

基于雷达组网拼图和 XGBoost 的雷达定量降水估测

赵华生¹ 李晓静^{2*}

(1 广西气象科学研究所, 南宁 530006; 2 广西幼儿师范高等专科学校, 南宁 530006)

摘要 针对传统方法采用天气雷达进行强降水的定量估测存在较大偏差问题, 论文以 1 h 累计雨量为估测对象, 基于雷达组网拼图资料, 采用 XGBoost(eXtreme Gradient Boosting)算法, 建立新的雷达估测降水模型。该模型设计以前 1 h 的雷达组合反射率因子作为输入, 进一步采用若干个剔除异常样本的策略有效清除建模样本中的部分噪声, 更好地构建了雷达组合反射率与估测对象之间的非线性映射关系。在 32 万个独立检验样本的估测结果中, 其均方根误差(RMSE)为 6.04 mm、平均绝对误差(MAE)为 3.50 mm、预报偏差(BIAS)为 1.05; 与目前业务系统上使用的 ZR(300, 1.4)关系方法相比, 前者的 RMSE 和 MAE 分别下降了 20.6% 和 10.3%, 而 BIAS 指标则显示后者对降水量级的估测明显低估。进一步对小时雨强大于 10 mm 样本的统计结果表明, 新方案的 RMSE、MAE 以及 TS 评分均大幅优于 ZR(300, 1.4)关系方法, 可进行实际业务应用。

关键词 雷达降水估测; XGBoost; 组合反射率; 雷达拼图

中图分类号: P415.2 **DOI:** 10.19517/j.1671-6345.20220301 **文献标识码:** A

引言

中国是世界上暴雨洪涝等强降水发生最严重的国家之一, 平均每年因此造成的经济损失高达几百亿元^[1], 精准监测和预测强降水落区对于解决该问题非常关键。虽然目前国内地面自动站数量在逐年增加, 但仍无法满足网格预报业务对精细化程度的需求。因此, 在短时临近的气象业务中, 基于雷达等遥感观测手段的定量降水估测和预报方法在网格降水量监测和临近预报中扮演着愈发重要的角色。建立一个精准的雷达定量降水估测(Quantitative Precipitation Estimation, QPE)方法, 运用在雷达外推的短时临近预报业务中, 就可以更好的进行短时临近降水预测, 从而成为防灾减灾的重要手段。

近年来, 利用新一代多普勒雷达进行定量降水估测成为雷达资料应用的主要研究方向之一^[2]。大量研究表明, 天气雷达的观测资料与地面的实况观测雨量的有效结合, 有助于提高雷达的降水估测精

度^[3]。雷达反射率因子(Z)与降水强度(R)的关系通常用 $Z=aR^b$ 表示, 其中参数 a 和 b 可以通过雨量和雷达的观测资料进行确定^[4]。目前美国和中国业务系统均采用了 $a=300$ 、 $b=1.4$ 作为雷达估测降水的默认参数。然而大量的研究表明, 参数 a 、 b 是与季节、降水条件以及地域等因素有关的时空函数^[5-6], 对雷达定量降水估测的准确性有很大影响, 单一的 $Z-R$ 关系估测降水方法往往会造成大雨估测偏小而小雨估测偏大的问题^[7]。针对不同地区 $Z-R$ 关系的系数确定问题, 国内外学者开展大量关于 $Z-R$ 关系降水估测模型的研究: 郑媛媛等^[8]针对不同距离和不同仰角的回波, 利用最优化方法对不同高度回波的 $Z-R$ 关系, 提出了不同的降水估测模型, 提高了雷达估测降水的精度, 然而这种细分区域和降水类型的模型难以推广使用^[5]; 高峰等^[9]采用 2 h 滑动时间窗口的方法对 $Z-R$ 关系进行实时订正, 该方法对中雨的估测具有较好的效果。结合地面雨量观测值和天气雷达观测进行降水估测的方

<http://www.qxkj.net.cn> 气象科技

广西高校中青年教师科研基础能力提升项目(2022KY0963)、广西科技厅重点研发计划项目(桂科 AB21196041)、广西职业教育教学改革研究重点项目(GXGZJG2022A001)资助

作者简介: 赵华生, 男, 1980 年生, 硕士, 高级工程师, 主要从事非线性气象统计预报, Email: 2006zhaohuasheng@163.com

收稿日期: 2022 年 7 月 18 日; 定稿日期: 2023 年 2 月 14 日

* 通信作者, Email: 50921626@qq.com

法,主要思路是通过前几个小时的雨量站降水资料来调整 $Z-R$ 模型中的 a 、 b 参数^[10,11],该方法可以有效提高雷达降水估测的精度,但雷达与地面观测雨量的数据质量会带来一些误差^[12],特别是在反射率因子 Z 值较大的情况下, Z 的随机误差造成的较小变化会对降水强度 R 造成较大的影响。同时,动态调整 $Z-R$ 关系要求自动站降水观测数据与雷达观测同步,对降水观测的频次和数据传输的及时性要求较高。因此,基于 $Z-R$ 关系的降水估测方法难以解决强降水估测误差较大的问题^[13-20]。

随着机器学习方法的不断发展和应用,基于其较强的模型表达和拟合能力,许多学者将机器学习模型应用于雷达的定量降水估测。Xia 等^[21]最早采用人工神经网络用于雷达降水估测,通过将雷达观测插值到不同的高度层,输入不同高度层的雷达反射率因子来估测降水,结果优于 $Z-R$ 等传统方法;Ju-Young Shin 等^[22]利用随机森林、随机梯度增强模型和极端学习机方法对 Gwangdeoksan 雷达站极化雷达资料进行了定量降水估算模型的适用性研究。结果表明基于机器学习的定量降水估计模型比传统的 $Z-R$ 关系模型具有更好的性能,特别是对于强降水事件;Zhang Changjiang 等^[23]利用平稳小波变换与支持向量机(SVM)相结合,进行雷达的定量降水估测,相比于 $Z-R$ 关系,该方法的降水估测精度有一定的提高;傅德胜等^[24]将采用 RBF 神经网络与变分法得到的 $Z-R$ 关系进行对比,发现可以得到比 $Z-R$ 关系更好的降水估测模型;殷志远等^[25]采用遗传神经网络进行了雷达定量估测和预报降水的实验,整体上提高了洪水的预报精度。但是,目前尚未见有利用 XGBoost 方法来构建雷达定量降水估测的研究报道。XGBoost 方法是 GBDT(梯度提升迭代决策树)的改进,具有强的非线性拟合能力、防止过拟合、算法效率高、泛化能力强等特点^[26-27]。本文尝试采用 XGBoost 方法构建雷达定量降水估测模型,并进行定量强降水估测研究。

1 数据和处理方法

1.1 雷达资料

本文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(Constant Altitude Plan Position Indicating, CAPPI),覆盖范围为 102°

$\sim 114^\circ\text{E}$ 、 $19^\circ\sim 28^\circ\text{N}$,水平分辨率 $0.01^\circ\times 0.01^\circ$,垂直分辨率 0.5 km , $0.5\sim 10.5\text{ km}$,共 21 层,6 min 输出一次。论文以 21 个层次的组合反射率作为定量降水估测的雷达资料,其中数据采集的时段为 2016—2020 年 6 月汛期(每年的 4—10 月),而非汛期(每年的 1—3 月和 11—12 月)雷达关机时段及故障导致数据连续缺失时段不予统计。

1.2 地面实况雨量观测资料

地面实况雨量观测资料来源于广西气象局,包括广西区域约 2800 个自动站的 1 h 雨量观测数据,统计时段与雷达资料的时段相对应。图 1 为本文研究的雨量观测点的地理位置和分布情况。

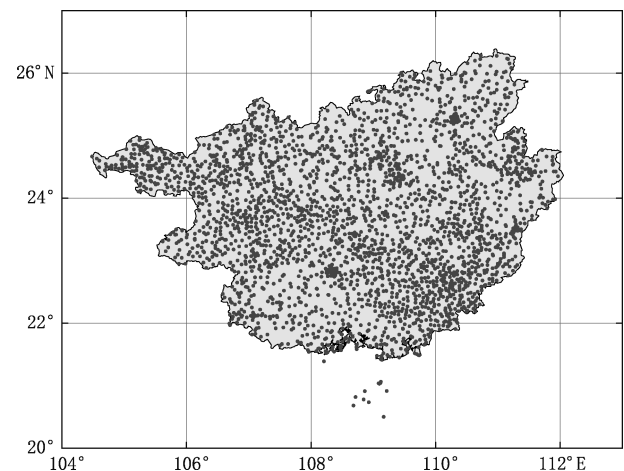


图 1 研究区域及其地面站点分布

1.3 数据分析及处理

针对一年中出现降水事件的概率比较小,出现大范围的强降水事件概率更小,同时,大多数 1 h 内的降水事件均发生在局部地区等情况。因此,为了使模型的训练更具有针对性,论文在训练样本时进行了如下处理:

(1)对 2016 年 4 月至 2019 年 6 月汛期内的整点时段,选取全区自动站点中,有超过 10 个(约 2850 个站点中的 10 个以上)以上站点的累计降水量超过 10 mm 的整点时段,共计 13402 个 1 h 时段。

(2)采用三次多项式插值法,将 CAPPI 组合反射率的网格数据插值到自动站的位置,代表该自动站的反射率因子,从而得到 CAPPI 组合反射率因子与地面自动站雨量观测之间的一一对应关系。

(3)计算该自动站在过去 1 h 内的平均 CAPPI

组合反射率因子 $D = \frac{1}{10} \sum_{j=0}^9 d_j$ (雷达反射率因子是 6 min 一帧, t 时刻其过去 1 h 内反射率因子为: $d(t), d(t-1), \dots, d(t-9)$)。

(4)构建样本矩阵:首先,论文以过去 1 h 的累计降水量作为预报对象,因此,对于(1)中选取的每个 1 h 时段所对应的 t 时($t=1, 2, \dots, 13402$),对于站点 s ($s=1, 2, \dots, 2850$),如果其过去 1 h D_s 大于或等于 20 dBz,则时段 t 选取的样本矩阵为: $\mathbf{A}_s = \{d_s(t), d_s(t-1), \dots, d_s(t-9), y_s(t)\} | s=1, 2, \dots, 2850$,其中, $y_s(t)$ 为站点 s 在 t 时的前 1 h 累计降水量。进一步,将所有样本组 \mathbf{A}_s 按时间 t 进行

堆叠,得到原始的样本数据矩阵 $\mathbf{B} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_N \end{bmatrix}$,其中 N

为 13402。

最终,样本矩阵 \mathbf{B} 的总样本量为 1342184 个。表 1 为所选样本中降水量分布情况。

表 1 所选样本的降水量分布情况

	1 h 降水量					
	<1	1~5	5~10	10~20	20~50	≥50
样本数	325714	538182	224663	159742	87632	6251

2 研究方法

XGBoost 算法^[26]于 2011 年提出,能够利用 CPU 多线程进行并行运算,提升模型的泛化性能。作为 Boosting 集成学习算法家族中的一员,XGBoost 是一个树集成模型,算法模型如下式:

$$\hat{y}_i = \sum_{j=1}^K f_j(x_i), \quad f_j \in \mathbf{F} \quad (1)$$

其中, K 为 CART 回归树的棵数; \mathbf{F} 为所有决策树的集合; f_j 为第 j 棵树模型; x_i 为样本 i ; \hat{y}_i 为 XGBoost 算法对样本 x_i 的预测值。与一般的集成决策树算法不同,XGBoost 可在节点内选择最佳分裂点,并且其候选分裂点计算增益采用并行计算。其迭代的目标函数 L 为:

$$\begin{aligned} L^T &= \sum_{i=1}^m l(\hat{y}_i, y_i) + \sum_h^K \Omega(f_h) \\ &= \sum_{i=1}^m l[\hat{y}_i^{T-1}, y_i + f_T(x_i)] + \sum_h^K \Omega(f_h) \quad (2) \end{aligned}$$

式中, l 为训练损失, Ω 为树的复杂度, m 为总的训练样本数, y_i 第 i 样本的真实值。即第 i 样本在第 T 轮迭代后模型的预测值 \hat{y}_i^T 是在第 $T-1$ 轮模型的模型预测值 \hat{y}_i^{T-1} 的基础上加了一个新树 $f_T(x_i)$ 。而树的复杂度函数 $\Omega(f) = \gamma M + \frac{1}{2} \lambda \|\omega\|^2$,其中 M 为每棵树的叶子节点数量, ω 为每棵树的叶子节点的分数组成的集合, γ 和 λ 为系数;由此可见,目标函数由训练损失和叶子节点权重或树的深度等两部分构成,一方面可以控制模型的复杂度,另一方面可以防止过拟合。同时,它使用二阶泰勒展开式逼近目标函数,使得目标函数近似优化更接近实际值,从而提高预测精度。

XGBoost 有包括正则化项、学习率和决策树数量和树结构属性等众多超参数,超参数定义了模型的复杂度或学习能力等属性,表 2 为 XGBoost 算法主要参数的常用默认值。

表 2 XGBoost 算法主要参数的默认值

参数	XGBRegressor	XGBClassifier
Max_depth	6	6
Learning_rate	0.3	0.3
Estimator	100	100
Objective	reg:linear	reg:logistic
Booster	gbtree	gbtree
N_jobs	1	1
Reg_alpha	0	0
Reg_lambda	1	1

3 预报试验

3.1 试验建模样本处理

由 1.3 节的第(4)点可知,本研究使用样本数据的矩阵 \mathbf{B} 的总样本为 1342184 个,每一个样本由 10 个反射率因子和 1 个地面雨量观测值对应组成。以样本 i 为例,假设该样本所对应的时间为 t ,则该样本地面雨量观测值为 $y_i(t)$ 、10 个反射率因子可表示为 $\{d(t), d(t-1), \dots, d(t-9)\}$ 。为了方便,论文以总样本的前 100 万个样本作为 XGBoost 算法模型的建模样本,以总样本最后的 342184 个样本作为独立样本,用于测试模型的性能。由于雷达反射率因子存在杂波等问题,建模样本中相应地也存在包含大量噪声的样本,即噪声样本,这类样本会对模

型的训练产生负面影响。为了使模型在训练学习中尽可能避免受到噪声样本的干扰,在建模计算前,需要将建模样本中的噪声样本剔除:

首先, $Z-R$ 关系 ($z = aR^b$, 其中 $a = 300, b = 1.4$, 记为 $ZR(300, 1.4)$, 下同) 的雷达雨量估测方法, 作为 SWAN 系统的默认参数, 其估测能力应该是具有一定参考价值, 因此, 论文采用 $ZR(300, 1.4)$ 的估测作为参考, 判断该样本是否属于噪声样本。

其次, 用 $ZR(300, 1.4)$ 关系估测 1 h 的降水量时, 由于我国气象雷达观测是 6 min 一帧, 也就是 1 h 10 帧, 因此, 过去 1 h 累计降水量 \hat{R} 的估测计算式为:

$$\hat{R} = 6/60R(t-9) + 6/60R(t-8) + \dots + 6/60R(t) = \frac{1}{10} \sum_{j=0}^9 R(t-j) \quad (3)$$

其中 $R(t)$ 为 t 时刻的雨强。

最后, 记 D 为过去 1 h 的平均反射率因子; y 为过去 1 h 的实况累计降水量; 同时令 \bar{R} 为利用 $ZR(300, 1.4)$ 关系计算 D 得到的估测降水。则利用 $d = 10 \lg(z), z = aR^b$ 以及不等式的性质, 可以证明, 对同一时段内, $\hat{R} \geq \bar{R}$ 。

于是, 清除异常建模样本噪声的 5 条策略如下:

(1) 当 $D \geq 25$ dBz 且 $y = 0$ mm 时, 由于 $\hat{R} > \bar{R} > 1$ mm。经试验统计, 当 $\hat{R} > 1$ 时, 实况观测得到实况累计降水量 $y = 0$ mm 的概率小于 10%, 因此在采取的策略中, 将这部分样本去除。

(2) 当 $D \geq 40$ dBz 且 $y < 5$ mm 时, 由于 $\hat{R} > \bar{R} > 12$ mm 而 $y < 5$ mm, 因此, 这部分样本可视为不合理的样本, 应当去除。

(3) 当 $D \geq 45$ dBz 且 $y < 10$ mm 时, 由于 $\hat{R} > \bar{R} > 27$ mm 而 $y < 10$ mm, 因此, 可以认为这部分样本是不合理的, 可以去除掉。

(4) 当 $D \geq 50$ dBz 且 $y < 15$ mm 时, 由于 $\hat{R} > \bar{R} > 63$ mm 而 $y < 15$ mm, 估测与实况观测相差巨大, 因此, 这部分样本也应去除。

(5) 当 $D \geq 55$ dBz 且 $y < 20$ mm 时, 由于 $\hat{R} > \bar{R} > 144$ mm 而 $y < 20$ mm, 估测与实况观测相差甚远, 因此, 这部分样本应当去除。

按照这 5 个异常样本的剔除策略, 在 100 万个建模样本中, 共剔除了 27367 个样本, 即实际的建模样本数为 972633 个。

3.2 XGBoost 算法模型参数的确定

在进行建模试验计算前, 需要确定 XGBoost 算法模型的参数。为了避免过多参数调整而影响到模型的通用性问题, 论文对算法的众多参数, 除了其中的 Estimator, 其他参数均采用算法的默认参数值 (表 2)。为了确定出最佳的 Estimator 值, 以估计值和观测值之间的平均绝对误差 (MAE) 作为目标进行多次建模试验, 结果见图 2。

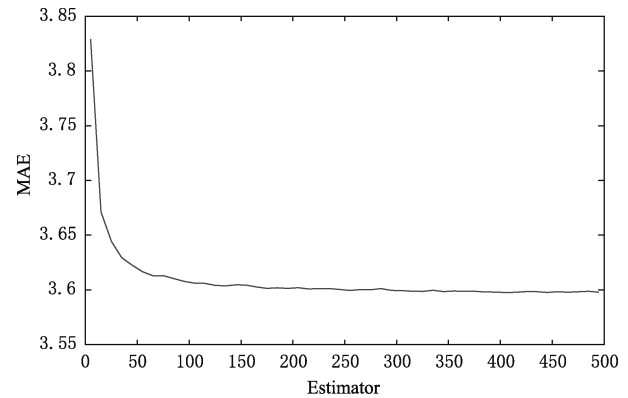


图 2 不同 Estimator 参数值下的 MAE 变化曲线

对图 2 进行分析可知, 当 Estimator 值增加到 200 以后, 模型的 MAE 指标变化不大。因此, 在建模试验部分, XGBoost 算法模型的参数最终确定为 Estimator=200, 其他参数均按照表 2 中的默认参数进行设置。

3.3 模型精度的评估指标

论文采用的估测精度评价指标分别为均方根误差 RMSE (Root Mean Square Error, R_{MSE})、平均绝对误差 MAE (Mean Absolute Error, M_{AE})、偏差 (BIAS, B_{IAS})^[28-29]、TS 评分 (Threat Score, T_S)^[30]。一般 RMSE 和 MAE 通常会受到强降水较大的影响, 主要反映强降水的估测效果。进一步利用 TS、空报比例 (PNB) 对各级降水区间进行分级评估, 以 \hat{y}_i 和 y_i 分别为样本 i 的预测值与观测值; n 为参与统计的样本量, 则这些评估指标分别定义如下:

$$R_{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (4)$$

$$M_{AE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (5)$$

$$B_{IAS} = \frac{\sum_{i=1}^n \hat{y}_i}{\sum_{i=1}^n y_i} \quad (6)$$

$$T_s = \frac{N_A}{N_A + N_B + N_C} \quad (7)$$

式(7)中 N_A 、 N_B 、 N_C 分别为正确、空报、漏报次数。该指标在 0~1 之间浮动,反应了估测降水的准确程度,这里 TS 值越大,表示降水估测的准确性越高。

4 试验结果及分析

论文先采用 XGBoost 算法对由第 3.1 节中得到的建模样本进行建模训练,随后利用训练好的模型对没有剔除异常样本的独立样本(因为在实际的定量降水估测中,无法判别测试样本是否是异常样本,所以独立样本不做异常样本的剔除)进行测试,测试结果采用本文 3.3 节的评价指标进行客观评估。同时为了更好地了解本文提出的雷达定量降水估测方法的性能,论文同时还利用目前美国和中国业务系统均采用的 Z-R 关系法对独立样本也进行了同样本的估测试验,具体结果见表 3。

表 3 两种估测方法对独立样本的估测精度对比

估测方法	RMSE	MAE	BIAS
XGBoost	6.04	3.50	1.05
ZR(300,1.4)	7.61	3.91	0.61

注:偏差越接近 1,效果越好;表中,黑色加粗数字为每一类的最好结果。

分析表 3, RMSE 和 MAE 的对比结果表明, XGBoost 估测降水方法相较于 ZR(300,1.4) 关系方法能取得更好的效果。32 万的独立样本估测的 RMSE 和 MAE 分别下降了 20.6% 和 10.3%。由于 RMSE 和 MAE 能够更好地反映强降水的估测效果,说明 XGBoost 方法在强降水条件下估测效果更好。同时, BIAS 指标上也表明, ZR(300,1.4) 关系方法对降水的量级明显低估。

为了进一步了解两种方法具体性能的对比情况,论文按照实况降水量强度 ≤ 10 mm/h、10~20 mm/h、20~30 mm/h、30~40 mm/h、40~50 mm/h、 ≥ 50 mm/h 划分为 6 个等级,并利用 3.3 节的评估指标分别对 6 个等级进行统计,结果见图 3。

由图 3a~b 可以看出, ZR(300,1.4) 关系方法在小于 10 mm/h 的降水估测上取得了更好的效果。而对于大于 10 mm/h 的降水估测, XGBoost 估测降水方法比 ZR(300,1.4) 关系方法在 RMSE 和 MAE 上的误差分别降低了 24%~31% 和 17%~21%,这可能是由于 XGBoost 模型能更好地拟合反射率因子与降水量的关系,从而降低了强降水估测的误差。

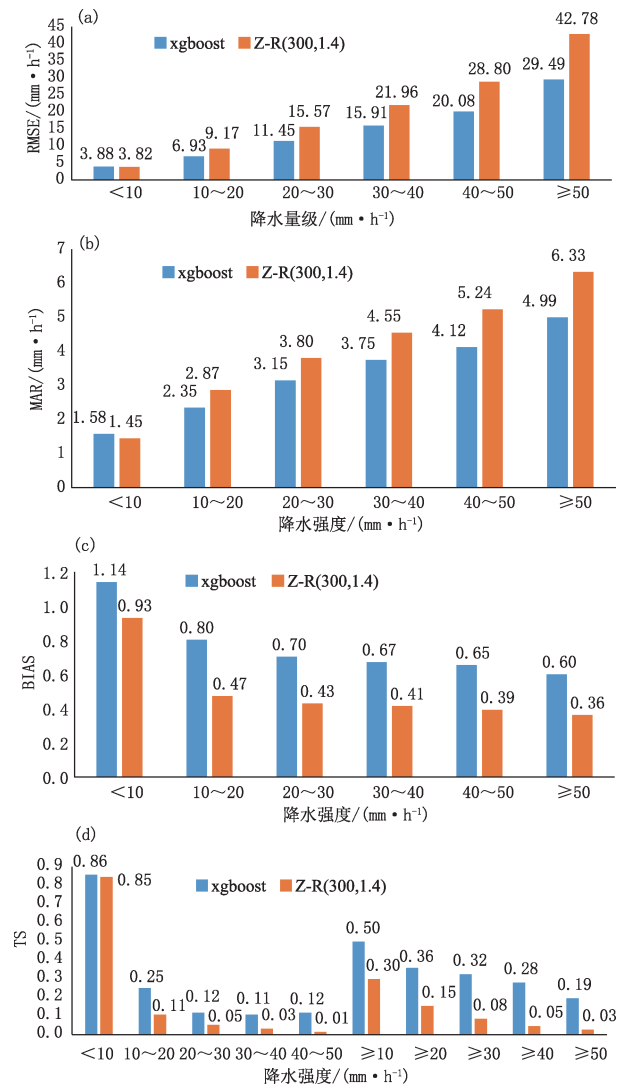


图 3 两种估测方法在独立样本上的估测精度检验

由图 3c 的 BIAS 结果可以看出: ZR(300,1.4) 关系方法在小于 10 mm/h 的降水估测上偏差较小,更接近于 1;而在大于 10 mm/h 的降水强度下, XGBoost 估测降水方法的估测偏差则更接近于 1,显示了更好的估测效果。从图 3d 中可以看出,除了在 >10 mm/h 范围外, XGBoost 估测降水方法不论是在 10~20 mm/h、20~30 mm/h、30~40 mm/h、 ≥ 50 mm/h 这 4 个等级,还是对于 ≥ 10 mm/h、 ≥ 20 mm/h、 ≥ 30 mm/h、 ≥ 40 mm/h 以及 ≥ 50 mm/h 这 5 个区间的雨量估测,其 TS 值均大幅优于 ZR(300,1.4) 关系方法。本文还进一步选取了独立样本中 2020 年 6 月 7 日 00:00:00 (UTC) 这一个例进行估测对比分析(图 4)。该个例有 53 个气象观测站的 1 h 降水量超过 20 mm,其中最大降水量为 93.5 mm。

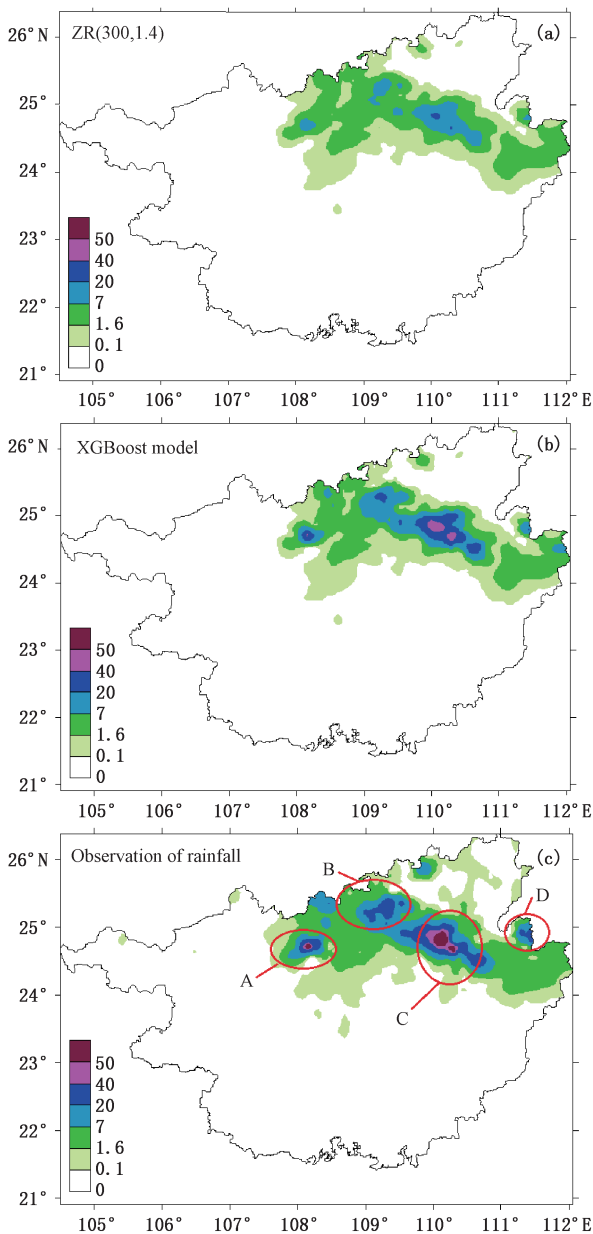


图 4 2020 年 6 月 7 日 00:00:00(UTC)降水估测对比图:
(a) ZR(300,1.4)方法预测的雨量分布,(b) XGBoost 模型预测的雨量分布,(c)实况观测雨量分布

由图 4 可见,观测的实况雨量(图 4c)分别有 4 个降水中心,分别为图中的 A、B、C 以及 D 区。通过对比可知,XGBoost 模型基本上能够完整估测出 C 区(主降水落区),并在 A 区和 B 区也稍有体现。而 ZR(300,1.4)方法(图 4a)得到的估测降水仅在主降水落区的 C 区稍有体现。进一步地,在降水量级的估测上,XGBoost 模型(图 4b)能够估测出超过 40 mm 以上的降水,并与实际雨量比较吻合,与之对

应的 ZR(300,1.4)关系法则没有估测出大于 40 mm 的降水。

综上所述,XGBoost 估测降水方法可以有效提高雷达估测强降水的精度,在降水强度大于 20 mm/h 时,其误差相对传统业务方法降低了 26%~31%,同时 TS 评分对于 ZR(300,1.4)关系方法有了较大提高。分析原因,可能是因为 XGBoost 方法能够在训练中更好地提取雷达回波与降水之间的非线性关系。对于降水强度小于 10 mm/h,论文提出的 XGBoost 估测降水方法与 ZR(300,1.4)关系方法的估测效果相当或略差,其原因可能是这一类的样本量大,包含的噪声也大,干扰了模型的训练效果。

5 结论

目前,在短时临近降水业务预报中,存在着利用各种方法外推出 20 帧(2 h)的雷达回波,然而,纵然外推出来的雷达反射率很接近实况,但在将雷达反射率转化为降水时,由于雷达的估测降水误差较大,往往无法较准确预测未来 1~2 h 强降水的落区。为此,本文利用雷达反射率因子和实况观测雨量资料,提出了一种基于雷达组网拼图和 XGBoost 的估测降水方法,该方法首先针对天气雷达存在多种难以避免的杂波问题,本文根据过去 1 h 的平均反射率值与地面实况降水量的异常关系提出了异常样本的剔除方法,从而尽可能地排除了异常样本对模型训练的影响。其次,针对雷达估测降水方法在强降水量级的估测存在较大偏差的问题,该方法利用 XGBoost 模型算法强大的非线性拟合能力,可以从训练样本中构建降水与雷达反射率因子之间的良好的非线性映射关系,从而提高了天气雷达估测强降水的效果。该方法在未进行异常样本剔除的独立样本测试中,本文提出的 XGBoost 估测降水方法与 ZR(300,1.4)关系方法相比,当降水强度大于 20 mm/h 时,XGBoost 模型较 ZR(300,1.4)方法在 RMSE 上降低了 26.48%~31.06%,MAE 降低了 17.17%~21.26%,TS 评分提高 100%以上。可见,XGBoost 估测降水方法明显提高了雷达定量测量强降水的精度。同时,目前业务系统提供的 Z-R 关系方法(ZR(300,1.4))在弱降水上的估测则优于 XGBoost 方法,因此,在实际的业务中可以将两则相结合,为雷达定量估测降水提供更好的精度。

参考文献

- [1] 李翠金. 中国暴雨洪涝灾害的统计分析[J]. 灾害学, 1996, 11(1): 59-63.
- [2] 李柏, 古庆同, 李瑞义, 等. 新一代天气雷达灾害性天气监测能力分析及其未来发展[J]. 气象, 2013, 39(3): 265-280.
- [3] 曾正茂, 刘黎平. 山西地区 C 波段新一代天气雷达(CC)联合雨量计降水效果对比与分析[J]. 成都信息工程学院学报, 2016, 31(1): 54-64.
- [4] Mapiam P P, Sriwongsitanon N. Climatological Z-R relationship for radar rainfall estimation in the upper ping river basin [J]. Scienceasia, 2008, 34(2): 215-222.
- [5] 魏鸣, 宋羽轩, 张雷. 定量测量降水量的回波结构自适应算法研究[J]. 大气科学学报, 2018, 41(5): 102-109.
- [6] 赵淑芳, 张立文, 王倩, 等. 山东极端强降雨分钟雨量、Z-R 关系和风暴结构演变特征[J]. 干旱气象, 2017, 035(005): 806-814.
- [7] 郑永光, 张小玲, 周庆亮, 等. 强对流天气短时临近预报业务技术进展与挑战[J]. 气象, 2010, 36(7): 33-42.
- [8] 郑媛媛, 谢亦峰, 吴林林, 等. 多普勒雷达定量估测降水的三种方法比较试验[J]. 热带气象学报, 2004, 20(2): 192-197.
- [9] 高烽, 周新尧, 孔凡超, 等. 滑动时间窗长度对太行山区雷达反演降水的影响[J]. 气象, 2018, 44(4): 72-81.
- [10] Goudenhoofd E, Delobbe L. Evaluation of radar-gauge merging methods for quantitative precipitation estimates [J]. Hydrology and Earth System Sciences, 13, 2(2009-02-18), 2009, 13(2): 195-203.
- [11] Hasan M M, Sharma A, Mariethoz G, et al. Improving radar rainfall estimation by merging point rainfall measurements within a model combination framework [J]. Advances in Water Resources, 2016, 97: 205-218.
- [12] Seo D J. Conditional bias-penalized kriging (CBPK) [J]. Stochastic Environmental Research and Risk Assessment, 2013, 27(1): 43-58.
- [13] 赵华生, 金龙, 黄小燕, 等. 基于 CNN 和 RF 算法的 ECMWF 降水分级订正预报方法[J]. 气象科技, 2021, 49(3): 419-426.
- [14] 朱岩, 翟丹华, 吴志鹏, 等. 基于 Xgboost 算法的短时强降水预报方法[J]. 气象科技, 2021, 49(3): 406-418.
- [15] 方楠, 姜舒婕, 闫晓敏, 等. 基于长短期记忆神经网络的风速超短期快速滚动预报技术[J]. 气象科技, 2022, 50(6): 842-850.
- [16] 赵生昊, 覃彬全, 杜乐. 基于机器学习和单站地面气象要素的雷电临近预警方法[J]. 气象科技, 2022, 50(1): 121-128.
- [17] 杜智涛, 姜明波, 杜晓勇, 等. 机器学习在气象领域的应用现状与展望[J]. 气象科技, 2021, 49(6): 930-941.
- [18] 龙亚星, 黄勤, 李成伟. 基于机器学习技术的蒸发皿蒸发量估算模型[J]. 气象科技, 2021, 49(2): 166-173.
- [19] 尹麒名, 甘建红, 漆慧, 等. 一种改进的循环神经网络雷达图像外推算法[J]. 气象科技, 2021, 49(1): 18-24+45.
- [20] 黄彬, 吴铭, 孙舒悦, 等. 基于深度学习的卫星多通道图像融合的海雾监测处理方法[J]. 气象科技, 2021, 49(6): 823-829+850.
- [21] Xiao R, Chandrasekar V. Development of a neural network based algorithm for rainfall estimation from radar observations [J]. IEEE Transactions on Geoscience and Remote Sensing, 1997, 35(1): 160-171.
- [22] Ju Y S, Ro Y, Cha J W, et al. Assessing the applicability of random forest, stochastic gradient boosted model, and extreme learning machine methods to the quantitative precipitation estimation of the radar data: A case study to Gwangdeoksan radar, South Korea, in 2018 [J]. Advances in Meteorology, 2019.
- [23] Zhang Changjiang, Wang Huiyuan, Zeng Jing, et al. Short-term dynamic radar quantitative precipitation estimation based on wavelet transform and support vector machine [J]. Journal of Meteorological Research, 2020, 34(2): 228-241.
- [24] 傅德胜, 肖晨, 谭畅, 等. RBF 神经网络在雷达定量估测降水中的应用研究[J]. 气象科学, 2015(2): 199-203.
- [25] 殷志远, 彭涛, 杨芳, 等. 基于 QPE 和 QPF 的遗传神经网络洪水预报试验[J]. 暴雨灾害, 2013, 32(4): 360-368.
- [26] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System [C]// the 22nd ACM SIGKDD International Conference, ACM, 2016.
- [27] 杜扬帆, 伍孝飞, 乔百友. 基于 XGBoost-PredRNN++ 的海表面温度预测[J]. 计算机系统应用, 2022, 31(10): 236-244.
- [28] Wang G, Liu L, Ding Y, et al. Improvement of radar quantitative precipitation estimation based on real-time adjustments to Z-R relationships and inverse distance weighting correction schemes [J]. Advances in Atmospheric Sciences, 2012, 29(3): 575-584.
- [29] Bennett N D, Croke B F, Guariso G, et al. Characterising performance of environmental models [J]. Environmental Modelling & Software, 2013, 40(2): 1-20.
- [30] 朱岩, 翟丹华, 吴志鹏, 等. 基于 XGBoost 算法的短时强降水预报方法[J]. 气象科技, 2021, 49(3): 406-418.

Radar Quantitative Precipitation Estimation Based on Radar Mosaic and XGBoost Algorithm

ZHAO Huasheng¹ LI Xiaojing²

(1 Guangxi Institute of Meteorological Sciences, Nanning 530006; 2 Guangxi College for Preschool Education, Nanning 530006)

Abstract: To address the problem of large bias in the quantitative estimation of heavy precipitation by traditional methods using weather radar, the thesis uses the 1-hour cumulative rainfall as the estimation object, a new model for radar precipitation estimation based on radar mosaic data and XGBoost (eXtreme Gradient Boosting) algorithm. The model is designed with the radar combined reflectance factor of the previous hour as the input factor, and further employs several rejection strategies of anomalous samples to effectively remove some of the noise from the modelling samples, thus better constructing a non-linear mapping relationship between the radar combined reflectance and the estimated object. The root mean square error (RMSE) is 6.04 mm, the mean absolute error (MAE) is 3.50 mm, and the forecast bias (BIAS) is 1.05 for the 320,000 independently tested samples; compared to the ZR(300,1.4) relational method currently used on operational systems, the RMSE and MAE of the former decrease by 20.6% and 10.3% respectively, while the BIAS indicators show a significant underestimation of precipitation magnitude by the latter. For samples with hourly rainfall intensity greater than 10 mm, further statistical results show that the new scheme's RMSE, MAE and TS scores are substantially better than the ZR (300, 1.4) relational method for practical operational applications.

Keywords: radar precipitation estimation; XGBoost; combined reflectivity; radar mosaic