

黄小燕,韦春霞,赵华生,等.地面-雷达-卫星资料的广西降水临近预报应用效果评估[J].气象研究与应用,2022,43(4):50-58.
Huang Xiaoyan, Wei Chunxia, Zhao Huasheng, et al. Evaluation of the application effect of Ground-Radar-Satellite data in Guangxi precipitation proximity forecast[J]. Journal of Meteorological Research and Application, 2022, 43(4): 50-58.

地面-雷达-卫星资料的广西降水临近预报应用效果评估

黄小燕^{1,2}, 韦春霞^{1*}, 赵华生¹, 何立¹, 吴玉霜^{3,4}

(1.广西壮族自治区气象科学研究所, 南宁 530022; 2.防城港暴雨实验室, 广西 防城港 538001;
3.广西壮族自治区气象台, 南宁 530022; 4.广西壮族自治区气象灾害防御技术中心, 南宁 530022)

摘要: 基于地面实况观测数据、雷达组合反射率、风云四号 A 星波段数据 3 种实况观测资料, 以 2018—2021 年广西 2850 个站点的小时累计雨量为预报对象, 采用随机森林算法建立未来 1—3h 的降水临近预报模型, 并分别进行了单类型观测资料预报因子、3 种类型观测资料多种组合预报因子输入的预报试验。对各预报试验结果综合采用 TS 评分、命中率、虚警率和漏报率进行点对点的评估表明, 地面资料在 1—3h 的小雨和中雨量级、1—2h 大雨量级的预报能力较好; 雷达资料对未来 1h 暴雨量级的预报较其他两种观测资料优势明显; 卫星资料在小雨量级上有一定的预报能力, 但其他量级各时效的预报均不理想。3 种观测资料在第 2—3h 的暴雨预报能力都偏低。3 类预报资料因子组合预报结果的评估表明, 各量级的多数预报时效均能比单类型资料预报取得更高的预报精度, 其中大雨量级的预报提高了 10% 以上, 暴雨则提高了 25% 以上。大雨和暴雨 1h TS 评分的空间分布情况表明, 组合因子高分区域分布均最广, 地面和雷达大雨量级的空间分布相当, 雷达的暴雨 TS 评分在 0.2 以上空间分布范围较地面和卫星资料广, 卫星的 TS 评分的空间分布显示其预报能力均最弱。

关键词: 地面实况观测数据; 雷达组合反射率; 风云四号 A 星波段数据; 降水临近预报; 评估

中图分类号: P457

文献标识码: A

doi: 10.19849/j.cnki.CN45-1356/P.2022.4.09

引言

数值模式预报技术和外推技术是目前业务上高分辨率定量降水短临预报的 2 种主要技术方法。因为数值模式在起初的数小时存在“模式起转”问题, 0~2h 模式临近降水预报效果不理想, 因此, 基于各种观测数据的外推是目前降水业务上临近预报所使用的最主要手段之一^[1-3]。

近年来, 随着观测技术的不断进步, 卫星和雷达等遥感技术也快速发展, 各种非常规数据呈爆炸式增长, 这些观测数据及其产品数据等形成了气象大数据^[4], 高效地利用数据成为了行业抢先发展的重要方向。目前气象业务中常使用的多源观测数据,

包括卫星、雷达以及自动气象站等观测数据, 对于降水天气的观测特性差异巨大。卫星云图中包含大量引发降水对流系统初生的有效信息, 雷达回波能够清晰地看到对流系统内部的降水粒子分布状况, 自动气象站能够提供降水系统的温压湿风等影响环境。单独使用各类观测数据, 均能一定程度上实现降水的有效预报, 如雷达回波外推、卫星云图亮温阈值判断等。目前的雷达降水外推预报研究发展较为成熟, 应用最广, 其中的单体质心法、交叉相关法和光流法是目前最常用的雷达回波外推算法^[5-10]。近年来, 随着深度学习技术的发展, 很多基于雷达资料的精细尺度降水临近预报试验也陆续开展^[11-13]。在利用卫星识别和预报潜在降水方面, 目前国内外也有

收稿日期: 2022-11-02

基金项目: 广西重点研发计划项目(桂科 AB22035016 和桂科 AB21196041)、广西气象局重点基金项目(桂气科 2021Z02)、国家自然科学基金项目(41765002)、广西人工智能预报技术创新团队项目

作者简介: 黄小燕(1978—), 女, 博士, 正研级高级工程师, 从事智能计算预报技术在天气预报中的研究应用。

E-mail: Gx_huangxy@163.com

* 通讯作者: 韦春霞(1968—), 女, 本科, 高级工程师, 从事气象应用研究工作。E-mail: wcx_hc@163.com

了较多相关研究,例如建立复杂算法,通过处理卫星图像数据对降水量进行估算^[14]及建立数学模型,分析验证采用卫星数据与降水过程之间的联系等^[15-18]。自动站的地面实况观测资料在分析强降水预报特征时运用较多^[19-20];也有学者将其运用于临近定量降水预报的研究,取得较好的预报效果^[21-22]。可见,采用地面、雷达和卫星等观测数据进行定量降水预报已经取得较多的研究进展,但对这些观测资料的预报效果进行评估对比的分析却较少涉及。

降水是复杂多变的物理变化过程,降水与气象要素之间存在非线性关系,另外,降水还具有显著的地域特征。目前,国内学者对广西典型喀斯特地貌的低纬度区域基于地面实况观测、卫星数据以及天气雷达数据的降水临近客观预报研究较少,而广西区域短时降水具有突发性、强度大、局地性明显等特点,监测和预报的难度比较大,对高质量、高时空分辨率的精细化、网格化的降水临近预报产品有强烈需求。随着广西气象观测网布控的日渐完善,观测数据剧增,资料量大,类型多样,哪一类型的观测资料对广西降水临近预报的贡献最大?不同观测资料在广西降水临近预报中的预报精度如何?融合多种资料的组合预报是否能取得更好的预报效果?都是需要研究解决的问题。本文在此尝试基于地面实况观测、雷达和风云四号卫星的多波段数据资料,采用非线性的随机森林算法^[23]建立广西区域降水临近客观预报模型,分析和评估3类观测数据的降水临近预报效果,为广西的降水临近预报业务提供一些思路。

1 资料和方法

1.1 资料

首先定义降水预报试验样本入选标准为:广西区域内地面自动气象观测站点小时累计雨量>10mm,同时出现的站点数 ≥ 100 个。为了增加预报样本数量,并且使降水与非降水的预报对象有更好的样本均衡性,该入选降水样本个例的前3h和后2h的雨量样本数据也同时选入预报试验样本中。2018年3月12日国家卫星气象中心开始下发风云卫星四号(FY4)资料,对应卫星资料开始时间,本文选择的预报试验时间为2018年3月12日—2021年12月31日,试验选用的观测类预报资料包括地面实况观测数据、雷达和卫星共3类:

(1)地面实况观测数据(以下简称地面):地面实况观测数据来源于广西壮族自治区气象局,为广西

气象站点的逐小时地面观测数据,时间段为2018年3月12日00时—2021年12月31日23时(北京时,下同)。根据数据的完整性以及与降水的相关性,选择的主要气象要素包括:过去3h变压(PRS_Change_3h)、过去24h的变压(PRS_Change_24h)、24h变温(TEM_Change_24h)、露点温度(DPE)、相对湿度(RHU)、过去1h降水量。

(2)雷达组合反射率(以下简称雷达):所用的雷达资料是中国气象局的强天气分析预报系统对雷达基数据进行采样处理,形成格点化的反射率因子等高平面拼图(CAPPI),覆盖范围为 $102^{\circ}\sim 114^{\circ}\text{E}$ 、 $19^{\circ}\sim 28^{\circ}\text{N}$,水平分辨率 $0.01^{\circ}\times 0.01^{\circ}$ (约 $1\text{km}\times 1\text{km}$),垂直分辨率 0.5km , $0.5\sim 10.5\text{km}$,共21层,6min输出一次,以21个层次的组合反射率作为广西临近降水预报模型的雷达预报因子。

(3)风云四号A星(FY-4A,以下简称卫星)波段数据:来源于风云卫星遥感数据服务网(<http://satellite.nsmc.org.cn/PortalSite/Default.aspx>)。相较于我国上一代静止气象卫星FY-2系列,FY-4A卫星无论是在其有效载荷上、探测性能上都大幅度领先。另外,对比国际上同类静止气象卫星,目前FY-4A卫星所装载的多通道辐射成像仪,其14个成像通道都与国际领先水平相当,并且首次实现了静止轨道红外高光谱探测,能够获取到大气温湿度的三维结构,在国际上处于领先地位。本文所研究使用的主要是FY-4A的静止轨道辐射成像仪AGRI的L1级数据,选用其中第7~14共8个通道的数据,其进行地球全圆盘扫描的时间间隔为15min,进行区域扫描的时间间隔为1min。为充分利用原始数据精度且便于分析计算,预处理为研究区域间隔为 0.04° (约4km)的等经纬度网格数据。

1.2 方法

本文主要着重于客观比较和分析地面、雷达和卫星3种观测资料在广西临近降水中的预报效果,在建立未来1~3h预报时效的广西临近降水预报模型时,针对本预报试验大样本数据量的情况,同时为了更好地提取各类不同观测数据的预报信息以及基于计算效率方面的考虑,采用了非线性的随机森林算法(RF)来进行建模预报。这里对这种算法构建的广西降水临近预报模型和计算处理步骤进行简单介绍。

RF算法是集成学习方法的一种,它通过先产生一组CART决策树学习器,对于本文所研究的回归问题,采取将每一颗决策树的预测结果进行累加后

取均值的策略进行集成。RF 回归模型的计算速度较快并且它能够根据模型训练的过程给出样本中的特征因子的重要程度,可以为接下来的特征选择提供一个参考依据。其回归问题,采用的是最小均方差原则,对于任意划分特征 A 的任意划分点 S,将数据集划分为 S_1 和 S_2 ,要计算出使 S_1 和 S_2 各自的均方差最小且二者的均方差和最小特征以及特征值的划分点

$$\min[\min\sum(y_i-c_1)^2+\sum(y_i-c_2)^2] \quad (1)$$

其中, c_1 和 c_2 分别为 S_1 和 S_2 的样本输出均值, y_i 为输入样本。

采用 RF 算法的回归原理建立广西降水临近预报模型,将所选数据分为建模样本数据集、独立样本数据集,其中建模样本数据集为有效的模型输入数据集,独立样本数据集作为对模型预测结果的检验数据。

本文降水临近试验预报的计算和处理步骤如下:

(1)从训练数据中选择合适时间段的特征变量采用随机森林算法进行模型训练,分别形成未来 1h、2h 和 3h 预报时效的降水预报模型。

(2)将测试输入数据集输入至已训练的预报模型中,输出模型各时效的降水预报数据。

(3)对模型输出的各时效降水预报结果进行检验,评估预报效果。

2 地面-雷达-卫星资料的降水短时临近预报试验

2.1 资料预处理

(1)地面资料:根据资料的完整性,本文所选取的预报对象为广西 2850 个气象观测站点未来 1~3h 的累计雨量数据(图 1)。由于广西气象观测站点不

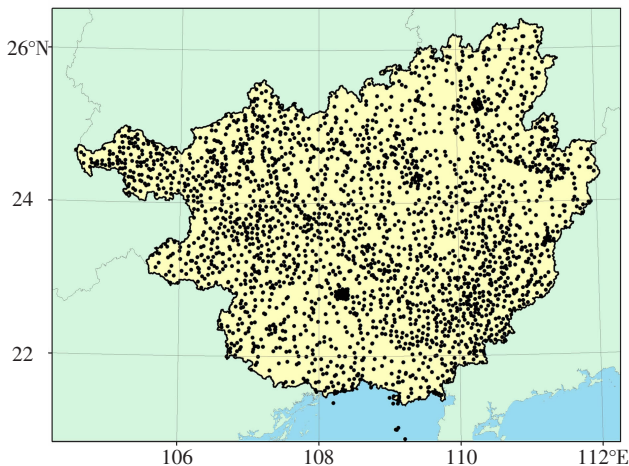


图 1 预报试验的广西 2850 个气象观测站点

断在增加,站点数量在本文试验统计时段内也是不相等的,为了更好地建立预报模型,需要对各要素离散的站点观测数据进行格点化后再站点化,统一到与雨量站点一致的站点上。具体处理为:针对每个时间样本的 6 个实况观测数据先进行 5km×5km 分辨率的格点转化,得到整个广西区域内所有格点的实况观测数据资料场,进一步对该格点数据场进行 2850 个气象站点的插值转化,处理后的数据在时间和空间上均一致。

(2)雷达:基于广西区域的雷达组合反射率资料,采用三次多项式的插值方法,把 1km×1km 分辨率的雷达资料插值到 2850 个气象站点上。

(3)卫星:利用国家气象卫星中心下发的 SWAP 处理平台进行卫星资料的投影和对标等预处理工作,并转存为 IMG 格式,进一步编程实现各通道数据的独立存储。同样是采用三次多项式的插值方法,把 4km×4km 分辨率的卫星资料插值到 2850 个气象站点上。

2.2 因子构建

(1)地面预报因子:起报时间点和前 1h 的观测要素预报因子,共 12 个实况观测预报因子。

(2)雷达预报因子:起报时间点前移 0.5h 内的 5 帧雷达回波图,各帧雷达数据图之间的差值,即第 5 帧图减去第 1 帧图,第 5 帧图减去第 2 帧图……以此类推,这样的处理方式,可以同时考虑到雷达的实况监测信息以及其随时间变化的信息,共得到 15 个雷达预报因子。

(3)卫星预报因子:选用起报时间点前移 1h 内的 4 帧云图数据,包括正点的 00 分、15 分、30 分、45 分共 4 个时间点上的数据,为了进行全天的预报,选择 7~14 通道数据,包括各通道数据的计算组合等,共获得 72 个卫星预报因子。

2.3 随机森林算法的降水临近预报试验

基于上述构建的 3 类观测资料的预报因子,采用 1.2 节介绍的随机森林算法,分别建立单类型观测资料因子、组合观测资料因子共计 7 组广西未来 1~3h 的降水临近预报模型。在模型的样本架构上,采用了叠加所有站点时间序列样本的方式进行,共计样本数量为 1855350 个。选取全部样本数据的 60% 共 1140000 个数据作为建模样本,进行模型的训练,其余 715350 个样本数据,作为独立样本,用于验证各类型观测资料所建模型的预报效果。

2.4 预报结果分析

2.4.1 短时降水分级及精度评定方法

为了客观检验地面、雷达和卫星资料未来 1h、2h 和 3h 广西降水短时临近预报能力,本文共分为 4 个雨量等级进行预报效果的评定:

- ①小雨: $0.1\text{mm} \leq R \leq 2.9\text{mm}$;
- ②中雨: $3.0\text{mm} \leq R \leq 9.9\text{mm}$;
- ③大雨: $10.0\text{mm} \leq R \leq 19.9\text{mm}$;
- ④暴雨和大暴雨: $20.0\text{mm} \leq R$ 。

其中, R 为雨量

精度评价指标主要采用真实技巧评分 TS 评分,同时结合命中率(POD)、虚警率(FAR)以及漏报率(PO)等指标对各级降水区间进行各站点的分级评估。4 类评估指标定义如下:

$$TS=NA/(NA+NB+NC) \tag{2}$$

$$POD=NA/(NA+NC) \tag{3}$$

$$FAR=NB/(NA+NB) \tag{4}$$

$$PO=NC/(NA+NC) \tag{5}$$

式中、和分别为正确、空报、漏报站(次)数,具体意义见表 1。

表 1 短时降水预报检验分类表

		预报	
		有	无
实况	有	N_A	N_C
	无	N_B	-

2.4.2 预报结果分析

2.4.2.1 单类型观测资料因子的预报

首先分别对地面、雷达和卫星进行单类型预报因子输入的降水临近建模预报试验。TS 评分试验结果表明(图 2),小雨量级:第 1h 最高为地面,TS 评分为 0.1541;卫星次之,雷达最差;第 2h,卫星最高,TS 评分为 0.1290;地面次之,雷达表现最差;第 3h,地面最高,TS 评分为 0.1482;卫星次之,雷达依旧表现最差。中雨量级:1~3h 均是地面最高,TS 评分分别为:0.1599、0.1114 和 0.0914;雷达次之,卫星最差。大雨级别:第 1h,地面和雷达均表现较好,TS 评分分别为:0.1048 和 0.1010;卫星较差。第 2h,3 种观测数据的预报精度均下降明显,最优地面的 TS 评分也仅有 0.0477;卫星和雷达评分相当;第 3h,3 种观测数据的预报精度进一步下降,表现最优的卫星,其 TS 评分也仅为 0.029;雷达下降最明显,TS 评分

只有 0.0169。暴雨级别:第 1h,雷达评分最高,TS 评分为 0.0930;地面实况次之,卫星在暴雨级别上的预报能力偏低;第 2h 和第 3h,3 种观测资料的预报均下滑明显,TS 评分均在 0.003 以下,预报精度很低,基本没有预报能力。分析还发现,各类型资料小雨量级预报的第 3h 均比第 2h 更高,似乎小降水级别的预报精度并不随预报时效的增加而有所下降,其原因值得进一步深入研究分析。

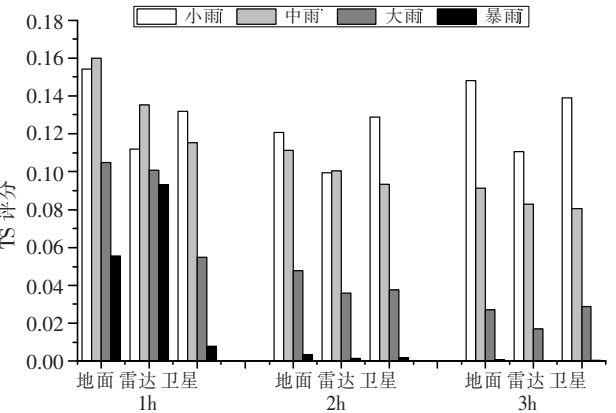


图 2 3 种观测资料因子降水临近预报的 TS 评分

命中率先分析发现(图 3),3 类观测资料均是中雨量级的命中率最高,大部分在 60%以上,最高命中率为雷达预报未来的第 3h,达到了 72.7%;小雨量级的命中率次之,各观测资料预报均在 30%~40%之间;进一步分析发现,小雨和中雨两个级别的命中率并未随着预报时效的增加而下降。大雨以上量级的命中率均偏低,分析 1h 的预报发现,地面和雷达在大雨量级上表现出比卫星高一倍的命中率;3 类观测资料在第 2h 的命中率均明显下降,特别是地面实况和雷达,下降至与卫星的命中率几乎一致;第 3h 中,地面实况和雷达进一步下降明显,3 种资料的命

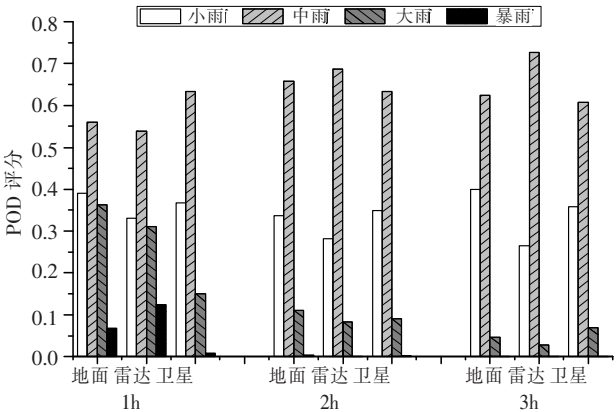


图 3 3 种观测资料因子降水临近预报的 POD 评分

中率都极低。暴雨量级上,雷达在第 1h 的命中率明显比地面和卫星资料的预报高一倍以上,为 12.5%,可以发现,3 类观测资料对于暴雨级别的命中率在第 2 和第 3h 基本都很低。

总体分析虚警率可知(图 4),除了卫星第 3 小时的暴雨预报,其他时效每类资料大雨量级的虚警率均最高。虚警率最低为雷达的 1h 暴雨预报,为 73.3%;地面次之,为 76.5%,卫星虚警率超过 80%。地面在第 2 小时的暴雨虚警率最低,为 81.5%,雷达和卫星第 2 小时的暴雨预报虚警率相对第 1 小时增长迅速,均超过 90%。值得关注的是,3 类数据资料

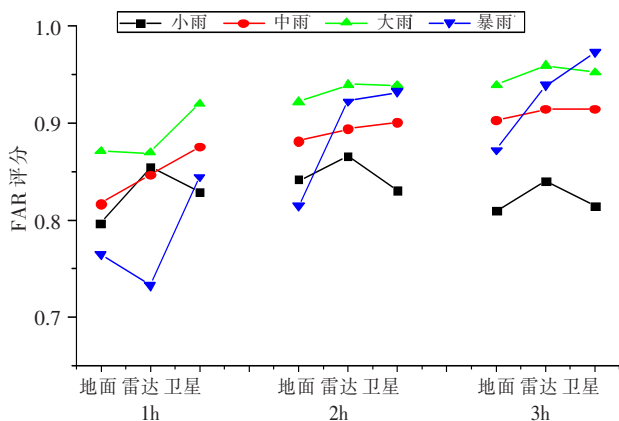


图 4 3 种观测资料因子降水临近预报的 FAR 评分

2.4.2.2 组合观测资料因子的预报

3 种类型观测资料在不同的量级和时效上所表现的预报能力各有优势和不足,若是把多种观测资料融合在一起进行预报计算,是否能取得更好的预报效果?为了验证这 3 类资料的组合预报效果,本文进一步进行分别建立各类资料的组合因子预报试验,包括地面预报因子+雷达预报因子、地面预报因子+卫星预报因子、雷达预报因子+卫星预报因子、地面预报因子+雷达预报因子+卫星预报因子,共 4 种情况。

限于篇幅关系,这里仅给出各类资料组合因子 TS 评分的预报情况,见表 2。分析可知,地面+雷达资料的组合预报因子在第 1 小时的前 3 个量级上均取得比其他资料组合更好的预报效果;第 2 小时,地面+卫星资料在小雨量级上表现最优,地面+雷达资料则是中雨量级预报效果最好;在更长的预报时效上(第 3 小时),3 类资料组合在一起的预报效果更优。进一步分析也可以发现,3 类资料组合在大雨以上量级的预报效果优势更明显。

的小雨量级虚警率上,第 3 小时均比前 2 小时有所降低,这与小雨量级在 TS 评分上的表现结果较为一致。

漏报率上(图 5),3 类资料在暴雨级别上各预报时效均最高,大雨级别次之,中雨的漏报率最小,这与中雨的命中率最高的统计结果较为一致。具体分析可知,第 1 小时,雷达的暴雨漏报率最低,为 66.9%;第 2 和第 3 小时,漏报率与地面和卫星数据基本一致,预报能力很弱。大雨级别上,第 1 小时,地面漏报率最低,为 63.7%;雷达次之,卫星最高。第 2 和第 3 小时,3 类资料的漏报率都偏高。

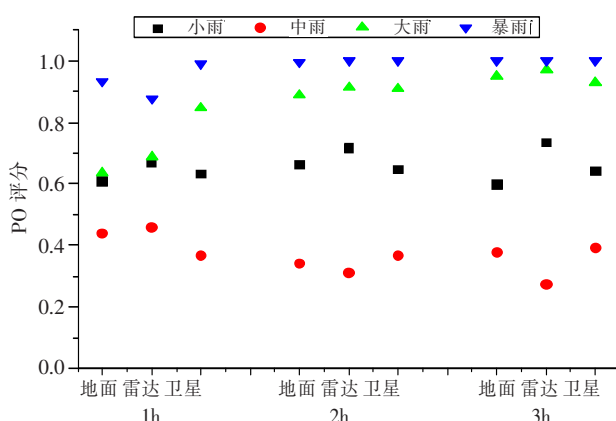


图 5 3 种观测资料因子降水临近预报的 PO 评分

进一步与前述单类型观测资料预报因子的 TS 最优预报进行对比分析可知(表 3),融合多种类型观测数据预报因子的预报 TS 评分,除了小雨量级的第 2 小时和第 3 小时没有提高,其余各量级各预报时效在进行多资料信息的因子组合后能取得更好的预报精度,特别是大雨以上量级的强降水级别提高更明显,大雨的 3 个预报时效都提高了 10%以上,暴雨提高最明显,达到 25%以上。这说明大雨以上量级的强降水,随着预报时效的增加,需要加入更多的预报信息才能取得更好的预报效果。

2.4.2.3 TS 评分的空间分布

3 种资料以及组合因子在各雨量等级上均表现为不同的预报能力,那么其是否有地域空间分布方面预报能力的差别?限于篇幅关系,我们在此仅给出 1h 大雨和暴雨量级 TS 评分的空间分布情况。

大雨量级上(图 6),3 种资料组合因子(图 6a)的 TS 评分在 0.1 以上的空间分布范围最广;其次为地面和雷达,空间分布范围相当(图 6b、6c);卫星能达到 0.1 以上评分的地域非常少。进一步分析发现,

表 2 组合观测资料预报因子降水临近预报的 TS 评分

输入资料	预报时效	小雨	中雨	大雨	暴雨
地面+雷达	1	0.1736	0.1679	0.1182	0.1170
	2	0.1182	0.1139	0.0523	0.0049
	3	0.1424	0.0930	0.0363	0.0021
地面+卫星	1	0.1625	0.1581	0.1044	0.0624
	2	0.1234	0.1110	0.0536	0.0045
	3	0.1462	0.0916	0.0361	0.0013
雷达+卫星	1	0.1310	0.1389	0.1009	0.1076
	2	0.1158	0.1037	0.0471	0.0035
	3	0.1329	0.0855	0.0307	0.0004
地面+雷达+卫星	1	0.1704	0.1649	0.1171	0.1251
	2	0.1221	0.1108	0.0543	0.0054
	3	0.1425	0.0931	0.0381	0.0032

注:表中加粗数字表示该预报时效上各量级的最优预报结果

表 3 组合因子各量级上最优 TS 与单个资料预报因子最优 TS 的比较 (%)

预报时效	小雨	中雨	大雨	暴雨
1	12.7	5.0	12.8	25.8
2	-4.3	2.2	13.8	58.8
3	-3.8	1.9	31.4	357.1

3 种资料组合因子较高的 TS 评分主要分布在永福和鹿寨,武宣和桂平一带,浦北、横县以及武鸣等地,

桂西的局部地区评分也较高。地面资料在钦州,桂林的龙胜、全州、灌阳一带,乐业等地表现较好(图 6b)。雷达资料在桂西的隆林和西林局部,崇左,南宁和武鸣一带的 TS 评分较高(图 6c)。卫星资料则在崇左南部,玉林的北流等地有较高的预报精度,但总体空间分布的表现上,预报能力还是偏弱(图 6d)。

暴雨量级的 TS 评分空间分布和大雨量级的差异较大,表现为评分 0.1~0.2 的范围较少,而评分在 0.2 以上,甚至 0.25 以上较高评分的空间分布较广。

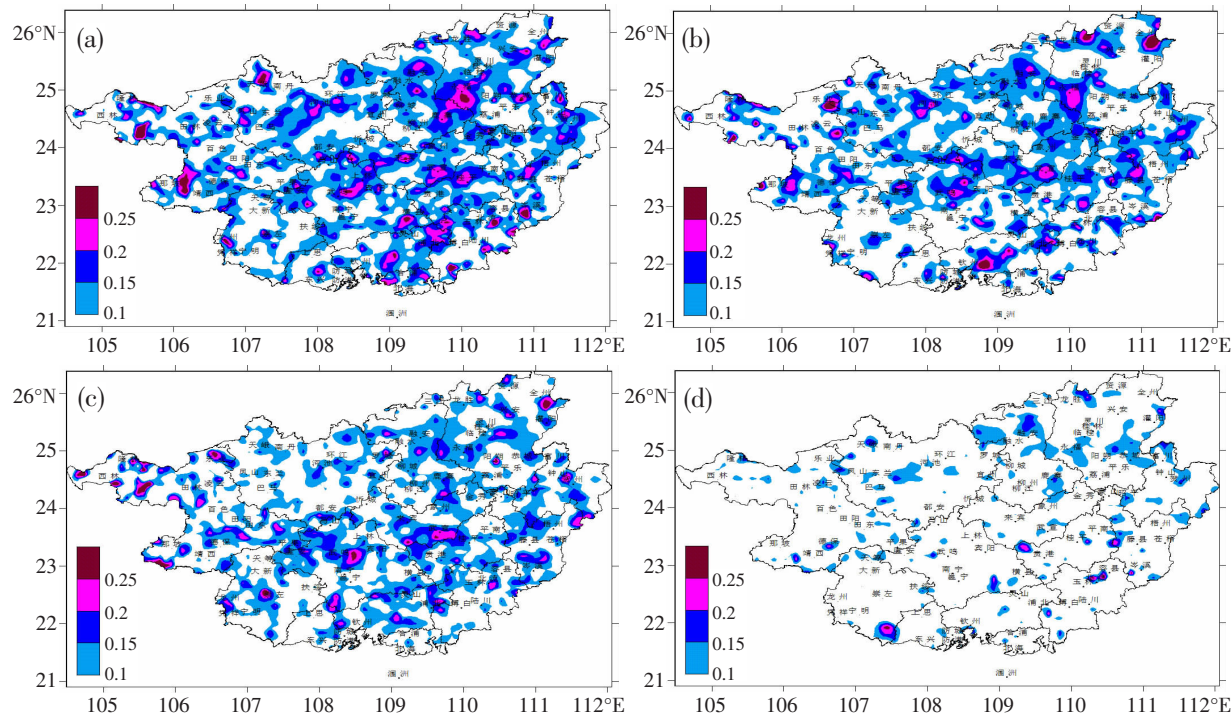


图 6 大雨量级各资料未来 1hTS 评分空间分布
(a)3 类资料组合;(b)地面;(c)雷达;(d)卫星

其中,3种资料组合因子(图7a)的TS评分在0.2以上的空间分布范围最广。其预报评分较高的地域主要为:平果-天等-隆安-扶绥一带,桂林永福和资源,钦州、防城和合浦等沿海地区。预报能力偏弱的地区主要分布在桂东南玉林的大部分地区、桂林的全州和灌阳、崇左和上思一带。雷达的暴雨TS评分在0.2以上空间分布范围较地面和卫星资料广(图

7b),其评分高的地域分布与组合因子多处较为一致,预报表现最好为桂林永福、临桂和资源,钦州、防城和合浦等沿海地区,但是平果-天等-隆安-扶绥一带表现不佳。地面资料在巴马、东兰地区评分较高(图7c),卫星的暴雨预报能力,空间分布显示很弱(图7d)。

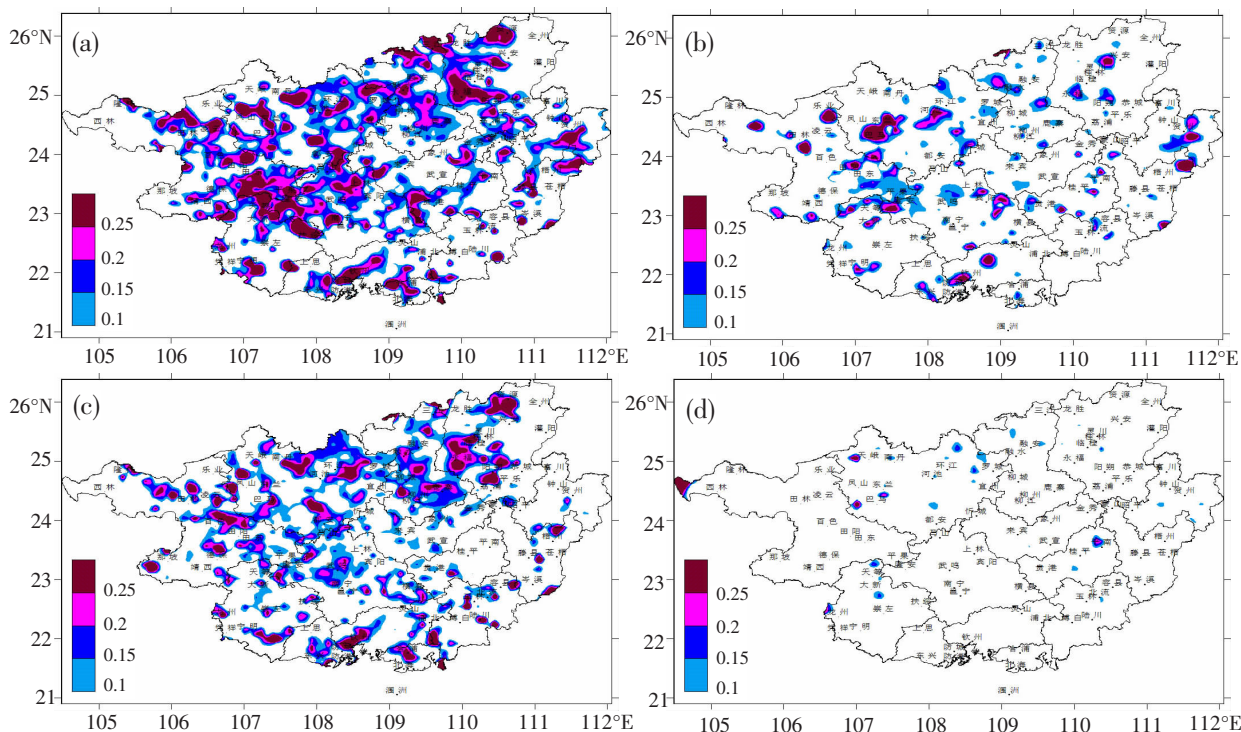


图7 暴雨量级各资料未来1hTS评分空间分布

(a)3类资料组合;(b)地面;(c)雷达;(d)卫星

3 结论和讨论

针对广西未来1—3h累计雨量,采用随机森林算法建立客观降水预报模型,分析和验证地面、雷达和风云四号静止气象卫星3类观测资料的广西降水临近预报效果,预报试验结果表明:

(1)TS评分上,地面资料对小雨和中雨2个量级3个时效雨量预报的TS评分都较高,第1和第2小时大雨量级的预报评分也较高;雷达资料对未来1h暴雨量级的预报TS评分优势较为明显,但是第2和第3小时的预报与其他两种资料一样,TS评分下降明显,预报能力低;卫星资料除了在小雨量级上预报效果与其他两种观测资料相当外,其他量级各时效的预报均不理想。

(2)3类观测资料均是中雨量级的命中率最高,

大部分在60%以上,小雨量级的命中率次之,各观测资料预报均在30%~40%之间;大雨以上量级的命中率均偏低,其中,地面和雷达1h大雨量级的命中率比卫星高一倍;而雷达在第1小时的暴雨命中率明显比地面和卫星资料预报高一倍以上。3种观测资料大雨量级的虚警率均较高,而暴雨的第1小时虚警率在所有时效和量级中均最低,表现最好。3种资料的暴雨级别各预报时效漏报率均最高,大雨级别次之,中雨的漏报率最小,与中雨的命中率最高的统计结果较为一致。

(3)组合观测资料因子的预报结果表明,在进行3类预报资料的因子组合后各量级的多数预报时效均能取得更好的预报精度,特别是大雨以上的量级,融合多种类型观测数据预报因子的预报精度提高更为明显,其中暴雨提高是最明显的,说明大雨以上量

级的降水,随着预报时效的增加,需要加入更多的预报信息才能取得更好的预报效果。

值得注意的是,本试验的检验方法是点对点进行的,这对雷达数据和卫星资料的检验效果会有所影响,未来可用空间诊断检验进行评估分析。另外,可以看到,本文试验中的卫星资料预报因子有 72 个,是 3 类资料中最多的,但是预报效果却相对不理想,其预报信息需要进一步深入挖掘才能更好进行预报应用。

本文目前只采用了 3 类观测数据进行预报试验,虽然各资料都有一定的预报能力,但是预报性能与业务需求之间依然有较大差距,特别是第 2 小时以后的预报,效果都比较差,因为对流系统通常会发生较大的变化,有的对流系统会消散,也会有新的对流系统生成,利用观测数据进行直接建模计算的方法,往往不能对上述变化做出较好的判断。目前的高分辨率数值天气预报使用了多种同化手段,已经显著提升了预报能力,将观测数据和高分辨率数值天气预报进行融合将成为一个有效的解决办法。如何实现多源观测数据和模式产品的有效综合应用,提取各类数据中的有效预报信息,进而实现强降水等灾害天气的更有效预报,是我们下一步需要深入研究解决的问题。

参考文献:

- [1] 俞小鼎,周小刚,王秀明. 雷暴与强对流临近天气预报技术进展[J]. 气象学报,2012,70(3):311-337.
- [2] 范水勇,王洪利,陈敏,等. 雷达反射率资料的三维变分同化研究[J]. 气象学报,2013,71(3):527-537.
- [3] 郑淋淋,邱学兴. 一种改进的降水临近外推预报技术方法研究及效果检验[J]. 气象科技,2020,48(1):97-106.
- [4] 沈文海. 再析气象大数据及其应用[J]. 中国信息化,2016(1):84-96.
- [5] 韩雷,王洪庆,林隐静. 光流法在强对流天气临近预报中的应用[J]. 北京大学学报(自然科学版),2008,44(5):751-755.
- [6] 韩丰,魏鸣,李南,等. 反射率因子和径向风速共同约束反演多普勒雷达风场化 [J]. 遥感学报,2013,17(3):584-589.
- [7] 韩丰,龙明盛,李月安,等. 循环神经网络在雷达临近预报中的应用[J]. 应用气象学报,2019,30(1):61-69.

- [8] 曹春燕,陈元昭,刘东华,等. 光流法及其在临近预报中的应用[J]. 气象学报,2015,73(3):471-480.
- [9] Bechini R, Chandrasekar V. An enhanced optical flow technique for radar nowcasting of precipitation and winds [J]. Journal of Atmospheric and Ocean Technology, 2017, 34(12):2637-2658.
- [10] Chen Y Z, Lan H P, Chen X L, et al. A nowcasting technique based on application of the particle filter blending algorithm [J]. Journal of Meteorological Research, 2017, 31(5):931-945.
- [11] 梁振清,陈生. 基于深度学习和雷达观测的华南短临预报精度评估[J]. 气象研究与应用, 2020, 41(1):41-47.
- [12] 张佳洛,黄勇,刘传才. 基于卷积门循环单元和气象雷达图像的临近降水预报[J]. 计算机与数字工程, 2021, 49(8):1538-1542.
- [13] 曹伟华,南刚强,陈明轩,等. 基于深度学习的京津冀地区精细尺度降水临近预报研究[J]. 气象学报, 2022, 80(4):546-564.
- [14] 师春香,谢正辉. 基于静止气象卫星观测的降水时间降尺度研究[J]. 地理科学进展, 2008, 27(4):15-22.
- [15] 魏建苏,严明良,樊永富,等. 卫星云图和数值产品结合的汛期强降水预警系统[J]. 气象科学, 2001, 21(3):355-362.
- [16] John R, Mecikalsk, Kristopher M, et al. Forecasting convective initiation by monitoring the evolution of moving cumulus in daytime GOES imagery [J]. MECIKALSKI AND BEDKA, 2006(1):49-77.
- [17] 卢乃锰,吴蓉璋. 强对流降水云团的云图特征分析[J]. 应用气象学报, 1997(3):14-20.
- [18] 李嘉睿,卢乃锰,谷松岩. 青藏高原地区 TRMM PR 地面降雨率的修正[J]. 应用气象学报, 2015, (5):636-640.
- [19] 罗艳艳,吴雪贞,郑宝智,等. 基于区域自动站资料的闽中地区前汛期短时强降水特征分析[J]. 海峡科学, 2016(9):7-10.
- [20] 张功文,董方亮,谢详永. 基于区域自动站资料的邯郸市夏季短时强降水分布研究[J]. 现代农业科技, 2016(14):205-207, 220.
- [21] 石宏辉,顾欣,杨通荣,等. 利用区域自动站资料作短时临近预报预警[J]. 贵州气象, 2010, 34(1):29-30.
- [22] 丁立善,张长栋. 利用区域自动站资料作短时临近预报预警[J]. 农业与技术, 2018, 38(8):230.
- [23] Breiman L. Random forests [J]. Machine Learning, 2001, 45(1):5-32.

Evaluation of the application effect of Ground–Radar–Satellite data in Guangxi precipitation proximity forecast

Huang Xiaoyan^{1,2}, Wei Chunxia^{1*}, Zhao Huasheng¹, He Li¹, Wu Yushuang^{3,4}

(1. Guangxi Institute of Meteorological Sciences, Nanning 530022, China; 2. Fangchenggang Rainstorm Laboratory, Guangxi Fangchenggang 538001, China; 3. Guangxi Meteorological Observatory, Nanning 530022, China; 4. Guangxi Meteorological Disaster Prevention Technology Center, Nanning 530022, China)

Abstract: Based on three kinds of ground observation data, radar combined reflectivity and band data of FY4A satellite from the live observation data, taking the hourly accumulated rainfall of 2850 stations in Guangxi from 2018 to 2021 as the prediction object, the random forest algorithm is used to establish the rainfall approaching prediction model for the next 1–3 hours, and the prediction tests of single type observation data prediction factor and three types of observation data multiple combination prediction factor input are conducted respectively. The results of each prediction test were evaluated by TS score, hit rate, false alarm rate and false alarm rate, the ground data has a good prediction ability in light rain and moderate rain in the next 1–3 hours and heavy rain in the next 1–2 hours; the radar data has obvious advantages over the other two kinds of observation data in forecasting the rainstorm magnitude in the next hour; the satellite data has a certain forecast ability in the light rain magnitude, but the predictions of other magnitudes and time effects are not ideal. The rainstorm forecast capacity of all three observations was low at the 2nd–3rd hour. The evaluation of the combined forecast results of three types of forecast data factors shows that the prediction accuracy of most forecasts of all magnitudes is higher than that of single-type data. The prediction of heavy rain increases by more than 10%, and that of heavy rain increases by more than 25%. The spatial distribution of 1-hour TS scores of heavy rain and rainstorm shows that the regions with high combined factor scores are the most widely distributed, and the spatial distribution of ground and radar heavy rain magnitude is similar. The spatial distribution range of radar heavy rain TS scores above 0.2 is wider than that of ground and satellite data, and the spatial distribution of satellite TS scores shows that their forecasting ability is the weakest.

Key words: Ground observation data, radar combined reflectivity, band data of FY4A satellite, precipitation approaching forecast, evaluate