

DOI: 10.14042/j.cnki.32.1309.2022.05.008

考虑有雨无雨辨识的多源降水融合方法

李伶俐^{1,2}, 王银堂^{1,2}, 唐国强³, 高 轩⁴, 王磊之¹, 胡庆芳^{1,2}

(1. 南京水利科学研究院水文水资源与水利工程科学国家重点实验室, 江苏 南京 210029; 2. 长江保护与绿色发展研究院, 江苏 南京 210098; 3. University of Saskatchewan Coldwater Lab, Canmore, Alberta T1W 3G1, Canada; 4. 西北农林科技大学旱区农业水土工程教育部重点实验室, 陕西 杨凌 712100)

摘要: 多源降水融合是精准估计降水时空分布的重要途径, 多聚焦降水量或降水强度的误差订正, 对短历时降水雨区辨识的重视不足。提出考虑有雨无雨辨识的多源降水融合框架, 耦合地理加权逻辑回归与地理加权回归模型, 构建兼顾雨区辨识及雨量估计的降水融合方法, 并应用于汉江流域 MSWEP V2.1 与地面站网观测日降水融合。结果表明: 所提方法成功再现有雨无雨空间格局并刻画了降水中心, 整体强化了 MSWEP V2.1 对地面降水的表征能力, 降低误报率和误报降水量的幅度超过了 60%, 提高临界成功指数和 Kling-Gupta 效率系数达 40% 以上; 较降水空间插值数据, 削减误报降水量并提升 Kling-Gupta 效率系数高于 10%; 另外, 较参考数据, 降水融合改善强降水事件(雨强 ≥ 50 mm/d)分辨精度的增益不低于 60%。所提方法有效改善了降水估计效果, 为多源降水融合提供了新思路。

关键词: 多源降水融合; 有雨无雨辨识; 地理加权逻辑回归; 地理加权回归; 汉江流域
中图分类号: P339 **文献标志码:** A **文章编号:** 1001-6791(2022)05-0780-14

获取高精度降水时空分布信息对于提升水旱灾害防御能力、促进水土资源高效配置与利用等具有重要意义。然而, 受天气形势、地理位置、地形地貌及人类活动等因素的复合影响, 降水呈现复杂变异性^[1], 精准估计降水时空分布一直是气象水文研究的难点之一。

目前, 观测和估计降水的方式日趋多元, 主要有地面观测、气象卫星与天气雷达遥感反演、大气再分析等^[2]。地面观测可获得点尺度高精度信息, 但站网密度和空间分布模式影响了降水空间格局的刻画。遥感反演和大气再分析的栅格降水信息具有覆盖范围广、空间连续性强等优势, 但受传感器及反演算法的影响, 分类和定量误差相对突出, 并且在不同气候、地形及降水强度等条件下呈现复杂分异特征^[3-4]。集成地面观测、遥感反演和大气再分析及其他相关辅助信息(如位置、地形、植被、微波链路信号等)的多源信息融合方法, 已成为获取精细化降水时空分布的重要途径^[5-6]。目前, 降水融合方法已取得一系列进展, 根据不同信息源的地位, 可分为 3 类: 一是以遥感反演或大气再分析数据为主, 采用若干种栅格降水数据构建初估场, 然后利用地面站点观测降水在一定优化准则下修正初估场, 得到降水融合场, 如最优插值^[7]、地理加权回归^[8]等; 二是以地面站点观测信息为主, 以遥感、再分析和其他相关信息作为辅助变量, 采用协克里金插值^[9]、广义可加模型^[10]等融合多源信息, 也有贝叶斯滤波^[11]和条件融合^[12]等方法在对站点观测降水插值的基础上, 利用其他栅格降水修正插值场; 三是信息地位平等的多源配准集成, 核心是率定各源数据权重, 典型方法有贝叶斯加权平均^[13]、概率密度函数叠置^[14]、TC(triple collocation)^[15]和深度神经网络^[16]等。然而, 这些方法重点关注了累积降水量或降水强度的误差订正, 提高年、月降水估计精度的效果显著, 但对于空间非连续性特征显著的短历时降水(如日、小时尺度等), 融合过程中各种降水数据的有雨无雨状态信

收稿日期: 2022-02-28; 网络出版日期: 2022-08-29

网络出版地址: <https://kns.cnki.net/kcms/detail/32.1309.P.20220826.1534.002.html>

基金项目: 国家重点研发计划资助项目(2021YFC3000104); 国家自然科学基金资助项目(52009081)

作者简介: 李伶俐(1992—), 男, 山西吕梁人, 工程师, 主要从事水文水资源研究。E-mail: ljli@nhri.cn

通信作者: 王银堂, E-mail: ytwang@nhri.cn

息相互干扰,难以有效改善雨区空间位置和范围的辨识效果,从而影响降水频率分析,并增大枯水流量模拟预报的不确定性。近年来,学界针对该问题开展了有益的探索。如 Cornes 等^[17]以 0 和 1 分别表示有雨和无雨,采用薄板样条插值方法将站点观测降水的状态变量插值到各栅格,以降水概率阈值 0.5 判断每个栅格的有雨无雨状态,该方法虽有一定效果,但未考虑降水概率阈值在不同地理气候区、不同日期的动态变化;Zhang 等^[18]构建了具有有雨无雨状态辨识和降水量回归功能的双层机器学习模型,将多种卫星反演降水和站网插值降水进行融合,有效订正了卫星反演降水的分类误差;Xiao 等^[19]以 ERA5 数据集中植被、土壤水、气温、蒸散发等为解释变量,采用人工神经网络构建了有雨无雨辨识模型,并将其与分位数映射法耦合构建了融合模型,也显著提高了分类辨识精度。显然,考虑有雨无雨辨识有助于提高降水融合精度,但集分类辨识与定量估计于一体的融合方法仍存在较大探索空间,新方法研究对于推动降水融合进步具有积极意义。

本文提出考虑有雨无雨辨识的多源降水融合框架,采用地理加权逻辑回归与地理加权回归方法分别构建降水概率估计模型和降水融合模型,提出一种新型多源降水融合方法;应用该方法开展汉江流域遥感反演和地面观测日降水融合研究,综合总体精度、不同季节、不同子流域及不同强度的降水估计效果,验证所提方法的有效性。

1 研究方法

1.1 考虑有雨无雨辨识的多源降水融合框架

多源降水融合框架包括有雨无雨状态辨识、多源降水融合及降水融合结果修正 3 个环节:

(1) 有雨无雨状态辨识。提取地面站点观测数据中隐含的有雨无雨状态,构建降水概率估计模型,以地面站点状态的正确辨识率最高为目标优选降水概率阈值,由此辨识各栅格降水状态。

(2) 多源降水融合。构建集成地面观测、遥感反演和再分析等数据及辅助信息的多源降水融合模型,估计各栅格降水量。

(3) 降水融合结果修正。各栅格有雨无雨状态与降水量估计值对应相乘,将无雨区降水量修正为 0,得到各栅格的最终降水融合结果。

该框架内降水概率估计与降水融合模型的构建是核心,可选用克里金插值、地理加权回归、广义可加、贝叶斯最大熵、机器学习等模型,演化出不同融合方法,具有较强灵活性。

1.2 基于地理加权逻辑回归与地理加权回归的多源降水融合方法

地理加权回归(Geographically weighted regression, GWR)是一种典型的局部多元回归分析方法,通过计算邻近点空间权重和局部回归系数,综合考虑了因变量空间自相关和因变量与解释变量互相关性,可有效刻画因变量空间异质性^[20]。Chao 等^[21]和 Chen 等^[22]基于 GWR 构建了地面观测和遥感反演降水融合模型,取得了不错的降水估计效果。本文也应用该方法构建多源降水融合模型,采用集局部回归与逻辑回归于一体的地理加权逻辑回归(Geographically weighted logistic regression, GWLR)模型^[23]实现有雨无雨辨识,估计降水概率。由此,提出联合 GWLR 和 GWR 的降水融合方法,记为 M-GWLR-GWR(Precipitation merging method based on GWLR and GWR)。M-GWLR-GWR 属于第 1 类降水融合方法,具体融合思路如下:首先在站点位置上,构建集成地理信息、遥感反演降水、再分析降水的多源融合模型,即 GWLR 和 GWR;在此基础上,将模型超参数(即最优空间邻域带宽)移用至各栅格,建立各栅格降水概率估计与多源融合模型;最后结合有雨无雨状态修正降水融合结果。

(1) 数据预处理。确定降水估计的目标分辨率,协调站点与栅格降水的时间分辨率、不同栅格降水产品的空间分辨率。准备地面站点与待估计栅格的经纬度、高程及其他辅助信息。

(2) 栅格降水概率估计。选取判断有雨无雨状态的降水量阈值(T),若观测降水量大于等于 T ,则表示

有雨,记降水状态变量 $I_o = 1$,反之表示无雨,记 $I_o = 0$ 。以地面站点有雨无雨状态为因变量,在优选站点地理信息解释变量组合的基础上,在站点尺度构建基于 GWLR 的降水概率估计模型(式(1))^[23],优化确定最佳邻域带宽($q_{1\text{-best}}$)。迁移应用该带宽,推求各栅格降水概率估计模型参数,计算各站点和栅格的降水概率估计值。

$$\ln\left(\frac{p(i)}{1-p(i)}\right) = \beta_{i0} + \sum_{k=1}^{K_1} \beta_{ik}(x_i, y_i) G_k(i) \quad (1)$$

式中: $p(i)$ 为 i 点(站点或栅格,坐标为 x_i, y_i)处降水概率,当 i 点为站点时,结果记为 $p_{m\text{-gauge}}$,当 i 点为栅格时,结果记为 $p_{m\text{-grid}}$; β_{i0} 为常数项参数; $G_k(i)$ 为 i 点处第 k 个地理因子(坐标、高程、坡度、坡向等, $k = 1, 2, \dots, K_1$); $\beta_{ik}(x_i, y_i)$ 为相应回归参数。

(3) 栅格降水状态辨识。初始化降水概率阈值为 $C_{\text{pop}0}$,据式(2)判断地面站点降水状态($I_{m\text{-gauge}}$),以所有站点有雨无雨状态的正确辨识率(R)最高为目标函数,采用 SCE-UA 求解最优降水概率阈值($C_{\text{pop}\text{-best}}$),由此辨识各栅格有雨无雨状态。

$$I_{m\text{-gauge}} = \begin{cases} 1, & p_{m\text{-gauge}} \geq C_{\text{pop}\text{-best}} \\ 0, & p_{m\text{-gauge}} < C_{\text{pop}\text{-best}} \end{cases} \quad (2)$$

$$R = \frac{n_0 + n_1}{n} \quad (3)$$

式中: $I_{m\text{-gauge}}$ 为基于降水概率估计模型与降水概率阈值判断的站点有雨无雨状态; n_0 为无雨站点的正确辨识数量; n_1 为有雨站点的正确辨识数量; n 为研究区地面观测站点总数; R 为正确辨识率评分。

(4) 多源降水信息融合。利用地面观测、遥感反演和再分析产品在站点位置的观测或估计降水量及相关辅助信息,在站点位置构建基于 GWR 的多源信息融合模型(式(4)),优化最佳邻域带宽($q_{2\text{-best}}$),迁移应用该带宽推求各栅格融合模型参数,计算各栅格降水融合结果。

$$P_m(i) = \gamma_{i0} + \sum_{k=1}^{K_2} \gamma_{ik}(x_i, y_i) W_k(i) + \sum_{l=1}^L \gamma_{il}(x_i, y_i) S_l(i) \quad (4)$$

式中: $P_m(i)$ 为 i 点(坐标为 x_i, y_i)处降水融合结果, mm,当 i 点为站点时,融合结果记为 $P_{m\text{-gauge}}$,当 i 点为栅格时,融合结果记为 $P_{m\text{-grid}}$; γ_{i0} 为常数项参数; $W_k(i)$ 为 i 点处第 k 个地理因子($k = 1, 2, \dots, K_2$); $\gamma_{ik}(x_i, y_i)$ 为相应回归参数; $S_l(i)$ 为第 l 种栅格降水产品(卫星反演、再分析等, $l = 1, 2, \dots, L$)在 i 点处的降水量估计值, mm; $\gamma_{il}(x_i, y_i)$ 为相应回归参数。

(5) 降水融合结果修正。针对各栅格,采用 $P_{m\text{-cor}} = I_{m\text{-grid}} P_{m\text{-grid}}$ 修正降水融合结果,检测是否存在 $P_{m\text{-cor}}$ 与 $I_{m\text{-grid}}$ 不一致的栅格。若存在,针对不一致栅格,搜索满足“ $I_{m\text{-grid}} = 1$ 且 $P_{m\text{-cor}} \geq T$ ”或“ $I_{m\text{-grid}} = 0$ 且 $P_{m\text{-cor}} < T$ ”条件的最邻近 M 个(一般取 4 或 9)栅格,采用反距离加权法计算 $P_{m\text{-cor}}$,即为最终降水融合结果。

针对研究时期内每一个时段,重复上述步骤(2)–(5),实现降水概率阈值及最佳邻域带宽的动态估计,得到研究区各时段栅格降水量空间分布估计结果。M-GWLR-GWR 方法适用于集成遥感反演、再分析与地面观测降水及相关辅助地理信息的降水空间分布精准估计。需要说明的是,该方法并不局限于液态降水。由于近年来遥感反演算法、再分析数值模式的进步,相关数据产品已经包含降水相态分类(液态和固态)^[24-25],提供了降雪定量估计结果,本文方法也是多源降雪信息融合的一种新途径。

2 研究区试验方案

2.1 研究区与数据

汉江是长江中游左岸的最大支流,干流流经陕西、湖北两省,全长 1 577 km,流域范围介于 106°15'E—114°20'E、30°10'N—34°20'N 之间,总面积约 15.9 万 km²(图 1)。地势复杂,呈西高东低态势。流域地处东

亚副热带季风气候区, 多年平均降水量约 900 mm, 时空分布不均, 年降水量呈南岸大于北岸, 上游、下游大, 中游小的分布特征, 且全年 70% ~ 80% 的雨量集中于汛期(5—10 月)。结合地理气候特点, 全流域分为丹江口以上(Ⅰ)、唐白河(Ⅱ)和丹江口以下(Ⅲ)3 个子流域。

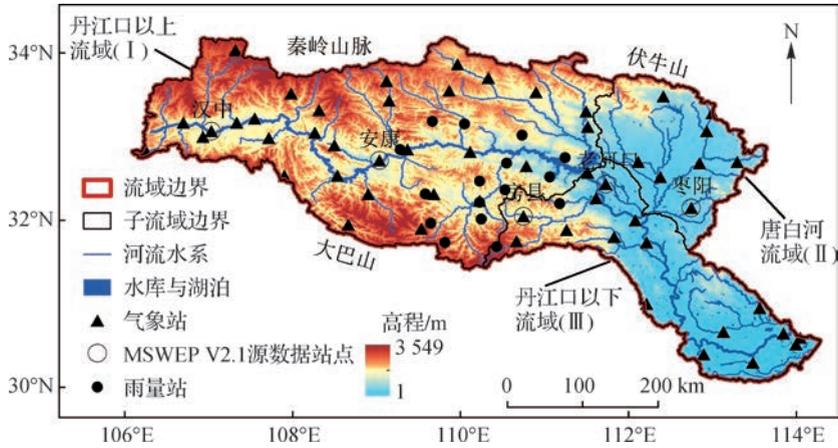


图1 汉江流域地形、水系及降水地面观测站

Fig. 1 Topography, river, and precipitation gauges locations over the Han River basin

所用栅格降水为 MSWEP (Multi-Source Weighted-Ensemble Precipitation Version), 是美国普林斯顿大学采用加权算法集成 4 种未经校正的卫星反演降水 (TMPA 3B42RT、CMORPH、GSMaP、GridSat)、2 种再分析降水 (EAR-interim、JRA-55) 和全球约 77 000 个雨量计观测降水形成的首个覆盖全球、横跨 1979—2021 年的高分辨率 ($0.1^\circ \times 0.1^\circ$, 3 h) 降水数据集, 详细研制算法见文献 [26]。降水数据为 MSWEP V2.1 的 3 h 数据 (<http://www.gloh2o.org/>), 时间范围为 2010—2016 年。地面站网观测降水数据为从国家气象科学数据中心和长江流域水文年鉴收集的 56 个气象站与 16 个雨量站 2010—2016 年逐日降水量。经分析, MSWEP V2.1 已集成了气象站网中安康、汉中、老哈河、房县和枣阳 5 站观测降水, 为确保 MSWEP V2.1 与地面观测信息相互独立, 地面观测数据中该 5 站数据予以去除。因此, 全流域共 67 个地面站点, 其中, 丹江口以上流域 (Ⅰ) 43 个、唐白河流域 (Ⅱ) 7 个、丹江口以下流域 (Ⅲ) 17 个。

2.2 降水融合试验方案

降水融合试验的空间分辨率取 $0.1^\circ \times 0.1^\circ$, 时间分辨率为日, 时间范围为 2010—2016 年。融合源数据为 MSWEP V2.1 和地面站点观测降水, 前者采用 UTC 时间, 沿时程累加得到日降水, 其对应北京时间为 08:00 至翌日 08:00, 将气象站北京时间 08:00—20:00、20:00—08:00 两时段降水求和得到日降水, 从而与雨量站观测日降水统计时段保持一致。采用 K-Means 聚类方法将地面站点划分为空间均匀分布的 4 组试验站网, 采用 4 折交叉验证的方式开展降水融合试验 (图 2)。

以传统基于 GWR 的降水融合方法 (记为 M-GWR) 作为对比, 阐明考虑有雨无雨辨识对降水融合精度的影响。另外, 还采用 GWLR-GWR 方法, 仅利用地理信息, 开展降水空间插值, 记作 I-GWLR-GWR, 与 M-GWLR-GWR 对比分析引入 MSWEP V2.1 的作用。所有降水融合与空间插值方法的地理因子均考虑经纬度坐标 (XY)、经纬度坐标 + 高程 (XYH) 2 种情形, 阐明高程对降水估计精度的影响; 经分析 2 种解释变量组合的方差膨胀因子较低, 共线性问题对模型参数估计的影响较小。本文 GWLR 与 GWR 模型采用自适应带宽, 权函数取 bi-square, 以留一交叉验证均方误差最小为目标^[20], 采用黄金分割法优选空间邻域带宽。降水融合模型构建通过编写 Matlab 代码实现。

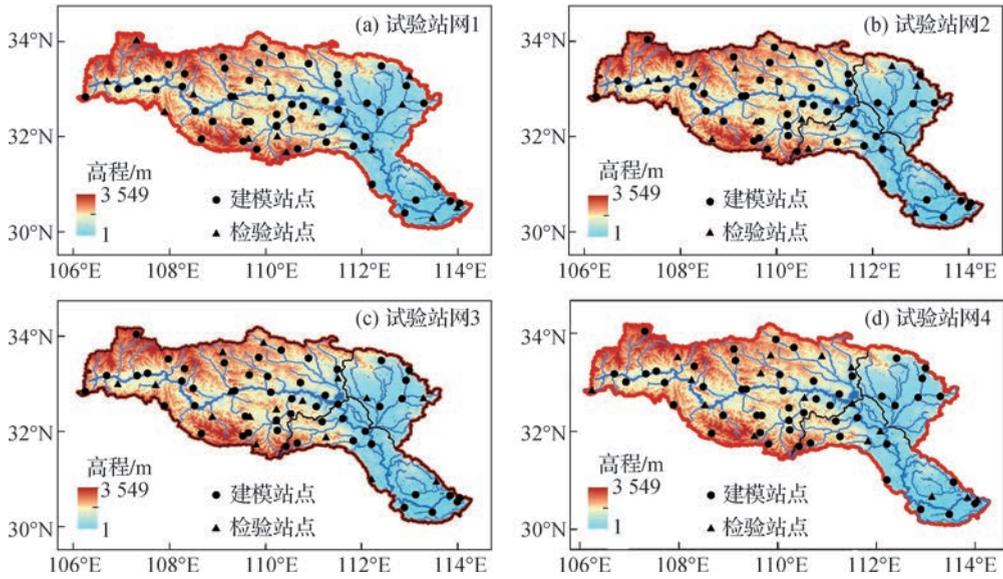


图 2 降水融合试验的地面观测站网分布

Fig. 2 Distribution of gauges networks for precipitation fusion experiment over the Han River basin

2.3 降水融合精度及增益评价方法

本文在站点位置上以观测降水为真值，采用分类和连续性 2 类指标评价降水融合精度。分类指标包括探测率 (Probability of detection, D_{PO})、误报率 (False Alarm Ratio, R_{FA}) 和临界成功指数 (Critical Success Index, I_{CS})，有雨无雨阈值取 $T = 0.1 \text{ mm/d}$ 。连续性指标包括绝对平均误差 (Mean absolute error, E_{MA}) 和相关系数 (Correlation Coefficient, C_C)。此外，还采用综合反映融合数据与观测数据的偏差、离散程度差异和相关系数的 Kling-Gupta 效率评价，其中偏差用均值比 (α) 表示，离散程度差异用变差系数比 (β) 表示，确保 α 与 β 独立 (式(5)一式(7))^[27]，Kling-Gupta 效率系数用 E_{KC} 表示。此外，为重点分析考虑有雨无雨辨识的影响，根据降水状态分类结果，将总误差 (Total Bias, B_T) 分解为击中误差 (Hit Bias, B_H)、漏报降水量 (Missed Precipitation, P_M) 和误报降水量 (False Precipitation, P_F) (式(8)一式(9))，评价不同性质误差的变化。

$$E_{KC} = 1 - \sqrt{(1 - C_C)^2 + (1 - \alpha)^2 + (1 + \beta)^2} \quad (5)$$

$$\alpha = \mu_m / \mu_o \quad (6)$$

$$\beta = (\sigma_m / \mu_m) / (\sigma_o / \mu_o) \quad (7)$$

$$B_T = B_H + P_M + P_F \quad (8)$$

$$B_H = \sum_{i=1}^H P_{m-gauge,i} - P_{o,i} \quad P_M = \sum_{i=1}^M -P_{o,i} \quad P_F = \sum_{i=1}^F P_{m-gauge,i} \quad (9)$$

式中： $P_{m-gauge,i}$ 、 $P_{o,i}$ 分别为第 i 个站点位置处降水融合数据和观测降水； μ_m 、 σ_m 分别为降水融合数据的均值和标准差； μ_o 、 σ_o 分别为观测降水的均值和标准差； H 为击中事件 (估计有雨且实测有雨) 的数量； B_H 为击中事件的累积误差； M 为漏报事件 (估计无雨而实测有雨) 的数量； P_M 为漏报事件的累积降水量； F 为误报事件 (估计有雨而实测无雨) 的数量； P_F 为误报事件的累积降水量。

评价融合增益是衡量降水融合方法有效性的重要内容。本文以降水融合精度较参考数据精度的相对改善程度作为融合增益，参考数据包括 MSWEP V2.1 和降水空间插值数据 (插值与融合方法对应)。以前者为参考，分析降水融合改善遥感反演降水精度的效果；以后者为参考，研究遥感反演降水捕捉空间分布格局方面的优势以及对于提高降水估计精度的作用。本文将精度评价指标按取值大小与表征精度是否更高的对应关系，划分为正向型 (指标值越大，精度越高)、逆向型 (指标值越小，精度越高) 和中间值最优型 (指标值越靠

近中间最优值, 精度越高)3 类指标, 提出针对各类指标融合增益评价的通用公式, 见表 1。对于中间最优值型指标, a 和 β 的最优值为 1, B_T 和 B_H 最优值为 0。

表 1 多源降水融合增益评价的通用公式

Table 1 General formulas for gains evaluation of multi-source precipitation fusion

类型	融合增益公式	适用指标
正向型	$\Delta S = \frac{S(P_m) - S(R_e)}{ S(R_e) } \times 100$	D_{PO} 、 I_{CS} 、 C_C 、 E_{KG} 、 P_M
逆向型	$\Delta S = \frac{-[S(P_m) - S(R_e)]}{ S(R_e) } \times 100$	R_{FA} 、 P_F 、 E_{MA}
中间值最优型	$\Delta S = -\frac{[S(P_m) - S_{best} - S(R_e) - S_{best}]}{ S(R_e) - S_{best} } \times 100$	a 、 β 、 B_T 、 B_H

注: P_m 表示降水融合数据; R_e 为参考降水; S 为评价指标; S_{best} 为指标最优值。

3 结果与讨论

本文涉及地面站点观测、MSWEP V2.1、M-GWR(XY)、M-GWR(XYH)、I-GWLR-GWR(XY)、I-GWLR-GWR(XYH)、M-GWLR-GWR(XY)和M-GWLR-GWR(XYH)共 8 种方式估计汉江流域降水空间分布。通过降水空间分布估计结果对比、本文方法与传统融合方法对比、本文方法融合结果与参考降水对比, 综合阐明所提方法的有效性。

3.1 降水空间分布估计结果

图 3 和图 4 分别给出了 8 种方式估计汉江流域日平均降水量和典型日期(2016 年 9 月 25 日)降水量的空间分布图。由图 3 可知, 汉江流域日平均降水量呈现东南、南部和西部高, 东北部低的总体特征。MSWEP V2.1(图 3(b))低估了东南部和西部降水; 传统融合方法(图 3(c)和图 3(d))较好刻画了降水量高值区空间格局, 但略微高估了东北部低值区降水量; M-GWLR-GWR 方法(图 3(f)和图 3(h))则较好呈现了高值区与低值区降水空间分布, 与空间插值方法(图 3(e)和图 3(g))相比, 改善了低估东南部和南部降水的不足。由图 4 可知, 该日降水主要发生在流域中部, 降水中心位于唐白河流域东部和丹江口以上流域南部。MSWEP V2.1(图 4(b))仅捕捉了唐白河流域降水中心; 传统融合方法(图 4(c)和图 4(d))虽能捕捉 2 个中心, 但在西北和东南部存在大量误报现象; 空间插值方法(图 4(e)和图 4(g))较好描述了雨区范围, 但明显低估了降水中心雨量; 而本文方法(图 4(f)和图 4(h))则兼顾了有雨无雨格局与降水中心, 较好刻画了日降水量空间分布。

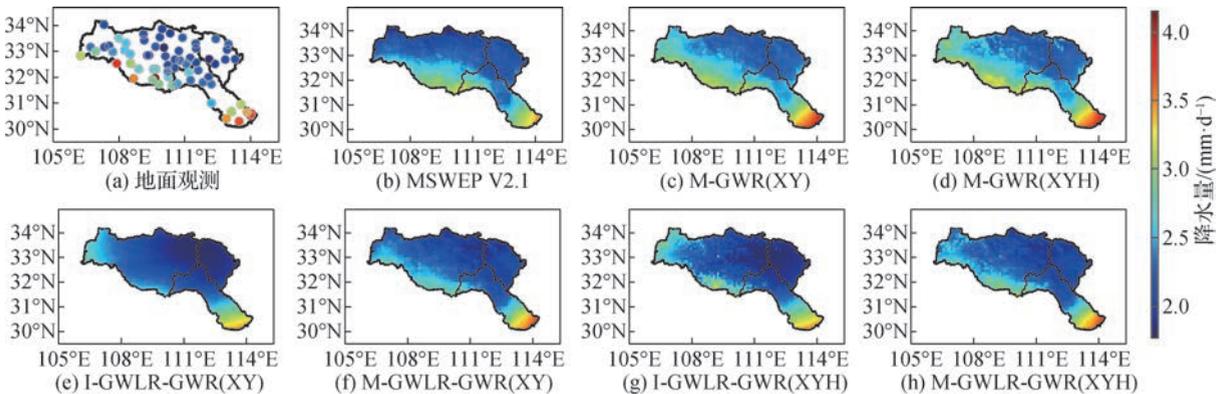


图 3 汉江流域日平均降水量空间分布

Fig. 3 Spatial distribution of mean daily precipitation over the Han River basin

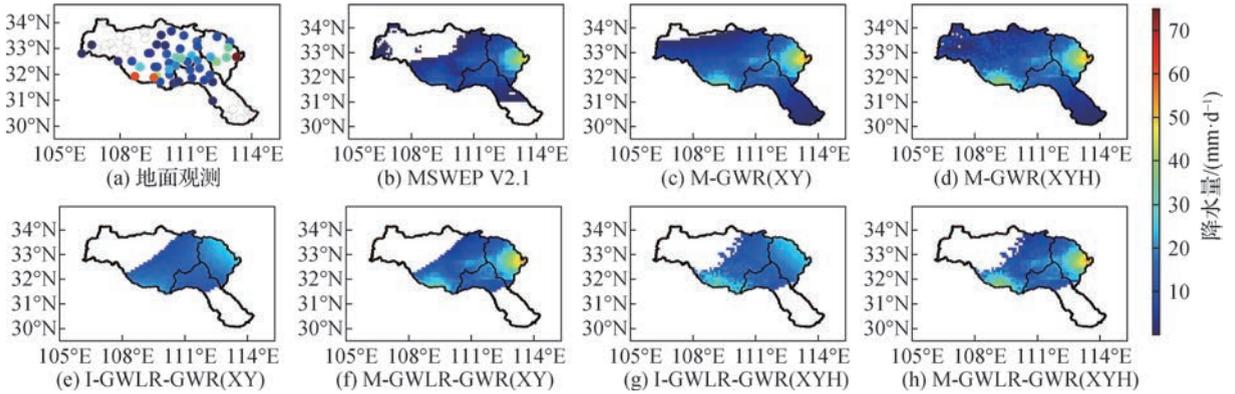


图4 汉江流域2016年9月25日降水量空间分布

Fig. 4 Spatial distribution of precipitation over the Han River basin on September 25, 2016

3.2 与传统降水融合方法的比较

从总体精度、不同季节、不同子流域及不同强度日降水估计精度4个方面,对比M-GWLR-GWR与传统M-GWR方法的优劣。受日降水随机性强及空间变异性复杂的影响,常出现部分站点降水估计精度为离群值,为减弱其对平均估计效果分析的影响,以中位数代替算数平均值。表2给出了不同融合方法估计汉江流域各站点日降水精度的中位数。由表可知,在分类精度方面,较传统方法,M-GWLR-GWR的探测率有所减小,误报率降低至0.1附近, I_{CS} 提升至近0.7。相应地,误报降水量得到了显著压缩,击中误差也有所减小,但漏报降水量绝对值明显增大,使总误差由高估转为低估,且绝对值也大幅增加。在连续性精度方面,M-GWLR-GWR的 C_C 略小于M-GWR,但前者估计日降水序列的变差系数更接近地面观测, E_{KC} 也小幅升高。对于2种融合方法,在解释变量中引入高程因子对各指标的影响总体较小,可能原因是日尺度降水的强随机性影响了降水—高程相关关系的稳健性。

表2 M-GWLR-GWR与M-GWR方法估计汉江流域各站点日降水精度的中位数

Table 2 Medians of metrics for daily precipitation estimates derived by M-GWLR-GWR and M-GWR at gauges over the Han River basin

方法	D_{P_0}	R_{FA}	I_{CS}	$B_T/$ mm	$B_H/$ mm	$P_M/$ mm	$P_F/$ mm	C_C	α	β	E_{KC}
M-GWR(XY)	0.908	0.396	0.568	239.4	-152.8	-69.1	467.2	0.755	1.056	0.753	0.622
M-GWR(XYH)	0.910	0.401	0.568	264.7	-106.4	-69.9	461.3	0.753	1.060	0.757	0.627
M-GWLR-GWR(XY)	0.762	0.117	0.687	-462.8	-111.5	-486.2	119.7	0.728	0.875	0.874	0.659
M-GWLR-GWR(XYH)	0.772	0.125	0.698	-456.9	-93.7	-431.9	126.3	0.732	0.884	0.863	0.668

图5对比了2种融合方法对不同季节日降水的估计精度。在分类辨识方面(图5(a)—图5(c)),与M-GWR相比,M-GWLR-GWR显著降低 R_{FA} 的同时,缩小了四季的差异,虽引起 D_{P_0} 中位数下降,但各季节 I_{CS} 均有所升高。这是由于地面站点分布稀疏不均,依据地面观测降水状态识别的有雨无雨分界线与真实边界仍存在偏差,使漏报区域增大,导致较传统方法减小了探测率。但结合3个分类精度指标的数量关系^[28],可知本文方法改善 R_{FA} 的正面效益超过了对 D_{P_0} 的负面影响。从季节分布看, I_{CS} 呈现春秋高于夏冬的特征,M-GWLR-GWR使分类指标的冬季变幅趋于扩大,即增加了空间差异性。在不同误差组分方面(图5(d)—图5(f)),M-GWLR-GWR降低了四季的 P_F ,增大了 P_M 绝对值, B_H 变化不大;受夏季降水量较大的影响, P_F 中位数和变幅的改善效果优于其他季节,而 P_M 则相反。从连续性精度来看(图5(g)—图5(i)),M-GWR与M-GWLR-GWR方法的 C_C 相差不大,但后者在四季均倾向于降低 E_{MA} , E_{KC} 均得到了小幅提升。

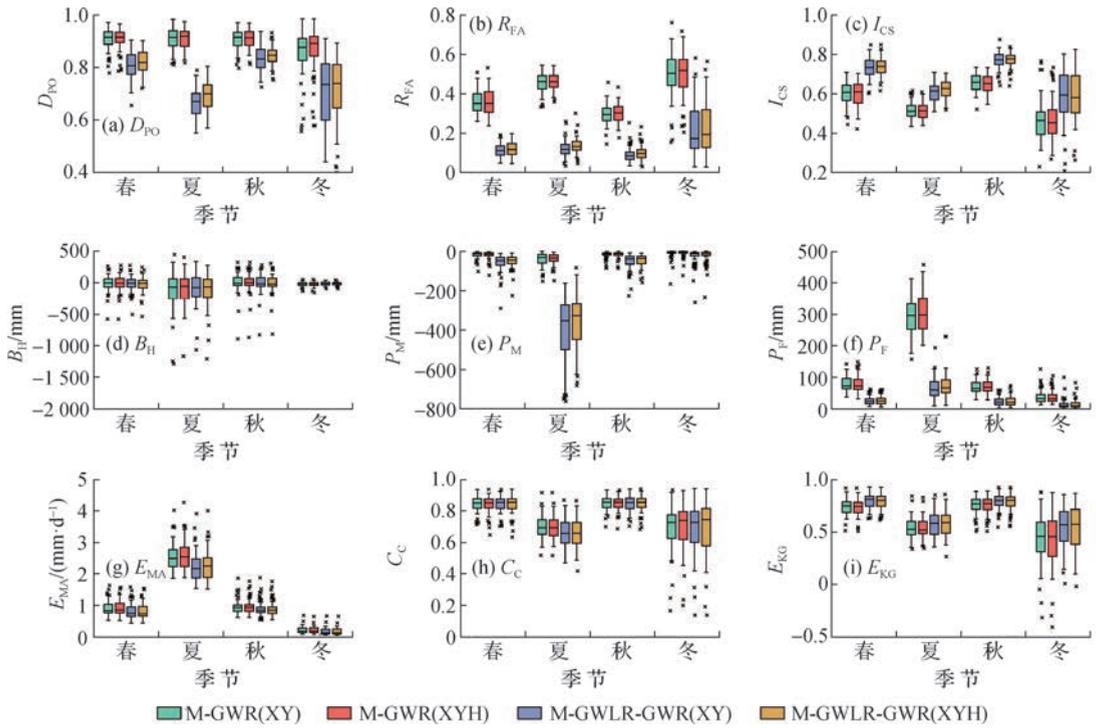


图 5 M-GWLR-GWR 与 M-GWR 方法估计不同季节日降水的精度指标箱线图

Fig. 5 Boxplots of metrics of M-GWLR-GWR and M-GWR for estimating daily precipitation in different seasons

图 6 对比了 2 种融合方法对不同子流域站点日降水的估计精度。由图可知, 相比于 M-GWR, M-GWLR-

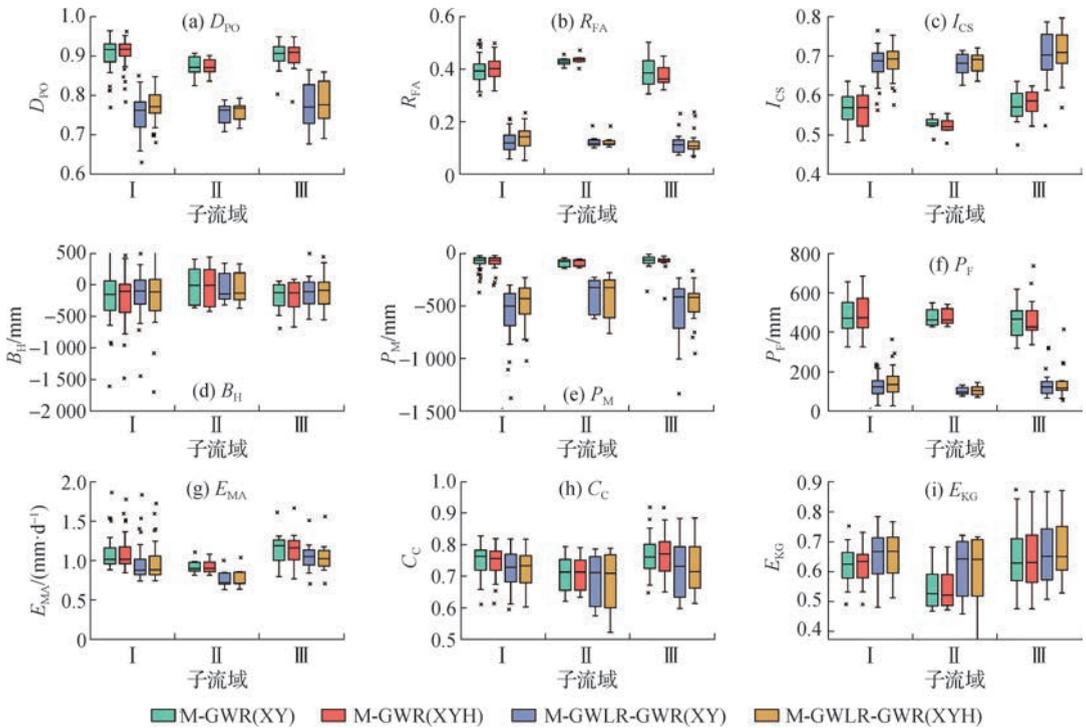


图 6 M-GWLR-GWR 与 M-GWR 方法对不同子流域站点日降水的估计精度指标箱线图

Fig. 6 Boxplots of metrics of M-GWLR-GWR and M-GWR for estimating daily precipitation at gauges over sub-basins

GWR 使各子流域的 R_{FA} 均得到了大幅削减,有效改善了 I_{CS} , 指标变幅在子流域间无显著差异。与分类指标对应,子流域的 P_F 中位数及变幅也有明显压缩,而 P_M 中则明显增大, B_H 变化相对较小。对于连续性精度,M-GWLR-GWR 小幅削减了 E_{MA} , 虽然 C_C 略有恶化,但各子流域 E_{KG} 中位数较 M-GWR 均有所增大,其中唐白河流域的改善效益相对突出,但站点差异性趋于扩大。

图7进一步对比了2种融合方法对不同日降水强度(Precipitation intensities, I_p)的估计精度。由图可知,随雨强增大,M-GWLR-GWR 较 M-GWR 降低 D_{PO} 、压缩 R_{FA} 、增大 I_{CS} 的效果趋于减小;M-GWR 的 R_{FA} 随雨强增大而先减后增,而 M-GWLR-GWR 显著减少了小雨误报事件,使 R_{FA} 呈现随雨强增大而增加的趋势。对于不同误差组分,在降水强度和分类精度的综合影响下,M-GWLR-GWR 较 M-GWR 削减 P_F 和增大 P_M 的幅度随雨强增大而减小,而 B_H 绝对值的减小幅度先增后减; P_M 对雨强变化的响应规律发生了改变,M-GWR 的 P_M 绝对值先增后减,而 M-GWLR-GWR 则随雨强增大而减小,这与小雨情形下 D_{PO} 有所降低相对应。

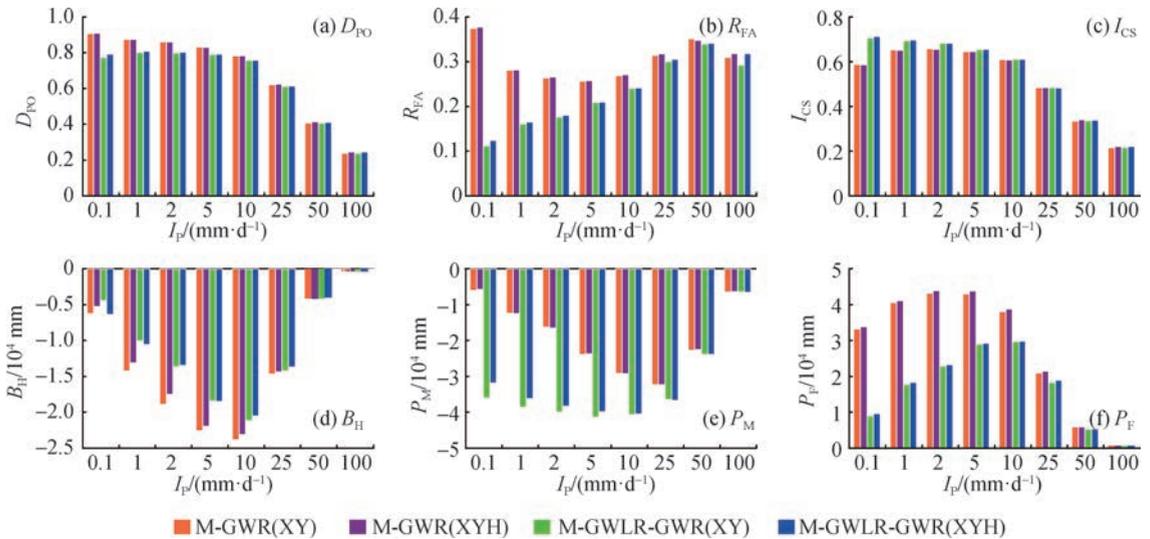


图7 M-GWLR-GWR 与 M-GWR 方法对不同强度日降水的估计精度

Fig.7 Performance of daily precipitation estimation with different intensities derived from M-GWLR-GWR and M-GWR

3.3 较参考数据的降水融合精度增益

表3给出了M-GWLR-GWR方法估计汉江流域各站点日降水精度增益的中位数。由表可知,所提方法有效改善了MSWEP V2.1精度,各项指标增益全部超过了10%,特别是 R_{FA} 和 P_F 增益超过了60%,综合精度 I_{CS} 和 E_{KG} 改善幅度分别高于40%和100%。由于插值降水精度总体高于MSWEP V2.1,以插值降水为参考的降水融合精度增益总体小于以MSWEP V2.1为参考的分析结果,降水融合对分类精度的影响十分微弱,订正 B_T 、 B_H 和 P_F 的效益均高于15%, C_C 、 α 、 β 和 E_{KG} 也得到了不同程度的改善,特别是综合精度 E_{KG} 提高了10%以上,高程因素的影响相对较弱。

图8给出了M-GWLR-GWR方法估计不同季节日降水的精度增益箱线图。以MSWEP V2.1为参考,所提方法使 ΔR_{FA} 和 ΔP_F 中位数分别高于40%和70%,冬季略小于其他季节;对于 D_{PO} 和 P_M ,除夏季有负面影响外,在春、秋、冬三季,2个指标改善效益的中位数分别不低于10%和60%;四季 ΔI_{CS} 中位数均在25%以上,夏季略低于其他季节。连续性精度也得到了有效改善, ΔE_{MA} 与 ΔC_C 中位数全部高于40%,而对于 E_{KG} ,则各季节均有半数以上站点的增幅超过了1倍,总体上连续性精度增益的季节性特征不显著。以插值降水为参考, ΔR_{FA} 中位数介于2%~7%, ΔP_F 中位数不低于10%,且冬季最小; ΔB_H 中位数均大于0,并在夏季达到20%以上; ΔD_{PO} 、 ΔP_M 和 ΔI_{CS} 中位数接近于0。连续性精度方面,春、夏、秋三

季的 ΔE_{MA} 箱体整体高于 0, 在冬季也有超过 50% 站点的误差得到订正, ΔC_C 和 ΔE_{KC} 中位数分别介于 2% ~ 8% 与 2% ~ 15%, 且夏季高于其他季节。引入高程除对指标增益的变幅略有影响外, 其中位数与不考虑高程的情形基本一致。

表 3 M-GWLR-GWR 方法估计汉江流域各站点日降水精度增益的中位数

Table 3 Medians of metrics gains for daily precipitation estimates derived by M-GWLR-GWR at gauges over the Han River basin

对比组	ΔD_{PO}	ΔR_{FA}	ΔI_{CS}	ΔB_T	ΔB_H	ΔP_M	ΔP_F	ΔC_C	$\Delta \alpha$	$\Delta \beta$	ΔE_{KC}
①	11.2	71.3	42.2	54.7	74.7	21.3	87.7	65.2	42.2	47.4	107.4
②	-0.2	3.6	0.2	15.3	26.9	-0.2	25.6	5.9	12.7	30.5	11.9
③	11.3	69.6	44.7	58.8	73.6	37.0	88.1	68.0	47.1	43.7	120.9
④	-0.3	4.3	0.2	16.7	23.1	-0.4	26.7	5.7	15.6	29.0	10.4

注: ① M-GWLR-GWR(XY)与 MSWEP V2.1; ② M-GWLR-GWR(XY)与 I-GWLR-GWR(XY); ③ M-GWLR-GWR(XYH)与 MSWEP V2.1; ④ M-GWLR-GWR(XYH)与 I-GWLR-GWR(XYH)。

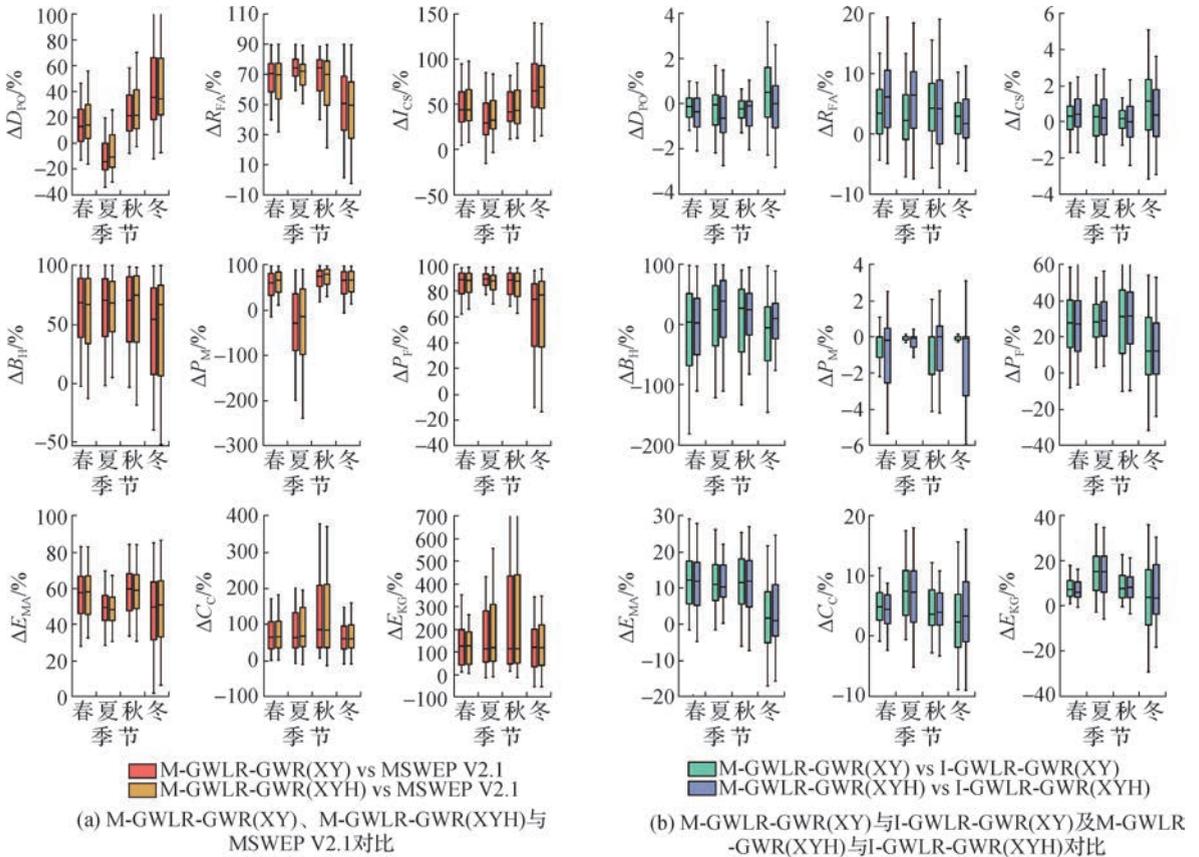


图 8 M-GWLR-GWR 方法估计不同季节日降水的精度增益箱线图

Fig. 8 Boxplots of metrics gains of M-GWLR-GWR for estimating daily precipitation in different seasons

图 9 为 M-GWLR-GWR 方法估计不同子流域站点日降水的精度增益箱线图。较 MSWEP V2.1, 各子流域 ΔR_{FA} 、 ΔI_{CS} 和 ΔP_F 的箱体整体高于 0, 并且中位数分别超过了 60%、30% 和 80%; ΔB_H 中位数高于 60%, 但考虑高程使唐白河流域部分站点 ΔB_H 小于 0; D_{PO} 和 P_M 增益相对较小, 并且 3 个子流域均存在部分站点的精度有所恶化。就连续性精度而言, ΔE_{MA} 、 ΔC_C 和 ΔE_{KC} 的中位数分别超过了 50%、40% 和 90%, 特别是丹江

口以下流域约 50% 站点的综合精度 E_{KC} 得到了巨幅提升。总体上, 唐白河流域 D_{PO} 、 B_H 、 P_M 精度增益低于另外 2 个子流域, 而 R_{FA} 、 I_{CS} 、 E_{MA} 精度增益恰好相反, 这主要与参考数据 MSWEP V2.1 的相关精度指标在不同子流域的空间分异特征有关。较插值降水, 降水融合数据的 P_F 、 B_H 、 E_{MA} 和 E_{KC} 改善效益相对较高, 中位数分别超过了 20%、7%、8% 和 10%, ΔR_{FA} 、 ΔD_{PO} 、 ΔI_{CS} 、 ΔP_M 和 ΔC_C 相对较小; 总体上各指标增益在不同子流域间无明显差别, 但受地势平坦、降水与高程相关性较弱等影响, 引入高程因子使唐白河流域部分站点 B_H 大幅增加, ΔB_H 中位数由 33% 降至 7%。

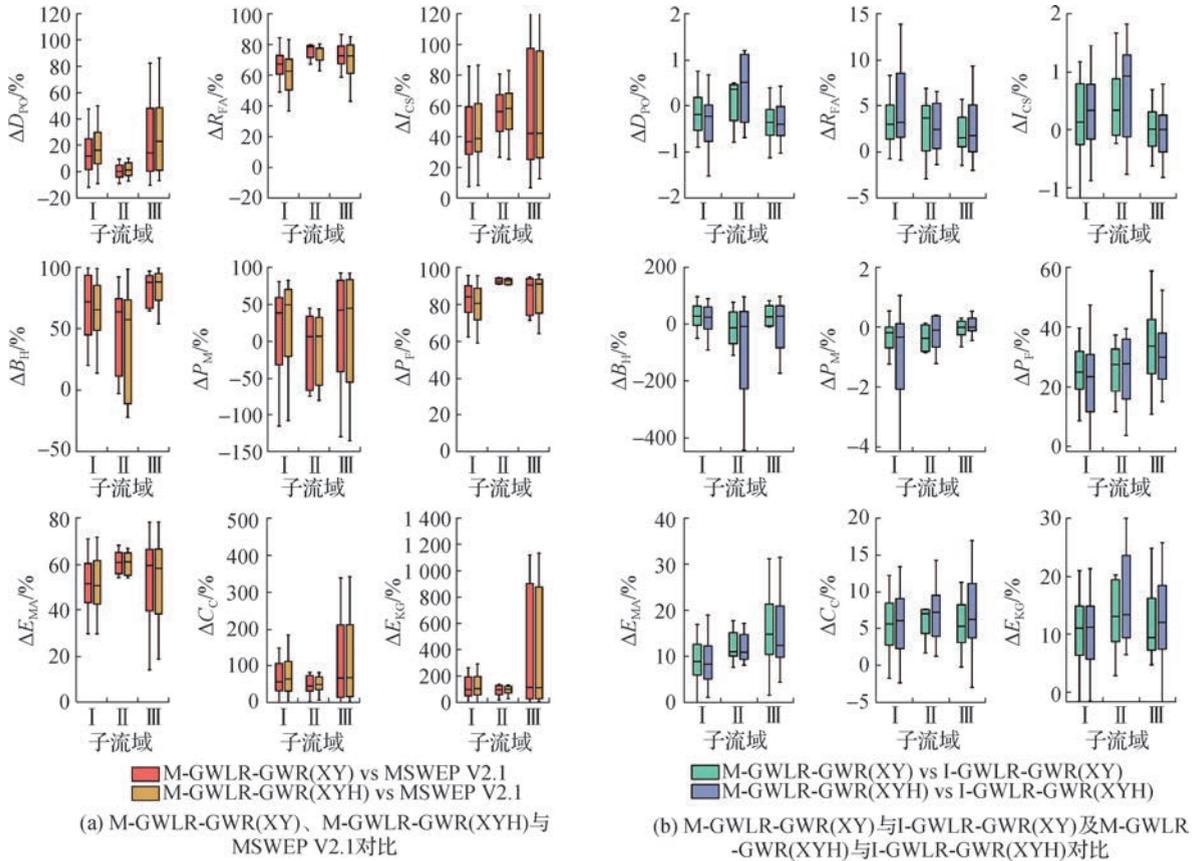


图9 M-GWLR-GWR方法估计不同子流域站点日降水的精度增益箱线图

Fig. 9 Boxplots of metrics gains of M-GWLR-GWR for estimating daily precipitation at gauges over sub-basins

图 10 分析了 M-GWLR-GWR 方法估计不同强度日降水的精度增益。以 MSWEP V2.1 为参考, 除 B_H 外, 各指标在不同雨强条件下均有所改善, ΔD_{PO} 和 ΔI_{CS} 随雨强增大而升高, ΔR_{FA} 和 ΔP_F 则先减后增。值得注意的是, 本文方法使强降水事件 ($I_p \geq 50$ mm/d) 的综合分类精度 I_{CS} 提高了 3 倍, 大幅降低了 P_M 和 P_F , 但增大了 B_H , 这与漏报降水事件转为击中降水事件有关。对于插值降水, 除 P_F 外, 各指标增益总体大于 0, ΔD_{PO} 、 ΔI_{CS} 和 ΔP_M 随雨强增大而增加, ΔB_H 、 ΔP_F 则相反; 同样地, 强降水事件捕捉能力 I_{CS} 提升幅度超过 60%, 但 P_F 有所增大。这主要是由于遥感反演产品通常低估强降水, 与地面观测融合总体使强降水区域的降水估计值增大, 从而减少了漏报事件数量、增加了击中事件数量, 但真实强降水区域边界处雨量站分布不均可能使降水融合数据的强降水区略大, 从而增加了误报事件数量和误报降水量。

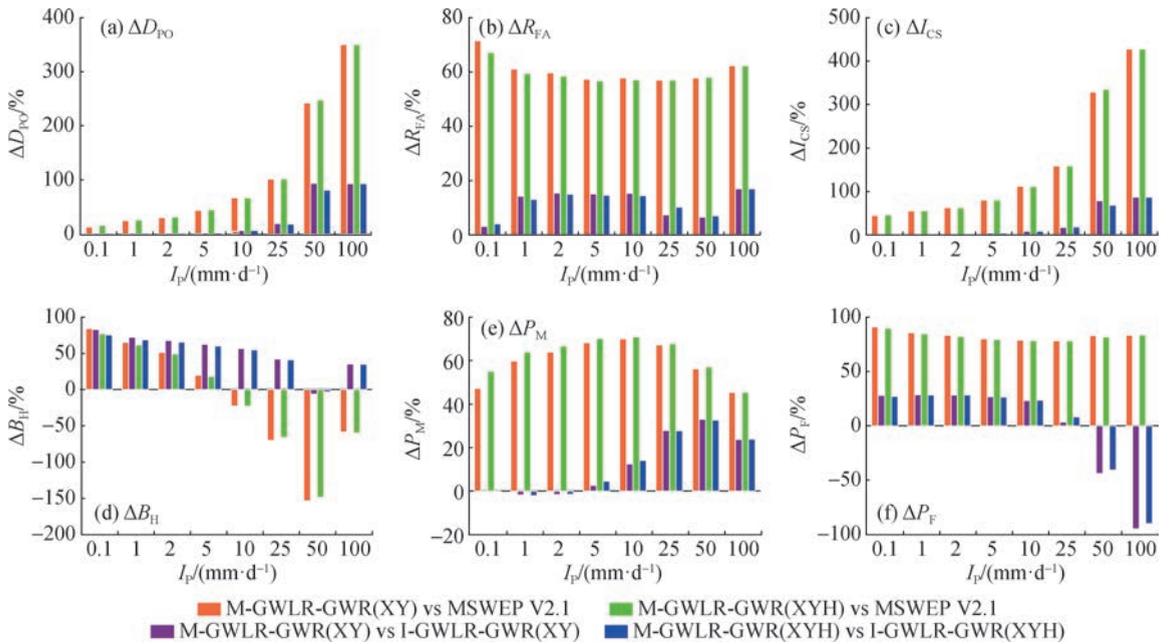


图 10 M-GWLR-GWR 方法估计不同强度日降水的精度增益

Fig. 10 Metrics benefits of M-GWLR-GWR for daily precipitation estimates with different intensities

4 结 论

本文基于地理加权逻辑回归与地理加权回归模型, 构建了考虑有雨无雨辨识的多源降水融合方法, 选择汉江流域, 开展了 MSWEP V2.1 与地面观测日降水融合试验, 综合对比了本文方法与传统方法的优缺点, 量化了较遥感反演降水数据及降水空间插值数据的统计精度增益。主要结论如下:

(1) 与传统方法相比, 本文方法有效刻画了有雨无雨空间分布及降水中心的位置, 大幅减小了误报率及误报降水量, 临界成功指数和 Kling-Gupta 效率系数均有所提高, 并且改善效果在夏季、小雨情形下更显著。引入高程对降水融合精度影响较弱。

(2) 较 MSWEP V2.1 数据, 本文方法改善误报率和误报降水量增益的中位数高于 60%, 提高综合精度 E_{KC} 幅度超过 100%。在夏季唐白河流域, 探测率和漏报降水量等指标增益相对较低。降水融合使强降水 ($I_p \geq 50$ mm/d) 事件的捕捉能力提高了 3 倍, 并降低了漏报和误报降水量。

(3) 较降水空间插值数据, 本文方法降低误报降水量和提高 Kling-Gupta 效率系数的幅度超过了 10%, 夏季 Kling-Gupta 效率系数的增益高于其他季节。此外, 降水分类辨识性能的改善幅度随雨强增大而升高, 提高强降水事件辨识精度的幅度高于 60%。

本文给出了考虑有雨无雨辨识的多源降水融合框架的一种实现方式, 纳入其他空间估计模型而演化出的不同融合方法是否更优值得继续探讨。所提方法与雨量站网的关联性较强, 在不同站网密度及分布模式下, 对于日及日以下尺度短历时降水的降水融合性能还需更多实例分析加以验证; 特别对于小时甚至更短时间分辨率的降水空间估计, 还可考虑引入卫星云图、雷达回波图像等信息, 进一步提高雨区边界的辨识精度。另外, CTC (Categorical triple collocation) 作为不依托地面观测信息的代表性融合方法^[29], 已在提升有雨无雨辨识技巧方面展示了较好的潜力。今后对比不同方法的优劣, 并考虑综合集成是重要的探索方向。

参考文献:

- [1] 任国玉, 战云健, 任玉玉, 等. 中国大陆降水时空变异规律: I: 气候学特征[J]. 水科学进展, 2015, 26(3): 299-310. (REN G Y, ZHAN Y J, REN Y Y, et al. Spatial and temporal patterns of precipitation variability over China's mainland: I: climatology[J]. *Advances in Water Science*, 2015, 26(3): 299-310. (in Chinese))
- [2] SUN Q H, MIAO C Y, DUAN Q Y, et al. A review of global precipitation data sets: data sources, estimation, and intercomparisons[J]. *Reviews of Geophysics*, 2018, 56(1): 79-107.
- [3] MAGGIONI V, MEYERS P C, ROBINSON M D. A review of merged high-resolution satellite precipitation product accuracy during the tropical rainfall measuring mission (TRMM) era[J]. *Journal of Hydrometeorology*, 2016, 17: 1101-1117.
- [4] 胡庆芳, 杨大文, 王银堂, 等. 赣江流域 TRMM 降水数据的误差特征与成因[J]. 水科学进展, 2013, 24(6): 794-800. (HU Q F, YANG D W, WANG Y T, et al. Characteristics and sources of errors in daily TRMM precipitation product over Ganjiang River basin in China[J]. *Advances in Water Science*, 2013, 24(6): 794-800. (in Chinese))
- [5] HU Q F, LI Z, WANG L Z, et al. Rainfall spatial estimations: a review from spatial interpolation to multi-source data merging [J]. *Water*, 2019, 11(3): 579.
- [6] 熊立华, 刘成凯, 陈石磊, 等. 遥感降水资料后处理研究综述[J]. 水科学进展, 2021, 32(4): 627-637. (XIONG L H, LIU C K, CHEN S L, et al. Review of post-processing research for remote-sensing precipitation products[J]. *Advances in Water Science*, 2021, 32(4): 627-637. (in Chinese))
- [7] 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.
- [8] 胡庆芳. 基于多源信息的降水空间估计及其水文应用研究[D]. 北京: 清华大学, 2013. (HU Q F. Rainfall spatial estimation using multi-source information and its hydrological application[D]. Beijing: Tsinghua University, 2013. (in Chinese))
- [9] SIDERIS I V, GABELLA M, ERDIN R, et al. Real-time radar-rain-gauge merging using spatio-temporal co-kriging with external drift in the alpine terrain of Switzerland[J]. *Quarterly Journal of the Royal Meteorological Society*, 2014, 140(680): 1097-1111.
- [10] HUANG C C, ZHENG X G, TAIT A, et al. On using smoothing spline and residual correction to fuse rain gauge observations and remote sensing data[J]. *Journal of Hydrology*, 2014, 508: 410-417.
- [11] 李哲. 多源降雨观测与融合及其在长江流域的水文应用[D]. 北京: 清华大学, 2015. (LI Z. Multi-source precipitation observations and fusion for hydrological applications in the Yangtze River basin[D]. Beijing: Tsinghua University, 2015. (in Chinese))
- [12] SINCLAIR S, PEGRAM G. Combining radar and rain gauge rainfall estimates using conditional merging[J]. *Atmospheric Science Letters*, 2005, 6(1): 19-22.
- [13] WANG Q J, SCHEPEN A, ROBERTSON D E. Merging seasonal rainfall forecasts from multiple statistical models through Bayesian model averaging[J]. *Journal of Climate*, 2012, 25(16): 5524-5537.
- [14] HASAN M M, SHARMA A, JOHNSON F, et al. Merging radar and in situ rainfall measurements: an assessment of different combination algorithms[J]. *Water Resources Research*, 2016, 52(10): 8384-8398.
- [15] CHEN C, HE M N, CHEN Q W, et al. Triple collocation-based error estimation and data fusion of global gridded precipitation products over the Yangtze River basin[J]. *Journal of Hydrology*, 2022, 605: 127307.
- [16] WU H C, YANG Q L, LIU J M, et al. A spatiotemporal deep fusion model for merging satellite and gauge precipitation in China [J]. *Journal of Hydrology*, 2020, 584: 124664.
- [17] CORNES R C, van der SCHRIER G, van den BESSELAAR E J M, et al. An ensemble version of the E-OBS temperature and precipitation data sets[J]. *Journal of Geophysical Research: Atmospheres*, 2018, 123(17): 9391-9409.
- [18] ZHANG L, LI X, ZHENG D H, et al. Merging multiple satellite-based precipitation products and gauge observations using a novel double machine learning approach[J]. *Journal of Hydrology*, 2021, 594: 125969.
- [19] XIAO S, ZOU L, XIA J, et al. Bias correction framework for satellite precipitation products using a rain/no rain discriminative model[J]. *Science of the Total Environment*, 2022, 818: 151679.
- [20] 覃文忠. 地理加权回归基本理论与应用研究[D]. 上海: 同济大学, 2007. (QIN W Z. The basic theoretics and application research on geographically weighted regression[D]. Shanghai: Tongji University, 2007. (in Chinese))
- [21] 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.
- [22] CHEN S L, XIONG L H, MA Q M, et al. Improving daily spatial precipitation estimates by merging gauge observation with multiple satellite-based precipitation products based on the geographically weighted ridge regression method[J]. *Journal of Hydrology*, 2020, 589: 125156.

- [23] NKEKI F N, ASIKHIA M O. Geographically weighted logistic regression approach to explore the spatial variability in travel behaviour and built environment interactions: accounting simultaneously for demographic and socioeconomic characteristics[J]. *Applied Geography*, 2019, 108: 47-63.
- [24] HUFFMAN G J, BOLVIN D T, BRAITHWAITE D, et al. NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG)[J]. *Algorithm Theoretical Basis Document (ATBD) Version*, 2015, 4: 26.
- [25] ORSOLINI Y, WEGMANN M, DUTRA E, et al. Evaluation of snow depth and snow-cover over the Tibetan Plateau in global re-analyses using in situ and satellite remote sensing observations[J]. *The Cryosphere*, 2019, 13(8): 2221-2239.
- [26] BECK H E, WOOD E F, PAN M, et al. MSWEP V2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment[J]. *Bulletin of the American Meteorological Society*, 2019, 100(3): 473-500.
- [27] GUPTA H V, KLING H, YILMAZ K K, et al. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling[J]. *Journal of Hydrology*, 2009, 377(1/2): 80-91.
- [28] ROEBBER P J. Visualizing multiple measures of forecast quality[J]. *Weather and Forecasting*, 2009, 24(2): 601-608.
- [29] DONG J Z, CROW W T, REICHLER R. Improving rain/no-rain detection skill by merging precipitation estimates from different sources[J]. *Journal of Hydrometeorology*, 2020, 21(10): 2419-2429.

An innovative multi-source precipitation merging method with the identification of rain and no rain*

LI Lingjie^{1,2}, WANG Yintang^{1,2}, TANG Guoqiang³, GAO Xuan⁴, WANG Leizhi¹, HU Qingfang^{1,2}

(1. *State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing 210029, China*; 2. *Yangtze Institute for Conservation and Development, Nanjing 210098, China*; 3. *University of Saskatchewan Coldwater Lab, Canmore, Alberta T1W 3G1, Canada*; 4. *Key Laboratory of Agricultural Soil and Water Engineering in Arid and Semiarid Areas, Ministry of Education, Northwest A&F University, Yangling 712100, China*)

Abstract: Multi-source precipitation merging is a crucial way to estimate the spatiotemporal distribution of precipitation accurately. The commonly used merging methods mainly focus on bias correction of the total precipitation amount or precipitation intensity but often neglect to identify short-duration precipitation. In this study, we proposed a merging framework of multi-source precipitation by identifying rain and no rain and constructed a precipitation merging method considering both rain area identification and rainfall estimation by coupling the geographical weighted logistic regression (GWLR) and geographically weighted regression models (GWR). Then, the merging experiments of the Multi-Source Weighted-Ensemble Precipitation Version 2.1 (MSWEP V2.1) and the daily precipitation observed by the ground gauges network over the Han River basin were implemented. The results show that the proposed method successfully reproduces the spatial pattern of rain and no rain and catches the precipitation center. It overall strengthens the performance of MSWEP V2.1 to estimate ground precipitation, reduces the false alarm rate (R_{FA}) and false precipitation (P_F) by more than 60%, and improves the critical success index (I_{CS}) and Kling-Gupta efficiency coefficient (E_{KG}) by more than 40%. Moreover, the gains of correcting P_F and improving E_{KG} are higher than 10% against the spatially interpolated precipitation. Meanwhile, compared with reference data, precipitation fusion enhances the classification accuracy of heavy precipitation events (intensity ≥ 50 mm/d) by not less than 60%. The innovative method effectively improves the performance of precipitation estimation and provides a new idea for multi-source precipitation merging.

Key words: multi-source precipitation merging; rain and no rain identification; geographically weighted logistic regression; geographically weighted regression; Han River basin

* The study is financially supported by the National Key R&D Program of China (No. 2021YFC3000104) and the National Natural Science Foundation of China (No. 52009081).