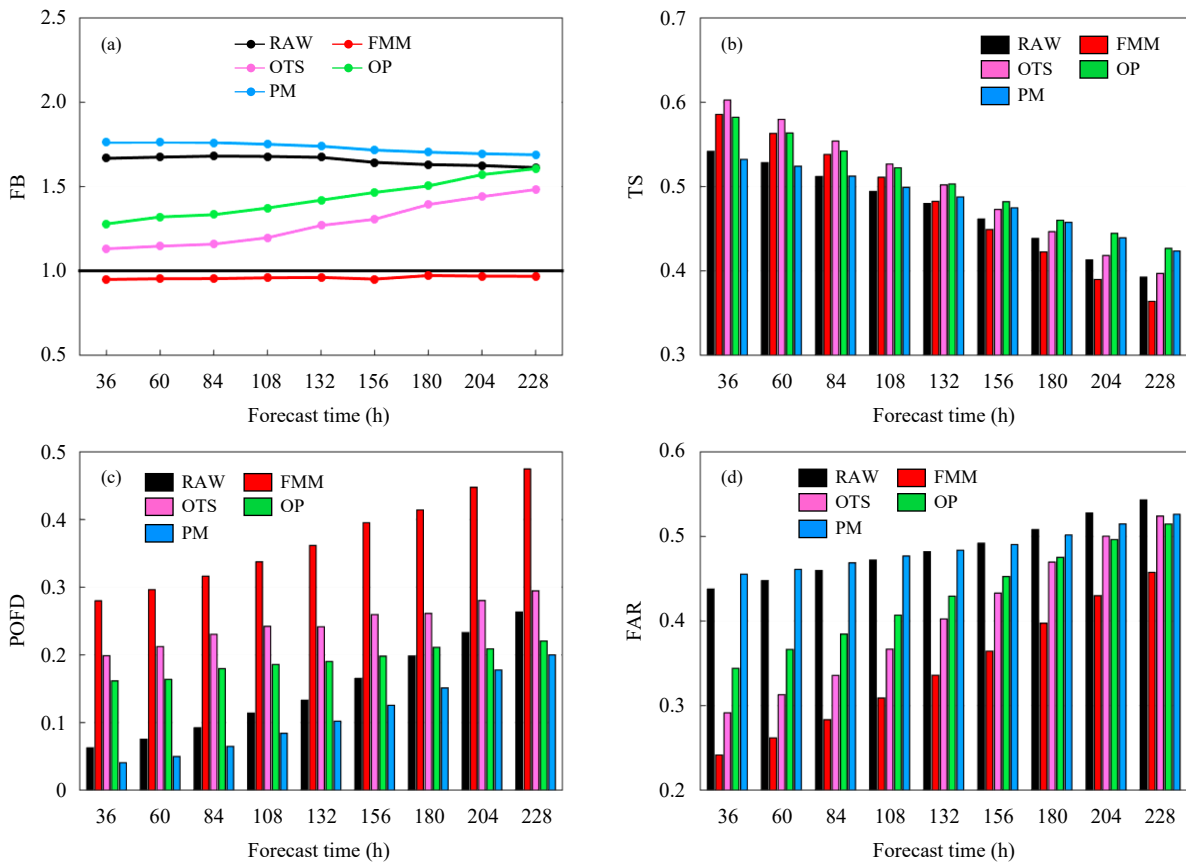


图 5 原始预报、频率匹配、最优 TS 评分、最优百分位和概率匹配法全年不同预报时效小雨量级 (0.1 mm/d) 的频率偏差 (a)、TS 评分 (b)、漏报率 (c) 和虚警率 (d)

Fig. 5 Annual frequency bias (a), threat score (b), probability of false detection (c) and false alarm ratio (d) of light rain (0.1 mm/d) for RAW, FMM, OTS, OP, and PM at different lead times

5.5 d 预报时效下正技巧最高, 而在更长的预报时效下 TS 评分不如最优百分位和概率匹配法, 这是由于随着预报时效的进一步延长, 预报的不确定性增大, 集合预报相比确定性预报的优势逐渐凸显。考虑到最优百分位法相比概率匹配法增加了历史实况信息, 因此最优百分位法在各个预报时效下 TS 评分的表现好于概率匹配法。

中雨量级(图 6), 原始模式预报存在较弱的湿偏差, 不同预报时效的频率偏差都在 1.3 左右。频率匹配法订正显著消除了中雨的弱湿偏差, 不同预报时效的频率偏差都接近 1。根据理想模型的结论可以判断频率匹配法无法改进弱湿偏差情形下的 TS 评分, 这一点从图 6b 中也得到了印证。概率匹配法的频率偏差与原始预报一致, 而最优 TS 评分

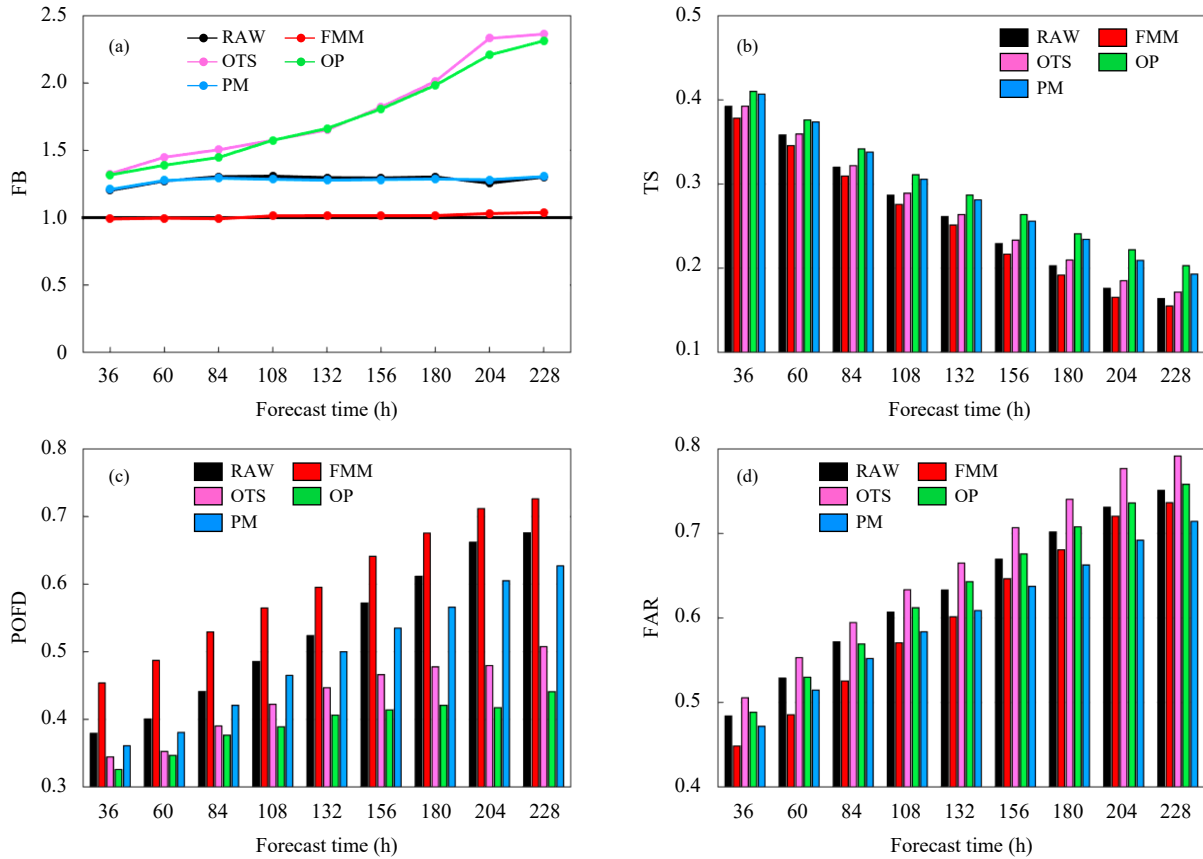


图 6 与图 5 相同, 但为中雨量级 (10 mm/d)
 Fig. 6 Same as Fig. 5 but for moderate rain (10 mm/d)

和最优百分位法均呈现较大的湿偏差, 且随预报时效的延长而增大。这一点也可以通过理想模型进行解释, 由于最优 TS 评分与最优百分位法都以 TS 评分最优为目标, 因此总是伴随着湿偏差, 且雨区的位移偏差越大, 湿偏差越大。虽然最优 TS 评分法对强降水的订正系数进行了范围限制 (35 mm/d 以上降水的订正系数限制在 0.8—1.6), 但雨区面积随降水阈值并非线性变化, 订正前、后雨区面积的增长倍数可大于降水量增长的倍数。此外, 还可以看出集合预报对于存在较弱湿偏差的中雨量级降水预报体现了明显的优势, 最优百分位和概率匹配法相比原始预报在没有增大虚警率和漏报率的情况下, 提升了所有预报时效的 TS 评分, 这得益于集合预报信息的加入, 而仅通过对确定性预报进行量级订正 (频率匹配和最优 TS 评分法) 无法显著改进原始预报。

大雨量级 (图 7), 原始预报频率偏差接近 1, 几乎不存在系统偏差, 此时频率匹配法无法改进原始

预报。最优 TS 评分法虽然略微改进原始预报的 TS 评分, 但也伴随着强的湿偏差和虚警率最高的问题。基于集合预报的最优百分位和概率匹配法再一次发挥了优势, 但两者的表现有所不同。最优百分位法倾向于增加一定的湿偏差来达到 TS 评分的最大, 与最优 TS 评分法相比, 最优百分位法的湿偏差与之量级相当, 但 TS 评分、漏报率和虚警率均好于最优 TS 评分法。这是由于集合预报信息的引入在一定程度上修正了确定性预报的落区。概率匹配法在不同评分上的表现较为平衡, 没有明显的干湿偏差, 且 TS 评分、漏报率、虚警率比原始预报均有改进, 尤其是虚警率的降低最明显, 这再一次体现了集合预报的优势。

暴雨量级 (图 8), 原始预报存在较弱的干偏差。频率匹配法仍然具有 4 种订正方法中最优的频率偏差, 但相比小雨、中雨和大雨那样近乎完美的频率偏差, 暴雨仍存在微小的湿偏差。这是由于暴雨量级的训练样本较少, 频率偏差存在年际变化,

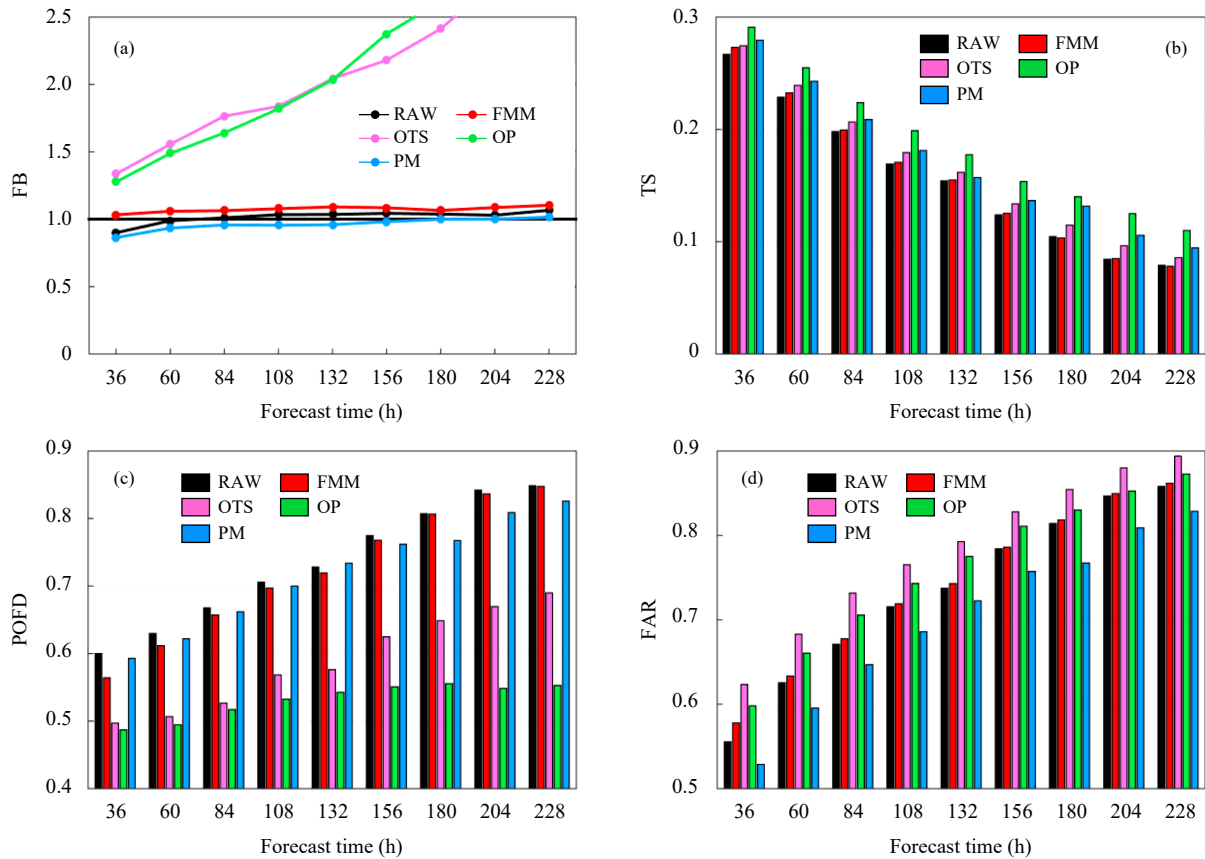


图 7 与图 5 相同, 但为大雨量级 (25 mm/d)

Fig. 7 Same as Fig. 5 but for heavy rain (25 mm/d)

或是由于模式升级更新导致在该量级降水预报的系统偏差出现了微小的变化。频率匹配法改进了 1.5—5.5 d 预报时效的暴雨 TS 评分, 与理想模型的结论“频率匹配可以改变干偏差情形的 TS 评分”一致。对于更长的预报时效, 原始预报的频率偏差接近 1, 但漏报率、虚警率都显著增大, 说明暴雨的位移误差也显著增大, 甚至出现较多原始预报与实况没有重叠雨区的情况, 此时频率匹配法因保持雨区面积与实况一致而无法提升 TS 评分, 最优 TS 评分法通过适当的扩大雨区范围增加与实况雨区的重叠面积在一定程度上提升了 TS 评分。需要注意的是, 由于最优 TS 评分法增加了订正系数的范围限制, 因此湿偏差并不是很大。同时, 测试了不加订正系数范围限制的最优 TS 评分, 发现对小雨、中雨、大雨都没有影响(因为都小于 35 mm/d), 而对暴雨的频率偏差与最优百分位法类似存在很大的湿偏差, 且可进一步提升 TS 评分, 但提升之后 TS 评分总体上仍低于最优百分位法(图略)。最优百分位法虽然显

著提升了不同预报时效的 TS 评分, 但也存在湿偏差和虚警率过大的问题。概率匹配法对暴雨的干偏差比原始预报还大, 说明在暴雨量级上集合预报本身相比实况仍然量级偏低, 因此无法改进暴雨的预报性能。

4.2 季节预报性能

评估不同定量降水预报客观订正方法在不同季节的预报性能表现, 对于实际业务更具指导意义。图 9 给出以 60—84 h 预报时效为例进行的性能检验评估, 同一张图上可同时展示 4 种检验评分的结果, 其中实曲线代表 TS 评分等值线, 虚直线的斜率为频率偏差, 横坐标为 1-虚警率(FAR), 纵坐标为 1-漏报率(POFD)。从实况基础覆盖率(BR)可以看出, 不同等级的降水覆盖率排序都是夏季 > 春季 > 秋季 > 冬季。总体来看, 频率匹配法在不同季节、不同降水等级情形下的频率偏差都是最优的, 最接近对角线, 有效减小了原始预报雨区面积的系统偏差。对于晴雨(同小雨), 实况样本量比较

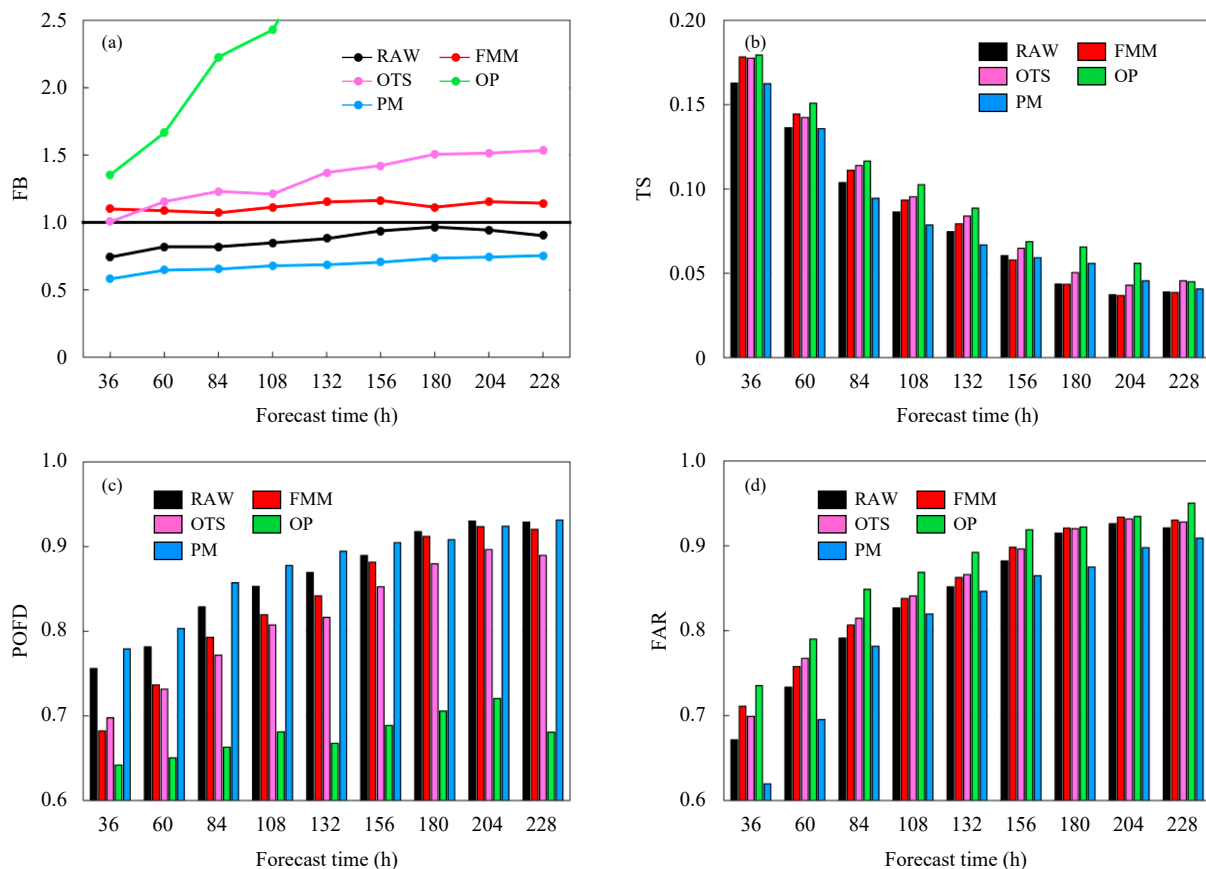


图 8 与图 5 相同, 但为暴雨量级 (50 mm/d)

Fig. 8 Same as Fig. 5 but for torrential rain (50 mm/d)

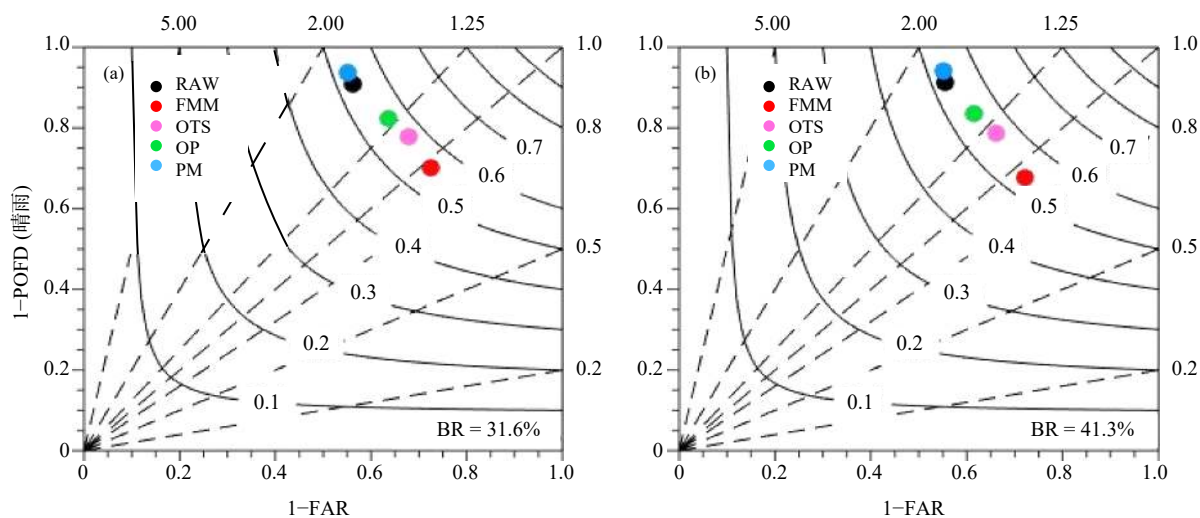
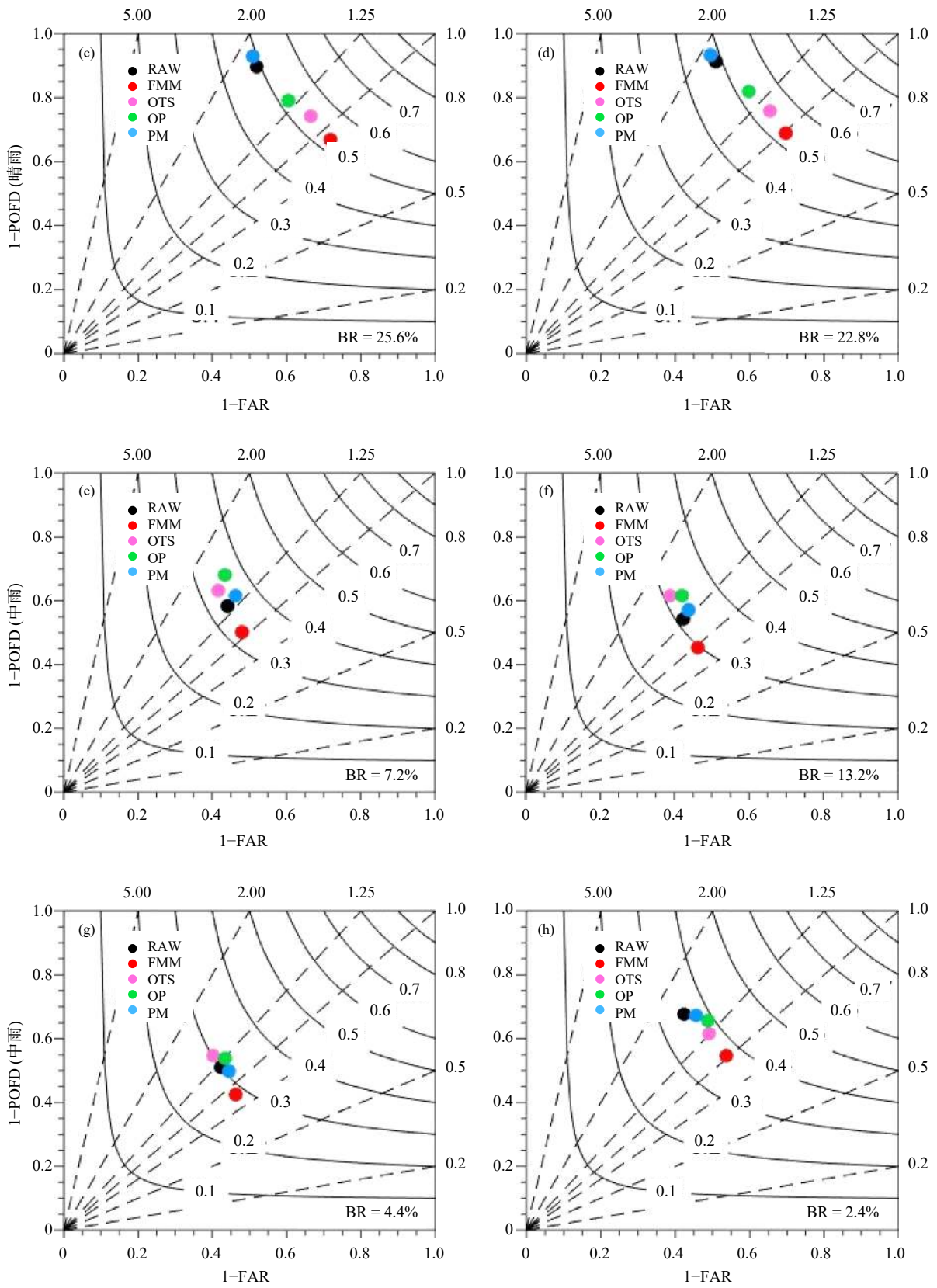


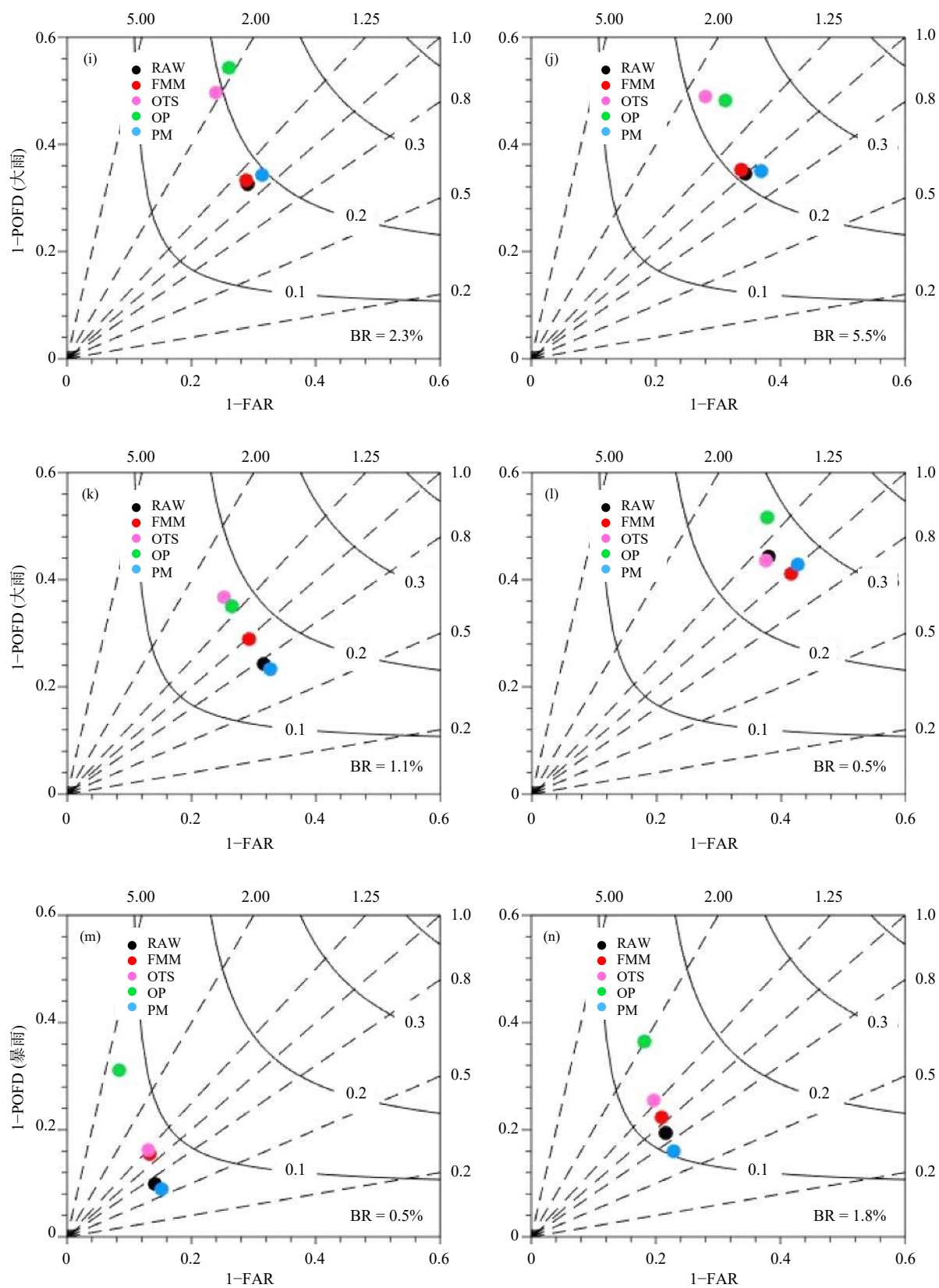
图 9 不同季节 (a, e, i, m. 春季, b, f, j, n. 夏季, c, g, k, o. 秋季, d, h, l, p. 冬季)、不同降水量级 (a—d. 小雨 (0.1 mm/d), e—h. 中雨, i—l. 大雨, m—p. 暴雨) 的原始预报、频率匹配、最优 TS 评分、最优百分位和概率匹配法的性能 (BR 为样本的基础观测覆盖率)

Fig. 9 Performance diagrams of RAW, FMM, OTS, OP and PM for different seasons (a, e, i, m. spring; b, f, j, n. summer; c, g, k, o. autumn; d, h, l, p. winter) and rain thresholds (a—d. light rain (0.1 mm/d), e—h. moderate rain (10 mm/d), i—l. heavy rain (25 mm/d), m—p. torrential rain (50 mm/d)) (BR is the sample base rate)

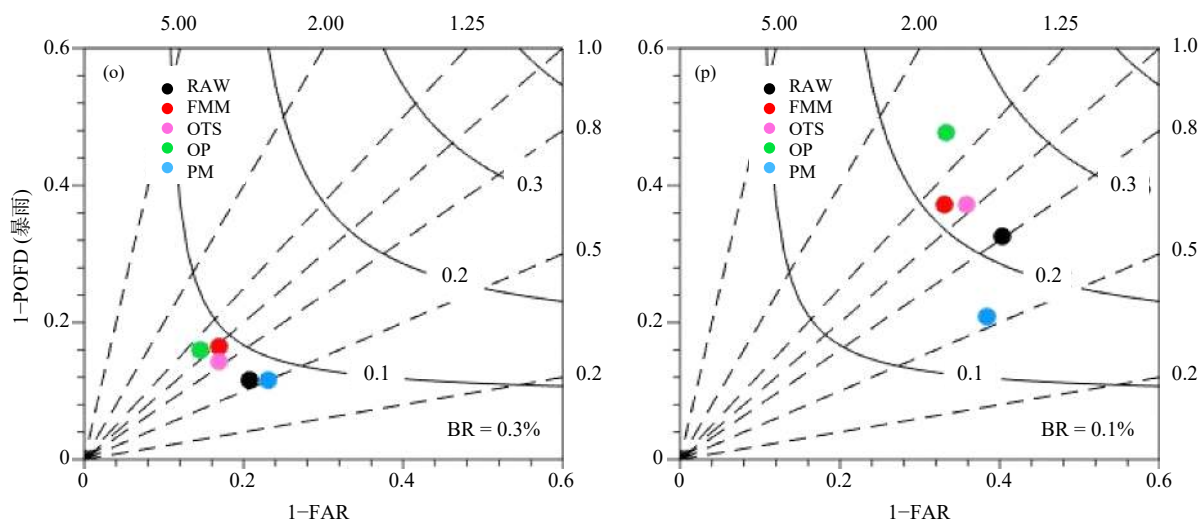


续图 9

Fig. 9 Continued



续图 9
Fig. 9 Continued



续图 9

Fig. 9 Continued

充足,即使样本量最低的冬季观测覆盖率也达到了 22.8%(图 9d),不同季节的评估结果也具有相似性。频率匹配、最优 TS 评分和最优百分位法都对原始预报的 TS 评分有小幅提升,同时降低了原始预报的湿偏差,减小了虚警率但增大了漏报率,其中频率匹配法的虚警率和漏报率达到平衡,结论与全年检验结果一致。对于中雨,最优百分位和概率匹配法对原始预报的 TS 评分有小幅提升,其中最优化百分位法的提升幅度略大一些。最优百分位法在降水较多的春、夏两季湿偏差比原始预报还大,秋季与原始预报一样,而降水较少的冬季减小了原始预报的湿偏差。对于大雨,原始预报在春、夏两季不存在干湿偏差,而秋季略偏干,冬季略偏湿。最优百分位法对原始预报的 TS 评分提升最大,但总是伴随着湿偏差,尤其在春季频率偏差达到了 2。对于暴雨,秋、冬两季出现的频率很低,观测基础覆盖率分别为 0.3% 和 0.1%。从春、夏两季来看,频率匹配和最优 TS 评分法在改进了原始预报的 TS 评分的同时,维持了较好的干湿偏差。虽然最优百分位法对夏季暴雨的 TS 评分提升最大,但湿偏差也明显,雨区总面积达到了观测的 2 倍。

4.3 不同订正方法的经济价值模型评估

图 10 以 60—84 h 预报时效为例,从参考不同预报产品进行风险决策所产生经济价值的角度,对 4 种定量降水预报客观订正方法进行评估。对于小雨(图 10a),当花费/损失在 0.05—0.2 时,只有概率

匹配法的经济价值高于原始预报,而当花费/损失在 0.5—0.7 时,另外 3 种订正方法产生了比原始预报更高的经济价值,从大到小依次是频率匹配、最优 TS 评分和最优百分位法。最优百分位法在中雨、大雨和暴雨的风险决策中体现了较大的参考价值(尤其是暴雨),经济价值分别在花费/损失区间 0.05—0.3、0.02—0.2、0.01—0.1 上高于其他预报方法。需要注意的是,当经济价值为负值时,所有预报产品在风险决策中都失去了参考价值。以暴雨为例,当花费/损失大于 0.2 时,无论参考那种预报产品进行风险决策,所产生的经济价值都是负效果。其他预报时效的结论与 60—84 h 时效一致(图略)。

5 结论与讨论

重点对比了频率匹配、最优 TS 评分、最优百分位、概率匹配 4 种定量降水预报客观订正方法的特性和预报性能。通过构建理想模型研究了频率匹配和最优 TS 评分法在不同位移和干湿偏差情形下的基本特征。选取典型降水个例,评估了 4 种定量降水预报客观订正方法在全年、不同季节、不同预报时效、不同降水量级下的表现,以及在风险决策中的经济价值。表 3 归纳了不同订正方法的基本原理,以及 2019 年在中国站点上的预报订正效果,包括不同降水量级下的频率偏差、预报技巧等信息。得出以下主要结论:

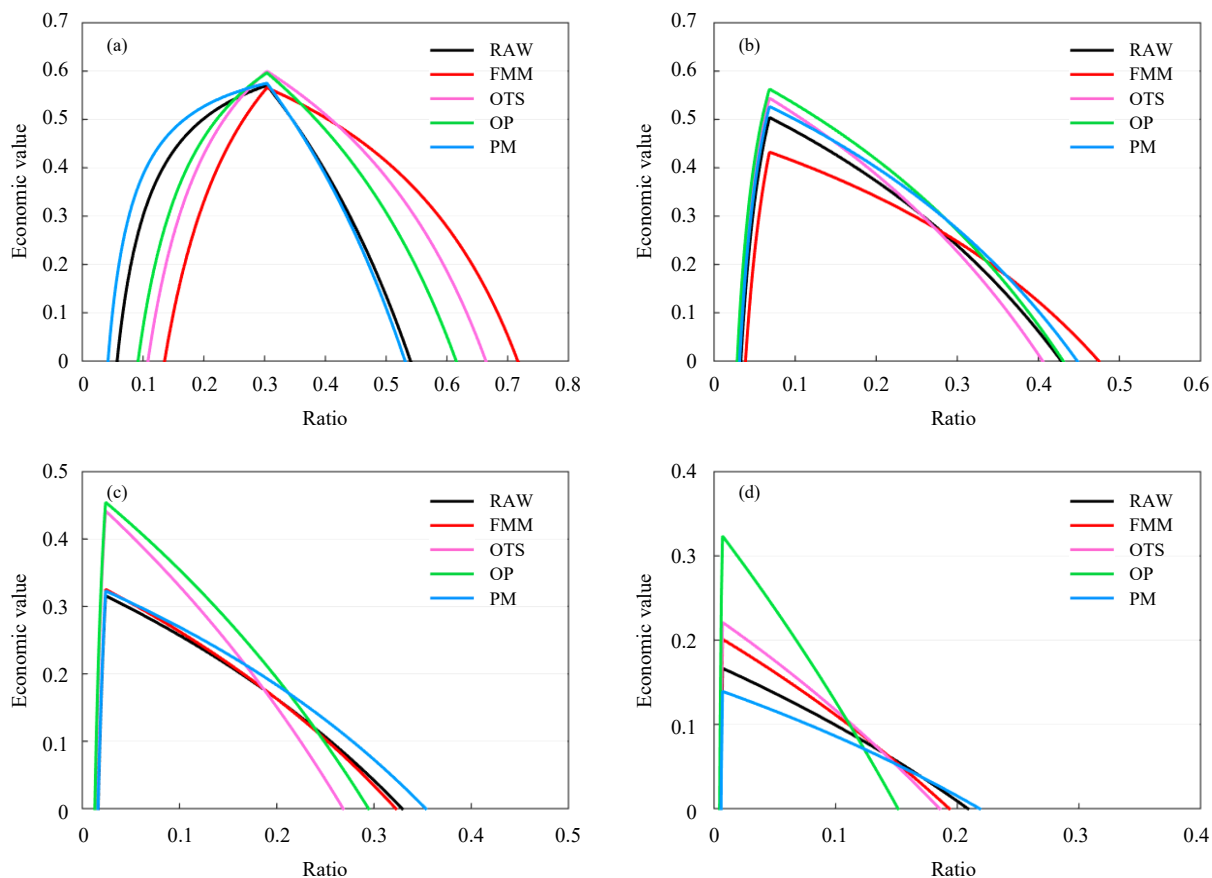


图 10 60—84 h 时效的原始预报、频率匹配、最优 TS 评分、最优百分位和概率匹配法在小雨 (a, 0.1 mm/d)、中雨 (b, 10 mm/d)、大雨 (c, 25 mm/d) 和暴雨 (d, 50 mm/d) 降水量级上的经济价值随花费/损失的变化

Fig. 10 Economic values as a function of cost/loss for 60—84 h precipitation forecasts of RAW, FMM, OTS, OP and PM at light precipitation forecasts of RAW, FMM, OTS, OP and PM at light (a, 0.1 mm/d), moderate (b, 10 mm/d), heavy (c, 25 mm/d) and torrential (d, 50 mm/d) rain thresholds

表 3 ECMWF 原始模式预报及不同订正方法在 2019 年中国 2411 个气象站的逐日降水预报中的表现 (TS 技巧是指预报产品的 TS 评分与 ECMWF 原始模式 TS 评分之差, 正值代表正技巧, 负值代表负技巧)

Table 3 Performance of the original ECMWF and four different calibration methods in the daily precipitation forecasts of 2411 meteorological stations during 2019 (The TS skill score indicates the difference of TS between the calibrated and the original ECMWF forecasts. Positive values represent skillful while negative values represent unskillful)

预报方法	ECMWF	频率匹配	最优TS评分	最优百分位	概率匹配
输入资料	确定性预报	确定性预报和实况	确定性预报和实况	集合预报和实况	集合预报
训练目标		频率偏差最优	TS评分最优	TS评分最优	
量级订正		有	有	有	有
落区订正		无	无	有	有
小雨频率偏差	湿	无	湿	湿	湿
中雨频率偏差	湿	无	湿	湿	湿
大雨频率偏差	无	无	湿	湿	无
暴雨频率偏差	干	无	湿	湿	干
小雨TS技巧		正	正	正	无
中雨TS技巧		负	无	正	正
大雨TS技巧		无	正	正	正
暴雨TS技巧		正	正	正	无

(1) ECMWF 确定性预报模式的整体降水系统偏差表现为: 小雨存在较强湿偏差、中雨存在较弱湿偏差、大雨无干湿偏差、暴雨存在弱的干偏差。

(2) 频率匹配和最优 TS 评分法基于历史训练资料直接对确定性预报进行量级订正, 无法订正原始预报的位移误差, 因此降水落区的形态也与原始预报相似; 最优百分位和概率匹配法均利用集合预报信息生成确定性预报产品, 可以改变降水落区的形态。

(3) 频率匹配法以频率偏差最优为训练目标, 能有效地消除模式的干湿偏差; 当雨区的位移误差较小且存在明显的干湿偏差时, 频率匹配法可以提升 TS 评分。以 TS 评分最优为训练目标的最优 TS 评分法和最优百分位法则表现为湿偏差, 特别是春、夏两季最优百分位的湿偏差较大; 最优 TS 评分可在一定程度上提升小雨、大雨和暴雨的 TS 评分, 但无法改善中雨的 TS 评分; 最优百分位法显著提升了所有量级降水的 TS 评分, 这主要得益于集合预报信息的引入, 在较长预报时效下的改进优势尤其凸显。概率匹配法虽然也使用集合预报信息, 但没有利用历史实况对集合预报进行订正, 因此仅能在一定程度上改进大雨、中雨和较长预报时效小雨的 TS 评分。

(4) 不同定量降水预报客观订正方法的风险决策经济价值模型评估表明: 对于小雨, 概率匹配和频率匹配法在不同花费/损失区间上具有较高的经济价值; 在中雨、大雨和暴雨的风险决策中, 最优百分位法体现了较大的参考价值, 尤其是暴雨。

以上 4 种定量降水预报客观订正方法在模式雨带位移偏差较大时仍存在一定的局限性, 此时最优 TS 评分和最优百分位法常常伴随着显著的湿偏差, 在实际业务使用时需要注意。虽然最优百分位和概率匹配法利用集合预报信息可在一定程度上改变降水落区的形态, 但对雨带的位移偏差调整幅度还是非常有限, 未来可尝试借助面向对象的检验评估方法 (Method for Object-based Diagnostic Evaluation, 简称 MODE) (Davis, et al, 2006, 2009) 来改进订正雨带的位移偏差。此外, 这 4 种订正方法都是基于降水量本身建模的, 未来可考虑引入更多的物理量和大数据信息, 探索基于机器学习的定量降水预报客观订正方法。

参考文献

- 毕宝贵, 代刊, 王毅等. 2016. 定量降水预报技术进展. *应用气象学报*, 27(5): 534-549. Bi B G, Dai K, Wang Y, et al. 2016. Advances in techniques of quantitative precipitation forecast. *J Appl Meteor Sci*, 27(5): 534-549 (in Chinese)
- 陈博宇, 代刊, 郭云谦. 2015. 2013 年汛期 ECMWF 集合统计量产品的降水预报检验与分析. *暴雨灾害*, 34(1): 64-73. Chen B Y, Dai K, Guo Y Q. 2015. Precipitation verification and analysis of ECMWF ensemble statistic products in 2013 flooding season. *Torrent Rain Dis*, 34(1): 64-73 (in Chinese)
- 陈博宇, 郭云谦, 代刊等. 2016. 面向台风暴雨的集合预报成员优选订正技术研究及应用试验. *气象*, 42(12): 1465-1475. Chen B Y, Guo Y Q, Dai K, et al. 2016. Research for the ensemble member optimization correction technique on typhoon rainstorm forecast and its application experiment. *Meteor Mon*, 42(12): 1465-1475 (in Chinese)
- 代刊, 曹勇, 钱奇峰等. 2016. 中短期数字化天气预报技术现状及趋势. *气象*, 42(12): 1445-1455. Dai K, Cao Y, Qian Q F, et al. 2016. Situation and tendency of operational technologies in short and medium range weather forecast. *Meteor Mon*, 42(12): 1445-1455 (in Chinese)
- 代刊, 朱跃建, 毕宝贵. 2018. 集合模式定量降水预报的统计后处理技术研究综述. *气象学报*, 76(4): 493-510. Dai K, Zhu Y J, Bi B G. 2018. The review of statistical post-process technologies for quantitative precipitation forecast of ensemble prediction system. *Acta Meteor Sinica*, 76(4): 493-510 (in Chinese)
- 杜钧, 李俊. 2014. 集合预报方法在暴雨研究和预报中的应用. *气象科技进展*, 4(5): 6-20. Du J, Li J. 2014. Application of ensemble methodology to heavy-rain research and prediction. *Adv Meteor Sci Technol*, 4(5): 6-20 (in Chinese)
- 李芳. 2012. 基于多模式集合方案的中国东部夏季降水概率季度预测. *气象学报*, 76(4): 493-510. Li F. 2012. Probabilistic seasonal prediction of summer rainfall over East China based on multi-model ensemble schemes. *Acta Meteor Sinica*, 76(4): 493-510 (in Chinese)
- 李俊, 杜钧, 陈超君. 2014. 降水偏差订正的频率(或面积)匹配方法介绍和分析. *气象*, 40(5): 580-588. Li J, Du J, Chen C J. 2014. Introduction and analysis to frequency or area matching method applied to precipitation forecast bias correction. *Meteor Mon*, 40(5): 580-588 (in Chinese)
- 李俊, 杜钧, 陈超君. 2015. “频率匹配法”在集合降水预报中的应用研究. *气象*, 41(6): 674-684. Li J, Du J, Chen C J. 2015. Applications of "Frequency-Matching" method to ensemble precipitation forecasts. *Meteor Mon*, 41(6): 674-684 (in Chinese)
- 栾菲. 2017. 六项目通过天气预报科技成果准入认证评审. *中国气象报*, 2017-03-30. Luan F. 2017. Six projects have passed the certification and evaluation of scientific and technological achievements in weather forecasting. *China Meteorological News*, 2017-03-30 http://www.cma.gov.cn/2011xwzx/2011xqxkj/2011xkjdt/201703/t20170330_403948.html (in Chinese)

- 吴启树, 韩美, 刘铭等. 2017. 基于评分最优化的模式降水预报订正算法对比. *应用气象学报*, 28(3): 306-317. Wu Q S, Han M, Liu M, et al. 2017. A comparison of optimal-score-based correction algorithms of model precipitation prediction. *J Appl Meteor Sci*, 28(3): 306-317 (in Chinese)
- 徐姝, 尉英华, 熊明明等. 2018. 频率匹配法在海河流域 ECMWF 集合预报融合产品中的应用研究. *气象与环境学报*, 34(4): 11-17. Xu S, Wei Y H, Xiong M M, et al. 2018. Application of frequency-matching method to ECMWF ensemble statistic fusing prediction products in the Haihe river basin. *J Meteor Environ*, 34(4): 11-17 (in Chinese)
- 徐姝, 熊明明, 王颖等. 2018. 改进的 ECMWF 集合预报融合产品在海河流域的检验与分析. *气象与环境科学*, 41(4): 41-46. Xu S, Xiong M M, Wang Y, et al. 2018. Verification and analysis of improved ECMWF ensemble forecast fusing product in Haihe river basin. *J Meteor Environ*, 41(4): 41-46 (in Chinese)
- 智协飞, 季晓东, 张璟等. 2013. 基于 TIGGE 资料的地面气温和降水的多模式集成预报. *大气科学学报*, 36(3): 257-266. Zhi X F, Ji X D, Zhang J, et al. 2013. Multimodel ensemble forecasts of surface air temperature and precipitation using TIGGE datasets. *Trans Atmos Sci*, 36(3): 257-266 (in Chinese)
- 周迪, 陈静, 陈朝平等. 2015. 暴雨集合预报-观测概率匹配订正法在四川盆地的应用研究. *暴雨灾害*, 34(2): 97-104. Zhou D, Chen J, Chen C P, et al. 2015. Application research on heavy rainfall calibration based on ensemble forecast vs. observed precipitation probability matching method in the Sichuan basin. *Torrent Rain Dis*, 34(2): 97-104 (in Chinese)
- 宗志平, 代刊, 蒋星. 2012. 定量降水预报技术研究进展. *气象科技进展*, 2(5): 29-35. Zong Z P, Dai K, Jiang X. 2012. The research progress of quantitative precipitation forecast. *Adv Meteor Sci Technol*, 2(5): 29-35 (in Chinese)
- Berenguer M, Surcel M, Zawadzki I, et al. 2012. The diurnal cycle of precipitation from continental radar mosaics and numerical weather prediction models. Part II: Intercomparison among numerical models and with nowcasting. *Mon Wea Rev*, 140(8): 2689-2705
- Buizza R. 2008. The value of probabilistic prediction. *Atmos Sci Lett*, 9(2): 36-42
- Chien F C, Jou B J D. 2004. MM5 ensemble mean precipitation forecasts in the Taiwan area for three early summer convective (Mei-Yu) seasons. *Wea Forecasting*, 19(4): 735-750
- Clark A J, Gallus Jr W A, Xue M, et al. 2009. A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles. *Wea Forecasting*, 24(4): 1121-1140
- Davis C A, Brown B G, Bullock R, et al. 2006. Object-based verification of precipitation forecasts. Part I: Methodology and application to meso-scale rain areas. *Mon Wea Rev*, 134(7): 1772-1784
- Davis C A, Brown B G, Bullock R, et al. 2009. The method for object-based diagnostic evaluation (MODE) applied to numerical forecasts from the 2005 NSSL/SPC Spring Program. *Wea Forecasting*, 24(5): 1252-1267
- Gallus Jr W A. 2010. Application of object-based verification techniques to ensemble precipitation forecasts. *Wea Forecasting*, 25(1): 144-158
- Guan H, Zhu Y J. 2017. Development of verification methodology for extreme weather forecasts. *Wea Forecasting*, 32(2): 479-491
- Hamill T M, Whitaker J S. 2006. Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application. *Mon Wea Rev*, 134(11): 3209-3229
- Jolliffe I T, Stephenson D B. 2003. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Chichester: John Wiley & Sons
- Roebber P J. 2009. Visualizing multiple measures of forecast quality. *Wea Forecasting*, 24(2): 601-608
- Surcel M, Zawadzki I, Yau M K. 2014. On the filtering properties of ensemble averaging for storm-scale precipitation forecasts. *Mon Wea Rev*, 142(3): 1093-1105
- Voisin N, Schaake J C, Lettenmaier D P. 2010. Calibration and downscaling methods for quantitative ensemble precipitation forecasts. *Wea Forecasting*, 25(6): 1603-1627
- Zhu Y J, Toth Z, Wobus R, et al. 2002. The economic value of ensemble-based weather forecasts. *Bull Amer Meteor Soc*, 83(1): 73-84
- Zhu Y J, Luo Y. 2015. Precipitation calibration based on the frequency-matching method. *Wea Forecasting*, 30(5): 1109-1124