

基于随机森林模型的长江流域分区多源融合降水模拟方法研究

宋蕾玥^{1,2}, 张珂^{1,2,3,4,5}, 晁丽君^{1,2}, 李曦^{1,2}, 牛杰帆^{1,2}, 黄轶铭¹

(1. 河海大学水文水资源学院, 江苏 南京 210098; 2. 河海大学水灾害防御全国重点实验室, 江苏 南京 210098;
3. 河海大学长江保护与绿色发展研究院, 江苏 南京 210098; 4. 中国气象局水文气象重点开放实验室,
江苏 南京 210024; 5. 水利部水利大数据重点实验室, 江苏 南京 210098)

摘要: 基于3种卫星降水产品, 提出了一种基于随机森林模型的长江流域分区多源融合降水模拟算法(FCM-RF算法)。采用模糊C均值算法, 结合地面观测站点资料对长江流域进行降水区域划分, 引入降水比降刻画降水空间性, 进一步通过普通克里金插值法优化融合结果, 得到一套长江流域空间分辨率为 $0.25^{\circ} \times 0.25^{\circ}$ 的多源融合降水产品, 并对其进行了评估。结果表明: FCM-RF算法在长江流域具有良好的表现, 可以有效提高原始卫星降水产品对于降水事件的捕捉能力, 在验证站点模拟降水量与实测降水量的相关系数可达到0.76; FCM-RF算法在年际上具有相似变化特征, 对于春秋降水的敏感性较高, 在夏季由于强降水影响表现欠佳, 冬季由于雨量稀少、存在固态降水, 呈现出误差小、相关系数较低的特点; FCM-RF算法在东南地区具有较强的降水捕捉能力, 在青藏高原地区的准确性较低。

关键词: 降水模拟; 模糊C均值算法; 随机森林模型; 分区多源融合方法; FCM-RF算法; 长江流域
中图分类号: P456.7 **文献标志码:** A **文章编号:** 1004-6933(2024)03-0125-08

Research on zoning multi-source fusion precipitation simulation method of the Yangtze River Basin based on random forest model// SONG Leiyue^{1,2}, ZHANG Ke^{1,2,3,4,5}, CHAO Lijun^{1,2}, LI Xi^{1,2}, NIU Jiefan^{1,2}, HUANG Yiming¹
(1. College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China; 2. The National Key Laboratory of Water Disaster Prevention, Hohai University, Nanjing 210098, China; 3. Yangtze Institute for Conservation and Development, Hohai University, Nanjing 210098, China; 4. China Meteorological Administration Hydro-Meteorology Key Laboratory, Nanjing 210024, China; 5. Key Laboratory of Water Big Data Technology of Ministry of Water Resources, Nanjing 210098, China)

Abstract: A multi-source fusion precipitation simulation method for the Yangtze River Basin based on the random forest model (FCM-RF algorithm) was proposed using three types of satellite precipitation products. Using the fuzzy C-means algorithm and ground observation station data, the precipitation regions in the Yangtze River Basin were divided. The precipitation ratio was introduced to characterize the spatial distribution of precipitation. The fusion results were further optimized using the ordinary Kriging interpolation method. A set of multi-source fusion precipitation products with a spatial resolution of $0.25^{\circ} \times 0.25^{\circ}$ in the Yangtze River Basin were obtained and evaluated. The results show that the FCM-RF algorithm has good performance in the Yangtze River Basin, which can effectively improve the capture ability of original satellite precipitation products for precipitation events. The correlation coefficient between simulated precipitation and measured precipitation at the validation stations can reach 0.76. The FCM-RF algorithm has similar interannual variation characteristics and is highly sensitive to precipitation in spring and autumn. It performs poorly in summer due to heavy rainfall, and exhibits small errors and low correlation coefficients in winter due to sparse rainfall and solid precipitation. The FCM-RF algorithm has strong precipitation capture ability in the southeast region, but its accuracy is low in the Qinghai Tibet Plateau region.

Key words: precipitation simulation; fuzzy C-means algorithm; random forest model; zoning multi-source fusion method; FCM-RF algorithm; Yangtze River Basin

基金项目: 国家重点研发计划项目(2023YFC3006505); 中央高校基本科研业务费专项资金资助项目(B240203007)

作者简介: 宋蕾玥(1999—), 女, 硕士研究生, 主要从事水文水资源研究。E-mail: 211301010101@hhu.edu.cn

通信作者: 张珂(1979—), 男, 教授, 博士, 主要从事水文水资源研究。E-mail: kzhang@hhu.edu.cn

随着全球气候变暖,水文循环加剧,全球的降水差异性更加明显^[14]。降水的量化估测对于气候监测以及水资源科学规划与管理起着至关重要的作用^[5]。目前,降水估测主要有地面雨量站、雷达测雨和卫星遥测3种手段^[6]。地面雨量站可以提供直接、准确性高的观测数据,但受地形等条件限制,无法获得大范围精确降水数据;天气雷达可以捕捉降水的空间分布,却要克服山区波束阻塞、异常传播误差和不精确的后向散射与降水率的关系^[7]。利用卫星观测可以捕获大尺度降水的时空变化,提供具有实时性的降水数据^[5,8,9]。利用现有降水捕捉手段与卫星产品进行融合开发出具有更高空间以及时间尺度的降水数据集,对不同地区的水文研究具有重要意义。多源集合降水数据,如 MSWEP (multi-source weighted-ensemble precipitation)^[10],为水文建模应用等提供了支持。目前,贝叶斯方法、最优插值法、卡尔曼滤波法等方法已经广泛应用于国内外降水融合研究^[11-15]。此外,克里金插值法作为一种常见的地统计方法,具有灵活处理不规则空间分布的优点,在海拔与降水关系较为密切的区域的模拟精度较高^[16-17]。

随着信息技术的进一步发展,机器学习方法由于其更强的学习泛化能力,在水文学研究中也得到了广泛关注^[18]。Wu 等^[19]利用卷积神经网络 (convolutional neural networks, CNN) 与长短期记忆 (long short-term memory, LSTM) 神经网络相结合,将中国地区热带降雨测量任务 (tropical rainfall measuring mission, TRMM) 原始相关系数提升至 0.72。而随机森林模型作为一种可以处理复杂变量之间的非线性关系的机器学习方法,能够较好地避免过拟合问题的出现,广泛运用于多源卫星融合^[20]。然而利用随机森林模型进行空间上的具体预测时,往往难以厘清观测值之间可能存在的空间相关性,将经纬度、到最近网格的距离、到海岸线的距离作为表征观测值地理联系的特征值结合随机森林模型使用是目前采用的解决方法^[21-22]。

过往研究多着眼于单个观测点的地理特征而鲜少注重降水区域相似性对于降水融合的影响,通过对降水区域的划分可以考虑到不同区域的气候、植被、地形等差异,将具有类似降水特征的区域划分为同一区域,可以使降水融合更具有针对性。本文提出了一种基于随机森林的分区融合 (fuzzy C-means and random forests, FCM-RF) 算法,利用 3 种卫星降水产品协同地面站点资料进行长时段的逐日降水模拟研究,选择表征降水空间连续性的降水比降作为影响因子,获得长江流域 2009—2019 年逐日融合降

水产品,以期更准确地捕捉降水的时空分布特征,为防洪抗旱和应急响应提供重要支持。

1 研究区概况与数据来源

1.1 研究区概况

长江流域是世界第三大流域,流域总面积 180 万 km²,位于 24°27'N ~ 35°54'N、90°33'E ~ 122°19'E。流域整体呈现出东高西低的特点,横跨我国的三大阶梯,流经青藏高原、横断山区、云贵高原、四川盆地以及长江中下游平原^[23-25]。流域内气候类型多变,以亚热带季风气候为主。由于长江流域地形复杂及其典型的季风气候,流域内降水时空分布呈现不均匀性,流域多年平均降水量为 1022.3 mm^[26]。长江流域大部分区域属于湿润带,干旱带、半湿润带、特别湿润带都在流域内有所分布,水文气象要素的空间差异性显著,流域雨量站分布如图 1 所示。

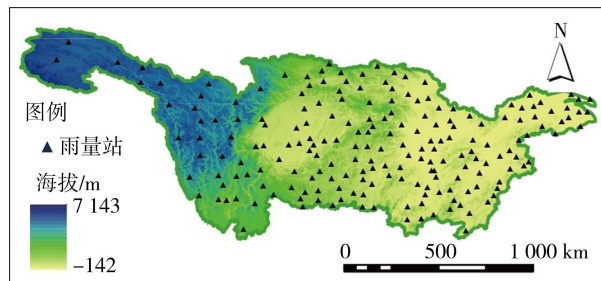


图 1 长江流域雨量站分布

Fig. 1 Distribution of rainfall stations in the Yangtze River Basin

1.2 数据来源

本文所用的地面降水资料来自中国气象数据网中的中国地面气候资料日值数据集 V3.0 (<http://data.cma.cn>),选用长江流域内 198 个站点 2009—2019 年逐日降水数据。选用的卫星数据包括 CMORPH (climate prediction center morphing technique)、IMERG (integrated multi-satellite retrievals for GPM) 以及 PERSIANN (precipitation estimation from remotely sensed information using artificial neural networks)。CMORPH 是一种利用地球同步卫星红外图像的运动矢量来传播被动微波数据中降水估计数据^[27]。全球降水观测 (global precipitation measurement, GPM) 是继 TRMM 之后由 NASA 和日本宇宙航空研究开发机构的合作任务, GPM 包括了一个核心卫星、GPM 微波成像仪、双波长降水雷达、Ku 降水雷达以及 Ka 降水雷达^[5,7]。集成多卫星反演产品 IMERG 是 GPM 生成的三级降水产品。PERSIANN 是由亚利桑那大学开发的基于人工神经网络估测降水的系统,主要利用地球静止红外图像

以及地面信息进行反演^[28]。本文采用的 CMORPH_V1.0_BLD 日降水数据具有 $0.25^{\circ} \times 0.25^{\circ}$ 的空间分辨率;IMERG-F 日降水数据具有 $0.1^{\circ} \times 0.1^{\circ}$ 的空间分辨率,数据由 NASA 官网获得 (<https://www.nasa.gov>);PERSIANN-CDR 日降水数据具有 $0.25^{\circ} \times 0.25^{\circ}$ 的空间分辨率。分别选取了 3 种卫星 2009—2019 年的降水资料,对 IMERG 数据使用双线性插值方法统一空间分辨率至 $0.25^{\circ} \times 0.25^{\circ}$,构建模型时所需的卫星产品月降水总量数据通过逐日数据累加获得。

地形高程数据采用了 NASA SRTM 数据,分辨率为 90 m,对长江流域海拔、坡度、坡向进行计算并提取至 $0.25^{\circ} \times 0.25^{\circ}$ 的栅格像元。归一化植被指数(normalized difference vegetation index, NDVI)数据来源于 MODIS 土地覆盖类型产品 MOD13A2,本文采用的 NDVI 数据时间分辨率为 16 d,空间分辨率为 1 km。地表温度数据来源于国家青藏高原科学数据中心(<http://data.tpdc.ac.cn>),采用中国大陆及周边逐日 1km 全天候地表温度数据集长江流域 2009—2019 年的逐日温度数据。利用时间插值方法将 NDVI 数据插值至时间分辨率为 1 d,同时地表温度数据及 NDVI 数据均重采样至 $0.25^{\circ} \times 0.25^{\circ}$ 的空间分辨率。构建模型时所需的 NDVI、地表温度月均值数据由逐日数据计算所得。

2 研究方法

2.1 分区多源融合方法

利用 FCM-RF 算法进行降水融合的步骤为:
①根据海拔、经纬度、年均降水量对长江流域进行降水分区;
②利用随机森林模型,协同气象地形因子以及 3 种卫星降水数据,对长江流域月降水进行估计;
③根据普通克里金方法优化后的降水日占比获得空间分辨率为 0.25° 的长江流域日模拟数据。

此外,采用不分区的随机森林-克里金(random forests and ordinary Kriging, RF-OK)算法和原始随机森林(original random forests, Ori-RF)算法对长江流域的降水进行预估,作为对照组与 FCM-RF 算法结果进行对比。RF-OK 算法以长江流域整个区域的海拔、坡度、坡向、NDVI 值、地表温度值、降水比降以及 3 种卫星降水产品逐月数据为输入数据,建立随机森林模型,模拟出整个区域的月降水总量,再进一步使用普通克里金插值优化得出最终模拟结果;Ori-RF 算法以地形、气象和卫星降水产品的逐日数据为输入数据,构建随机森林模型进行预测,直接获得流域整体的逐日模拟结果。

2.1.1 模糊 C 均值算法

模糊 C 均值算法是通过不断优化目标函数来获得样本对于聚类中心的隶属度矩阵,从而实现数据分类的模糊聚类算法。模糊 C 均值方法的目标函数需要获得各样本点到聚类中心的欧式距离的最小值,计算公式为

$$J_{\min}(U,V)=\sum_{i=1}^n\sum_{j=1}^k u_{ij}^M\|x_j-c_i\|^2 \tag{1}$$

式中: $J_{\min}(U,V)$ 为目标函数; U 为代表聚类结果的模糊矩阵; V 为聚类中心; M 为控制聚类模糊程度的参数; u_{ij} 为样本 x_j 对于第 k 类的隶属度; $\|x_j-c_i\|^2$ 为样本 x_j 与样本中心 c_i 的欧氏距离。误差平方和函数可以对聚类数进行选择,衡量数据点与聚类中心的距离。本文误差平方和函数值在聚类数取值为 6 之后下降程度变缓,因此选择划分为 6 个区域,这与 Chen 等^[29]对长江流域的分区数目一致。

2.1.2 降水比降计算

基于地理空间上的自相关原则,附近的观测点往往会呈现出相似的降水量。He 等^[21]利用了距离最近干网格的距离来描述降水在空间上的特征。在构建随机森林模型的过程中,为了展现降水的空间连续性,加入了降水比降作为新的协变量。降水比降为空间上两点的降水差值与两点的距离之比,这也是通过不同观测点之间的降水变量值来量化降水的空间相关性。当降水比降较小时,相邻站点间降水差异较小,即降水具有较强的空间自相关性,变化趋势较为平缓。降水比降计算公式为

$$v_0=\frac{p_{x_i}-p_{x_0}}{\|x_i-x_0\|^2} \tag{2}$$

式中: v_0 为 x_0 点的降水比降值; p_{x_i} 为距离 x_0 点最近的站点 x_i 的降水量; p_{x_0} 为站点 x_0 的降水量; $\|x_i-x_0\|^2$ 为站点 x_i 与 x_0 点的欧氏距离。在经纬网格中, p_{x_i} 取目标网格上、下、左、右侧相邻网格的降水量均值。

2.1.3 随机森林模型

随机森林算法是由 Breiman^[20]于 2001 年提出的集成学习算法,该算法是基于决策树的集成算法。随机森林模型采用了 Bagging 法从原始训练样本中抽取数据^[30],Bagging 法是一种有放回的抽样方法,可以建立独立的原始数据子集,并且子集中可以出现重复的元素,再利用被抽取的样本构建出多个决策树。在每轮随机采样过程中,未被抽样的数据被称为袋外数据,随机森林模型可以通过袋外数据对模型的泛化能力进行评估。而 Bagging 法的集合策略对于回归问题采用了平均的方法,得出模型的结果。随机森林算法具有实现简单、泛化能力强的特点,并且可以通过基尼系数得到特征重要性^[31]。

利用海拔、坡度、坡向、NDVI 值、地表温度值、降水比降,协同 3 种卫星的降水逐月数据作为输入,根据分区结果建立多个独立随机森林模型。在进行随机森林建模时,测试集和训练集的划分有多种方式,本文为了具体探究每个区域的降水特征,在每个分区中随机抽选 5 个站点作为测试集,剩下的站点数据作为训练样本,利用贝叶斯优化方法对模型进行调参。

2.1.4 普通克里金插值法

克里金插值法是一种利用已知点的加权值来对未知点进行预测的插值方法,权重值通过广义最小二乘法进行最优估计来得到。克里金插值法分为普通克里金插值法和协同克里金插值法。普通克里金插值法假设空间上属性值是平稳的,属性值在各个空间位置的数学期望都是同一个常数,计算公式为

$$\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \tag{3}$$

式中: $\hat{Z}(x_0)$ 为 x_0 点的预测值; λ_i 为 i 位置测量值的权重; $Z(x_i)$ 为 i 位置的测量值; n 为数据点个数。

根据卫星的日降水与月降水总量的占比可以得到卫星的月降水具体分配,而 CMORPH、IMERG、PERSIANN 3 种降水产品在站点位置处的降水日占比与站点实际降水日占比相关性均不高,本文选择 CMORPH 产品的降水日占比进行优化。将随机森林模型训练样本中的站点实际降水日占比与 CMORPH 降水日占比在站点尺度上的差值利用普通克里金插值法进行插值统一至 $0.25^\circ \times 0.25^\circ$ 经纬度网格,与原始 CMORPH 的降水日占比求和,得到优化后的网格化降水日占比。随机森林模型模拟得到的月降水总量值与降水日占比的乘积即生成最终的 FCM-RF 算法数据集。

2.2 精度评定方法和指标

将每个分区内随机抽取的站点作为随机森林模型测试集,利用未参加随机森林模型构建的普通克里金插值法的站点对最终生成的日降水融合产品进行验证。使用相关系数(R)、平均绝对误差(MAE)和均方根误差(RMSE)对融合产品进行评价。

3 结果与分析

3.1 分区结果

图 2 为利用模糊 C 均值算法对长江流域进行降水分区的结果。与以往按照水系分区不同,本文将长江流域分为 6 个区域,分别为西部青藏高原区(I 区)、西南部山区(II 区)、中北部山地丘陵区(III 区)、中南部山区(IV 区)、长三角平原区(V 区)、鄱阳湖平原区(VI 区)。图 3 为各分区年均降水量箱

型图,可见 II 区、III 区、VI 区的箱体上下下限范围之外出现了离群值,即该站点具有与区域内其余站点明显不同的极端降水值,其他分区的站点群未出现极端异常值。由图 2 可见, I 区位于高海拔的青藏高原地区,站点分布较为稀疏,年均降水量为 500 ~ 700 mm,是所有区域年均降水量最少的区域; II 区属于西南山区,包括了部分云南地区,区域内山地多,地形起伏,区域年均降水量为 800 ~ 1000 mm,有 1 个站点出现超出箱体上限的异常值,其年均降水量远高于其他站点,这可能是由西南部多山区短时强降水的发生导致。 III 区主要包括嘉陵江水系以及汉江水系,区域内站点年均降水量的分布与 II 区相似,该区域内存在年均降水量约为 500 mm 和 1400 mm 的异常站点。 IV 区主要包括乌江水系以及洞庭湖水系,年均降水量为 1100 ~ 1400 mm, IV 区与 V 区的年均降水量分布相似。 VI 区是年均降水量最高的区域,主要包括鄱阳湖平原,年均降水量为 1500 ~ 1800 mm,存在年均降水量超过 2500 mm 的异常站点。总体各分区内站点分布比较均匀,各分区年均降水量呈现由东南向西北逐渐减少的特点。该分区结果与 Chen 等^[29]的研究成果一致。

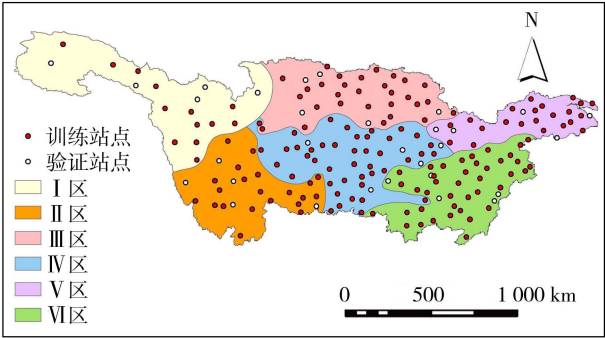


图 2 长江流域降水分区结果

Fig. 2 Results of precipitation zoning in the Yangtze River Basin

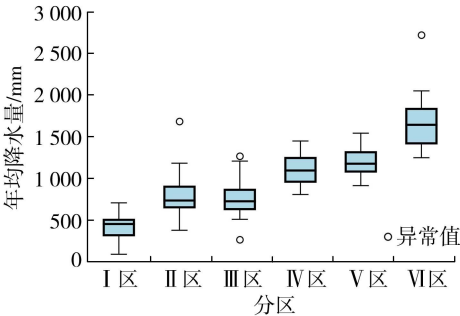


图 3 各分区年均降水量箱型图

Fig. 3 Box plots of annual average precipitation for each division

3.2 降水量月尺度模拟结果评价

图 4 为 FCM-RF 算法分区与 RF-OK 算法整体

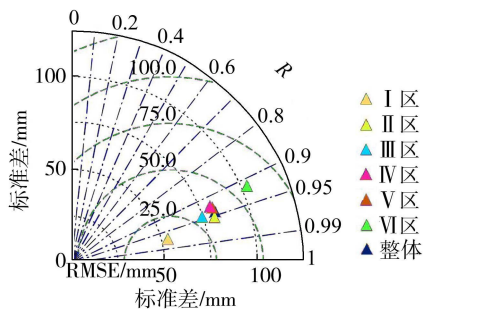


图4 FCM-RF 算法分区与 RF-OK 算法整体降水量月均值模拟评价结果

Fig.4 Evaluation results of monthly average precipitation simulation from zoning of FCM-RF algorithm and whole of RF-OK algorithm

降水量月均值模拟评价结果,可见 FCM-RF 算法的分区预测结果以及 RF-OK 算法的整体预测结果的相关系数均大于 0.9。I 区预测结果相关系数达到了 0.96,且具有较低的标准差,预测结果离散程度较低;II 区、III 区、IV 区、V 区以及整体预测结果相关系数均在 0.95 左右;VI 区的预测结果的相关系数稍低于其他组,标准差大于 100 mm, RMSE 在 50 mm 左右,预测结果与真实值有较大误差,可能由于 VI 区是流域内年降水量最丰沛的区域,受东南季风以及海洋环流的显著影响,随机森林模型对于该区域降水量的捕捉存在着不确定性。除 VI 区外,其他分区的随机森林模型中,卫星降水重要性都占据了重要性值的 80% 以上。在 I 区、II 区、VI 区的随机森林模型中,最重要的特征为 CMORPH 降水,其特征重要度分别为 67.2%、58.0%、53.6%;在 III 区、IV 区、V 区的随机森林模型中,最重要的特征为 IMERG 降

水,其特征重要度分别为 79.3%、46.1%、50.8%。

3.3 降水量日尺度模拟结果评价

3.3.1 站点结果评价

图 5 为 CMORPH、PERSIANN、IMERG 原始卫星降水产品和 Ori-RF 算法、RF-OK 算法、FCM-RF 算法等不同方法日尺度模拟结果与站点实测降水量拟合情况。由图 5 可见,CMORPH、PERSIANN、IMERG 3 种原始卫星降水产品与站点实测降水量的相关系数均处于较低水平,且 RMSE、MAE 较大。CMORPH 精度略高于 PERSIANN 和 IMERG,与站点实测降水量的相关系数超过了 0.5。3 种原始卫星降水产品中,PERSIANN 产品在长江流域的精度最低,与站点实测降水量的相关系数只达到了 0.34。Ori-RF 算法、RF-OK 算法、FCM-RF 算法模拟结果较 3 种原始卫星降水产品的精度都有所提高。Ori-RF 算法模拟结果的 RMSE 提升较大,模拟结果与站点实测降水量的相关系数提高不大,为 0.5 左右。RF-OK 算法与 FCM-RF 算法模拟的效果相似,FCM-RF 算法模拟结果与站点实测降水量的相关系数略高于 RF-OK 算法,达到了 0.76,同时 RMSE 降低到了 5.87 mm,说明分区可以将地形气候等较为相似的地区归为一类进行模拟,有利于模拟精度的提升。RF-OK 算法、FCM-RF 算法模拟精度相近的原因可能是研究区范围较大,气候、地形等影响降水的因素变化性强且复杂,估测逐日降水量较为精细,模型捕捉难度较大。

3.3.2 时间尺度分析

图 6 为 FCM-RF 算法评估指标变化情况。由

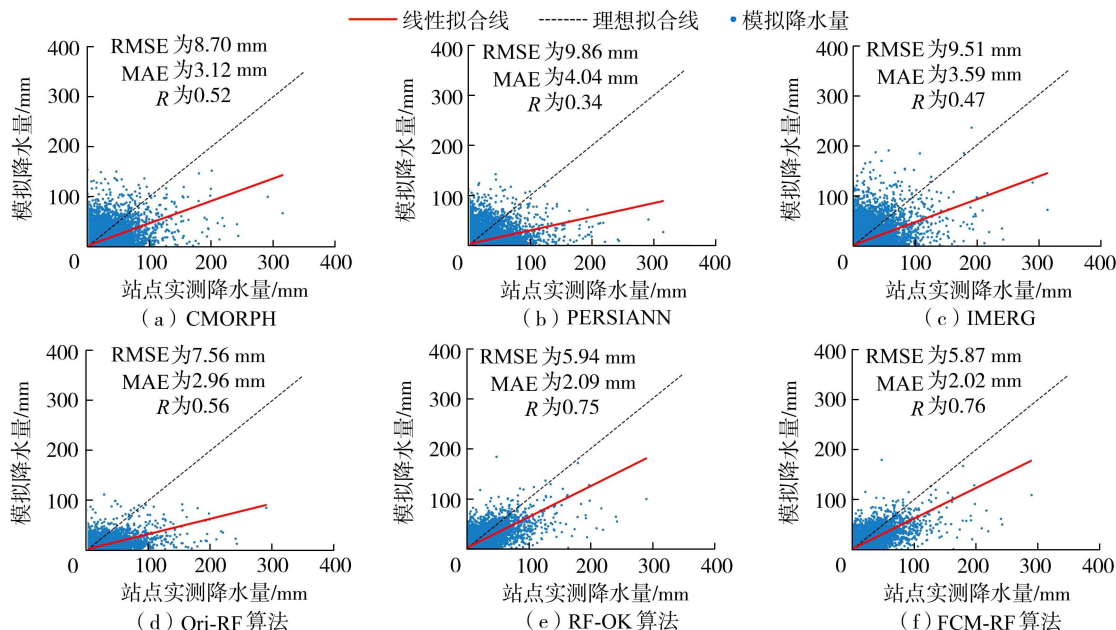


图5 不同方法日尺度模拟结果与站点实测降水量拟合情况

Fig.5 Fitting of daily scale simulation results using different methods with measured precipitation at stations

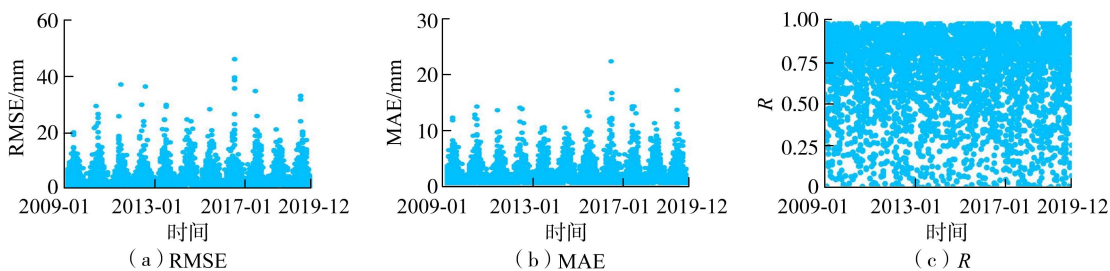


图6 FCM-RF 算法评估指标变化情况

Fig. 6 Changing of FCM-RF algorithm evaluation indexes

图6可见,整体上 FCM-RF 算法模拟结果与降水量实测值具有较高的相关性, RMSE、MAE 的变化趋势相似。冬季 RMSE、MAE 值都处于较低水平,夏季降水误差增大,出现这种现象的主要原因是夏季受到对流、水汽输送等复杂因素影响,具有较小时空尺度的强降水事件时有发生,因此对于夏季降水量的准确估测难度也增大。2016 年 7 月, FCM-RF 算法模拟结果与实测降水量的 MAE、RMSE 同时出现了极大值, RMSE 超过了 40 mm, MAE 超过了 20 mm, R 最大值为 0.9 左右,这与当时长江流域受强降雨影响,在中下游发生区域性大洪水的事实相符。FCM-RF 算法与站点实测降水量的 R 值在大部分时间内都处于 0.75 以上,没有明显随季节变化。2010 年 7 月、2014 年 7 月、2016 年 7 月、2017 年 7 月的 R 值最大值较其他时间有明显的下降,与对应时间的 RMSE、MAE 变化趋势吻合。由此可见, FCM-RF 算法整体上对于长江流域降水的预测具有较低的误差以及良好的相关性,可以对长江流域降水较好地估测,但对于夏季暴雨的预测还存在偏差,对于短时强降水的捕捉能力需要进一步提高。

3.3.3 空间尺度分析

图 7 为 FCM-RF 算法在验证站点不同季节模拟结果的相关系数空间分布。FCM-RF 算法的模拟结果在多数站点全年都具有较高的相关性, R 值都在 0.7 以上。位于 I 区的站点大多分布于青藏高原,全年的 R 值都呈现出一个较低的水平,为 0.4 ~ 0.6,这是由于青藏高原地区海拔差异明显,存在大规模的山脉和高原地形,受到复杂的局地效应和地形扰动的影响,算法对于降水事件的模拟精度会降低,可见 FCM-RF 算法在面临地形气候条件复杂的地区仍具有较大的不确定性。而对于东南沿海地区, R 值处于较高的水平,夏季、冬季的 R 值较春秋季节略低,一部分站点 R 值为 0.7 ~ 0.8,春秋季节 R 值均在 0.8 以上,这可能是由于夏季短时强降水具有不确定性,而冬季可能出现的固体降水对于卫星来说更难观测到,加大了融合降水产品对于降水的

捕捉难度。对于长江流域中部地区(Ⅲ区、Ⅳ区) FCM-RF 算法的表现也有夏季精度略低、冬季次之、春秋季节较好的特点。而从 R 值的空间分布来看,东南沿海地区 FCM-RF 算法模拟结果与实测降水量具有较好的一致性, R 值可以达到 0.9。在 FCM-RF 算法月尺度模拟结果的分析中,东南沿海地区较其他地区具有略大的误差;在 FCM-RF 算法日尺度模拟的过程中,降水日占比以及月尺度模拟结果的优劣共同主导了最终日尺度模拟结果的精度,造成这一现象的原因可能是东南沿海地区的日降水占比与其他地区相比准确度较低。西北地区与东南沿海地区相比,呈现出较低的 R 值,并且全年都存在由东南沿海向西北地区递减的趋势。

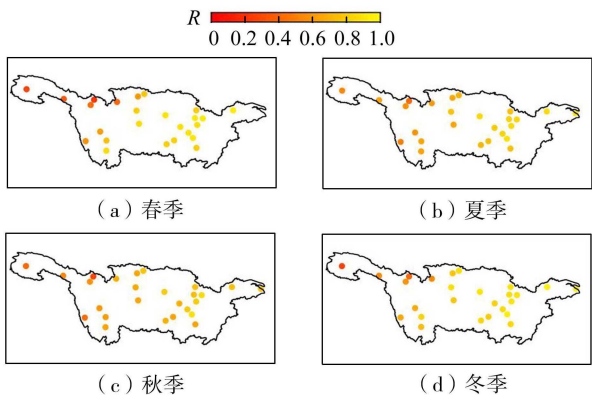


图7 FCM-RF 算法在验证站点不同季节模拟结果的相关系数空间分布

Fig. 7 Correlation coefficient spatial distribution of FCM-RF algorithm simulation results of different seasons at verification stations

图 8 为 CMORPH、PERSIANN、IMERG 原始卫星降水产品和 Ori-RF 算法、RF-OK 算法、FCM-RF 算法等不同方法 2009 年 6 月 29 日模拟降水量空间分布。在原始卫星降水产品中, CMORPH 可以捕捉到降雨中心, IMERG 对于降雨中心的捕捉能力略弱,但也可以大致捕捉到降水事件发生的范围,相较之下, PERSIANN 对于降水量的描述能力较弱。Ori-RF 算法可以捕捉到流域东北部降水事件发生的范

围,但对于降水量的预测能力弱,难以捕捉到降水中心的位置及实际降水。RF-OK 算法及 FCM-RF 算法在该日期都观测到了两个局部地区具有较大的降水量,FCM-RF 算法呈现出的降水中心位置的范围略大于 RF-OK 算法,且在青藏高原对于中雨事件的捕捉更为敏锐。对于日尺度降水的空间分布,FCM-RF 算法可以较完整地描述出降水事件的发生,并准确地捕捉到降水中心点位置,能够为一些无资料地区的降水勘测提供数据支持。

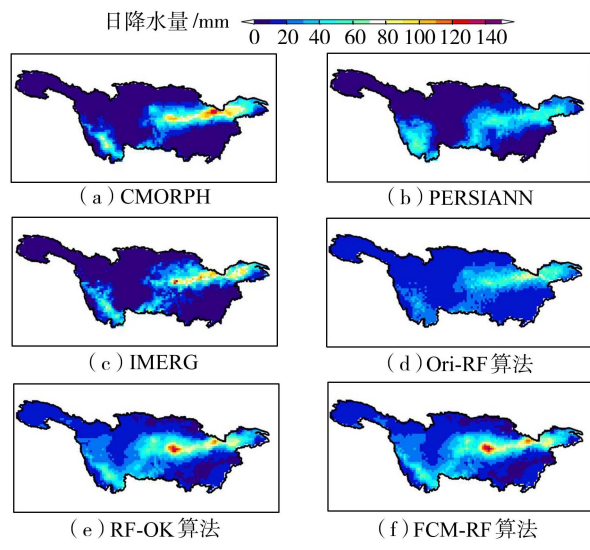


图 8 不同方法 2009 年 6 月 29 日模拟降水量空间分布
Fig. 8 Spatial distribution of simulated precipitation using different methods on June 29, 2009

4 结 论

a. 采用模糊 C 均值法对长江流域进行降水分区,共划分为 6 个区,分区内的站点具有较为相似的年均降水量,而分区结果也体现了我国因受东南季风影响所表现出来的降水量分布不均,并且从东南沿海向西北内陆递减的特点。

b. 验证站点的模拟结果表明,与 CMORPH、PERSIANN、IMERG 3 种原始卫星降水产品相比,多源融合降水产品在日尺度上的降水模拟精度都有较大的提升,相关系数可以达到 0.76,FCM-RF 算法模拟结果略优于 RF-OK 算法模拟结果。

c. 时间尺度上,FCM-RF 算法的模拟精度在长江流域降水每年呈现出相同的变化趋势,与实测降水有良好的相关性。夏季分区降水模拟结果与实测降水量之间有一定误差,表明在遭遇强降水事件时 FCM-RF 算法具有较大的不确定性;春秋降水模拟结果与实测降水量具有较高的相关系数,冬季相对较低。

d. 空间尺度上,FCM-RF 算法在东南沿海地区

的降水量预测更加准确,且全年精度都维持在较高水平,在青藏高原地区的降水捕捉能力较其他地区弱。

参考文献:

- [1] ZHANG W, FURTADO K, WU P, et al. Increasing precipitation variability on daily-to-multiyear time scales in a warmer world[J]. Science Advances, 2021, 7(31): eabf8021.
- [2] DONAT M G, LOWRY A L, ALEXANDER L V, et al. More extreme precipitation in the world's dry and wet regions [J]. Nature Climate Change, 2016, 6(5): 508-513.
- [3] LI Wentao, ZHANG Jiapeng, SUN Ruochen, et al. Evaluation of Tianji and ECMWF high-resolution precipitation forecasts for extreme rainfall event in Henan in July 2021[J]. Water Science and Engineering, 2023, 16(2): 122-131.
- [4] LIU L X, ZHANG K, CHAO L J, et al. Recent seasonal variations in ecosystem water use efficiency in China's key tropical-subtropical transitional zones in response to climate change [J]. Global Biogeochemical Cycles, 2022, 36(11): e2022GB007635.
- [5] KIDD C, HUFFMAN G. Global precipitation measurement [J]. Meteorological Applications, 2011, 18(3): 334-353.
- [6] CHAO L J, ZHANG K, LI Z J, et al. Geographically weighted regression based methods for merging satellite and gauge precipitation[J]. Journal of Hydrology, 2018, 558: 275-289.
- [7] HOU A Y, KAKAR R K, NEECK S, et al. The global precipitation measurement mission [J]. Bulletin of the American Meteorological Society, 2014, 95(5): 701-722.
- [8] JIANG Shanhu, WEI Linyong, REN Liliang, et al. Evaluation of IMERG, TMPA, ERA5, and CPC precipitation products over Chinese mainland: spatiotemporal patterns and extremes [J]. Water Science and Engineering, 2023, 16(1): 45-56.
- [9] 李艳忠, 庄稼成, 白鹏, 等. 中国不同气候区多源遥感降水融合与性能综合评估[J]. 地理研究, 2022, 41: 3335-3351. (LI Yanzhong, ZHUANG Jiacheng, BAI Peng, et al. Improving and comprehensive evaluation precipitation estimation by merging multi-satellite precipitation products in different climate regions of China [J]. Geographical Research, 2022, 41: 3335-3351. (in Chinese))
- [10] BECK H E, VAN DIJK A I J M, LEVIZZANI V, et al. MSWEP: 3-hourly 0.25° global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data [J]. Hydrology and Earth System Sciences, 2017, 21(1): 589-615.
- [11] MA Y, HONG Y, CHEN Y, et al. Performance of optimally merged multisatellite precipitation products using the dynamic bayesian model averaging scheme over the

- Tibetan Plateau [J]. Journal of Geophysical Research: Atmospheres, 2018, 123(2): 814-834.
- [12] FU Y, XIA J, YUAN W, et al. Assessment of multiple precipitation products over major river basins of China [J]. Theoretical and Applied Climatology, 2014, 123(1/2): 11-22.
- [13] 石羽佳, 王忠静, 索滢. 基于多源数据融合的海河流域降水资源评价[J]. 水科学进展, 2022, 33(4): 602-613. (SHI Yujia, WANG Zhongjing, SUO Ying. Evaluation of Haihe River basin precipitation resources based on multisource data fusion[J]. Advances in Water Science, 2022, 33(4): 602-613. (in Chinese))
- [14] SHEN Y, ZHAO P, PAN Y, et al. A high spatiotemporal gauge-satellite merged precipitation analysis over China [J]. Journal of Geophysical Research: Atmospheres, 2014, 119(6): 3063-3075.
- [15] LU X Y, TANG G Q, WANG X Q, et al. The development of a two-step merging and downscaling method for satellite precipitation products[J]. Remote Sensing, 2020, 12(3): 398.
- [16] XU S G, NIU Z, KUANG D, et al. Estimating summer precipitation over the Tibetan Plateau with geostatistics and remote sensing [J]. Mountain Research and Development, 2013, 33(4): 424-436.
- [17] GOOVAERTS P. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall [J]. Journal of Hydrology, 2000, 228(1/2): 113-129.
- [18] 沈嘉聚, 杨汉波, 刘志武, 等. 长江上游径流对气象要素变化的敏感性分析[J]. 水资源保护, 2023, 39(1): 119-126. (SHENG Jiaju, YANG Hanbo, LIU Zhiwu, et al. Sensitivity analysis of annual runoff to annual variation of meteorological elements in upper reaches of Yangtze River [J]. Water Resources Protection, 2023, 39(1): 119-126. (in Chinese))
- [19] WU H, YANG Q, LIU J, et al. A spatiotemporal deep fusion model for merging satellite and gauge precipitation in China [J]. Journal of Hydrology, 2020, 584: 124664.
- [20] BREIMAN L. Random forests [J]. Machine Learning, 2001, 45(1): 5-32.
- [21] HE X, CHANEY N W, SCHLEISS M, et al. Spatial downscaling of precipitation using adaptable random forests[J]. Water Resources Research, 2016, 52(10): 8217-8237.
- [22] HENGL T, NUSSBAUM M, WRIGHT M N, et al. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables[J]. PeerJ, 2018, 6: e5518.
- [23] 姜雨彤, 郝增超, 冯思芳, 等. 长江与黄河流域复合高温干旱事件时空演变特征[J]. 水资源保护, 2023, 39(2): 70-77. (JIANG Yutong, HAO Zengchao, FENG Sifang, et al. Spatiotemporal evolution characteristics in compound hot-dry events in Yangtze River and Yellow River basins [J]. Water Resources Protection, 2023, 39(2): 70-77. (in Chinese))
- [24] 郝增超, 张璇, 郝芳华, 等. 2022 年夏季长江流域复合高温干旱事件的影响及应对[J]. 水资源保护, 2023, 39(6): 46-52. (HAO Zengchao, ZHANG Xuan, HAO Fanghua, et al. Impacts and coping strategies of compound hot extremes and droughts during summer of 2022 in the Yangtze River Basin [J]. Water Resources Protection, 2023, 39(6): 46-52. (in Chinese))
- [25] 蒋艳, 贺新光, 章新平. 长江流域极端降水与当地平均气温的响应关系分析 [J]. 水资源与水工程学报, 2023, 34(1): 40-49. (JIANG Yan, HE Xinguang, ZHANG Xinping. Response of extreme precipitation to local average temperature in the Yangtze River Basin [J]. Journal of Water Resources and Water Engineering, 2023, 34(1): 40-49. (in Chinese))
- [26] 王小杰, 夏军强, 李启杰, 等. 长江中下游地区洪灾损失率函数构建及洪涝灾害损失评估 [J]. 水资源保护, 2023, 39(5): 88-98. (WANG Xiaojie, XIA Junqiang, LI Qijie, et al. Construction of flood loss rate function and flood loss assessment in middle and lower reaches of the Yangtze River [J]. Water Resources Protection, 2023, 39(5): 88-98. (in Chinese))
- [27] JOYCE R J, JANOWIAK J E, ARKIN P A, et al. CMORPH: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution [J]. Journal of Hydrometeorology, 2004, 5(3): 487-503.
- [28] HSU K L, GAO X, SOROOSHIAN S, et al. Precipitation estimation from remotely sensed information using artificial neural networks [J]. Journal of Applied Meteorology, 1997, 36(9): 1176-1190.
- [29] CHEN Y D, ZHANG Q, XIAO M Z, et al. Precipitation extremes in the Yangtze River Basin, China: regional frequency and spatial-temporal patterns [J]. Theoretical and Applied Climatology, 2014, 116(3/4): 447-461.
- [30] LI X, ZHANG K, NIU J F, et al. A machine learning-based dynamic ensemble selection algorithm for microwave retrieval of surface soil freeze/thaw: a case study across China [J]. Giscience & Remote Sensing, 2022, 59(1): 1550-1569.
- [31] 张珂, 牛杰帆, 李曦, 等. 洪水预报智能模型在中国半干旱半湿润区的应用对比 [J]. 水资源保护, 2021, 37(1): 28-35. (ZHANG Ke, NIU Jiefan, LI Xi, et al. Comparison of artificial intelligence flood forecasting models in China's semi-arid and semi-humid regions [J]. Water Resources Protection, 2021, 37(1): 28-35. (in Chinese))

(收稿日期: 2023-08-14 编辑: 王芳)