

# 利用空间随机森林方法提升GPM卫星遥感降水质量

胡保健<sup>1</sup>, 李伟<sup>2</sup>, 陈传法<sup>1</sup>, 胡占占<sup>1</sup>

1. 山东科技大学 测绘与空间信息学院, 青岛 266590;

2. 湖州中核勘测规划设计有限公司, 湖州 313000

**摘要:** 卫星遥感降水产品是当前获取大范围、连续性降水观测的主要来源, 但目前已有的卫星遥感降水产品空间分辨率粗糙, 且存在一定的系统偏差。为此, 本文充分考虑高分辨率环境变量(包括地形、NDVI、地表温度、经纬度)对降水影响以及邻近遥感降水(站点)空间相关性, 构建了一种双阶段空间随机森林SRF(Spatial Random Forest)方法(SRF-SRF)。以四川省2015年—2019年GPM(Global Precipitation Measurement Mission)月降水数据为例, 借助SRF-SRF对其质量提升, 并将计算结果与现有7种方法比较, 包括地理加权回归(GWR)、反向传播神经网络(BPNN)、随机森林(RF)、站点降水Kriging插值(Kriging)、经SRF降尺度后的地理差异分析校正(SRF-GDA)、经双线性插值降尺度后的SRF校正(Bi-SRF)以及年降水经SRF降尺度后按月比例分解并利用SRF校正(SRFdis)等。实验分析表明:(1)在月尺度上, 与原始GPM相比, SRF-SRF的平均绝对误差(MAE)降低了19.51%, 中误差(RMSE)降低了16.35%, 而且精度优于其他方法; 在季尺度上, SRF-SRF在冬季误差最小, 在夏季误差最大, 但其计算精度均优于其他方法; 在年尺度上, 基于SRF的4种方法(包括SRF-SRF、SRF-GDA、Bi-SRF和SRFdis)优于GWR、BPNN、RF, 并且SRF-SRF计算精度优于单阶段的Bi-SRF和SRF-GDA。(2)SRF-SRF降水产品空间分布连续性较好, 且局部降水细节得到明显提升。(3)借助RF对各自变量重要性分析得出, 降水空间相关性对卫星遥感降水质量提升具有重要作用。(4)基于月尺度的SRF-SRF融合校正效果优于基于年尺度的SRFdis, 表明NDVI可用于该区域月尺度降水质量提升。

**关键词:** 遥感, 降水, 降尺度, 点面融合, 随机森林, GPM, 机器学习

**中图分类号:** P426.6/P2

**引用格式:** 胡保健, 李伟, 陈传法, 胡占占. 2024. 利用空间随机森林方法提升GPM卫星遥感降水质量. 遥感学报, 28(2): 414-425  
Hu B J, Li W, Chen C F and Hu Z Z. 2024. Improving the quality of remotely sensed precipitation product from GPM satellites by using a spatial random forest. National Remote Sensing Bulletin, 28(2): 414-425[DOI:10.11834/jrs.20221222]

## 1 引言

降水是地球能量循环与物质交换重要组成部分, 是陆地生态系统重要驱动因素, 在气候变化中起着决定性作用(Markonis等, 2019; Zhou等, 2020b)。高分辨率和高精度的空间降水分布对区域水资源调度管理、灾害预测防控、生态研究等领域具有重要的应用价值(Brodeur和Steinschneider, 2020; Renard等, 2011; Zhang等, 2020)。然而, 降水数据估算误差在上述应用中会带来很大的不确定性(Xie和Xiong, 2011)。如何提高降水空间分辨率和准确性是近年来生态水文等领域研究热

点和难点(Shen等, 2014b)。

降水监测数据主要来自气象观测站、地基雷达测量、卫星遥感降水反演产品等。气象站观测降水可直接测得实际降水量, 但受地形限制, 难以反映降水的空间连续性分布(Sharifi等, 2019; Shen等, 2014a)。地基雷达观测可以提供连续高分辨率空间降水数据, 但只能反映局部区域降水, 并且复杂地形区域布设困难、易受环境因素干扰多。随着遥感技术的发展, 产生了大量基于星载探测技术的卫星遥感降水产品, 如TRMM、PERSIANN、GPM、CHIRPS、CMORPH(Hu等, 2014; Islam等, 2020; Sun等, 2016)。相对于气象站点观测与雷

收稿日期: 2021-04-28; 预印本: 2022-02-18

基金项目: 山东省自然科学基金(编号:ZR2020YQ26、ZR2019MD007); 山东省高等学校青创科技支持计划(编号:2019KJH007)

第一作者简介: 胡保健, 研究方向为多源遥感数据融合与处理。E-mail: baojian\_hu@foxmail.com

通信作者简介: 陈传法, 研究方向为数字地形建模、点云智能处理、资源环境模型与系统模拟等。E-mail: chencf@lreis.ac.cn

达测量降水，卫星遥感降水产品具有覆盖范围广、时空连续变化效果好的特点（Zhou等，2020a），在水文模型和生态环境研究中得到广泛应用（Jiang等，2021；Wei等，2020）。

然而，现有卫星遥感降水产品分辨率较低（最优仅为 $0.05^\circ$ ），而且受传感器误差与降水反演算法不确定性等影响，现有降水产品均含有一定的系统偏差，导致其不能满足局部尺度和生态流域尺度的应用研究（Immerzeel等，2009）。因此，国内外研究者们相继提出了多种点面融合方法，以借助地面观测站降水对卫星降水产品修正，如最优插值法（OI）（卢新玉等，2017；潘旻等，2012）、贝叶斯模型（Fu等，2016）、地理差异分析（GDA）（Jongjin等，2016）、地理加权回归（GWR）（Lu等，2019）、K最近邻模型（KNN）（王玉丹等，2016）等。这些方法虽然有效提高了卫星降水产品的精度，但仍有一些问题需要解决：（1）融合后的降水产品分辨率与原有的卫星降水产品相同，不利于局部地区水文模型研究；（2）一些方法只考虑了实测降水背景场，未考虑与降水有关的环境因素；（3）在进行偏差校正时，粗分辨率的卫星降水产品与基于点的观测站实测降水存在尺度差异，导致点面融合时捕捉不到局部降水细节信息。因此，在点面融合之前对遥感降水产品进行空间降尺度是十分必要的。

近年来，国内外学者们提出和发展了多种遥感降水产品降尺度方法，如指数回归（ER）（Immerzeel等，2009）、多元线性回归（MLR）（嵇涛等，2015；Jia等，2011；马金辉等，2013）、GWR方法等（胡实等，2020）。随着研究深入，一些研究者们开始尝试使用机器学习方法对遥感降水进行降尺度研究。大量研究表明，随机森林RF（Random Forest）模型在处理多维数据中具有良好的表现，且能够很好的捕捉自变量与因变量之间非线性关系（Baez-Villanueva等，2020；Njuki等，2020），具有良好的应用前景（Jing等，2016；Ma等，2018）。然而，上述降尺度研究均将RF模型作为一种简单的统计工具，忽略了相邻遥感（站点）降水数据的空间相关性。

基于以上讨论，本文提出了一种顾及空间相关性的双阶段卫星降水产品质量提升方法，即降尺度与点面融合相结合的思路，以提升卫星降水产品的空间分辨率与精度。该方法以RF为基础模

型，构造了顾及空间相关性的空间随机森林模型SRF（Spatial Random Forest）。与现有方法相比，新方法优势包括：（1）考虑了降水之间空间相关性；（2）降尺度与点面融合两阶段都采用了SRF模型，且均融入了环境变量影响。以四川省为研究区域，选取最新GPM IMERG V06B遥感降水数据作为数据源，使用SRF-SRF方法对GPM降水数据进行降尺度与点面融合处理，并将计算结果与经典方法比较，验证新方法的可行性和高效性。

## 2 研究区域和数据源

### 2.1 研究区概况

四川省位于中国西南部（ $97^\circ 21'E-108^\circ 31'E$ ， $26^\circ 03'N-34^\circ 19'N$ ），总面积约为 $4.86 \times 10^5 \text{ km}^2$ ，介于青藏高原与长江中下游，是两者的过渡地带。研究区域地形地貌复杂多样，涵盖了山脉、高原、丘陵、平原和盆地等地形，地形起伏走势西高东低。受地形因素和空间位置影响，该区域降水夏季多冬季少，其中夏季降水占全年80%—90%，且受川西垂直地势起伏影响，四川盆地降水充沛，年降水量800—1600 mm，呈现由东到西降水逐渐减少的变化特点（Yang等，2020；Lai和Gong，2017）。图1为研究区域地形及气象站点分布。

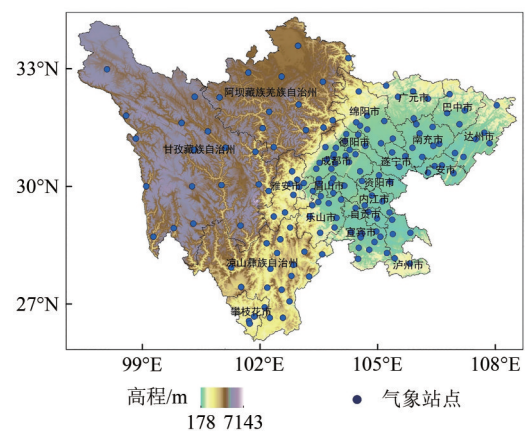


图1 四川省气象站分布

Fig.1 Distribution of weather stations in Sichuan Province

### 2.2 数据源

（1）气象站数据。气象站观测数据由中国气象局气象数据中心（<http://data.cma.cn/> [2021-01-01]）提供。选取四川省内2015年—2019年期间156个逐日气象站点降水数据（图1），所有数据经过内部一致性检查、极值检查以及空间一致性检查等

严格质量控制后 (Song 等, 2004), 经累积计算得到每月降水量。

(2) 卫星降水数据。GPM 降水数据从美国宇航局地球科学数据中心 (<https://pmm.nasa.gov/> [2021-01-01]) 获取, 是美国 NASA (美国国家航空航天局) 和日本 JAXA (日本宇宙航空研发机构) 共同研发的继 TRMM 之后全球降水测量卫星, 搭载全球先进的 DPR (Dual-frequency Precipitation Radar) 和 GPM GMI (GPM Microwave Imager) 系统, 对微小降雨探测能力比前一代产品 TRMM 更精准, 覆盖范围 ( $60^{\circ}\text{S}$ — $60^{\circ}\text{N}$ ) 更广阔。本文选用月尺度  $0.1^{\circ}\times 0.1^{\circ}$  分辨率的降水产品 GPM IMERG V06B (简称 GPM) 数据集作为本研究所采用的数据源, 时间跨度为 2015 年 1 月—2019 年 12 月。

(3) DEM、NDVI 以及地表温度 (LST) 数据。研究采用的 DEM 数据为 SRTM (Shuttle Radar Topography Mission) DEM V4.1 数据集, 从地理空间数据云 (<http://www.gscloud.cn/> [2021-01-01]) 获取, 空间分辨率为 90 m, 经像素平均重采样生成 1 km 分辨率, 然后在 Arcgis 中利用表面分析工具获

得坡度、坡向、地形起伏度数据。NDVI (MOD13A3)、LST (MOD11A2) 数据从美国航空局 (<https://ladsweb.modaps.eosdis.nasa.gov/> [2021-01-01]) 中获取, 空间分辨率为 1 km, NDVI 时间分辨率为月, LST 时间分辨率为 8 d。其中, LST 数据包括白天地表温度 ( $LST_d$ ) 与夜间地表温度 ( $LST_n$ ) 以及白天与夜间的温度差 ( $LST_{d-n}$ ), 由每 8 天 LST 求平均得出月平均温度 LST。

### 3 研究方法

本文构建了一种双阶段卫星降水产品质量提升方法。第一阶段, 通过考虑降水与其他因素 (如地形、NDVI、地表温度、经纬度等) 之间的关系, 借助 SRF 将 10 km GPM 卫星降水产品降尺度到 1 km 空间分辨率 (D\_GPM)。第二阶段, 将实测站点降水数据与 D\_GPM 及上述自变量 (如地形、NDVI、地表温度、经纬度等) 相结合, 再次借助 SRF 生成高精度高空间分辨率的降水数据。该方法的流程图如图 2 所示。

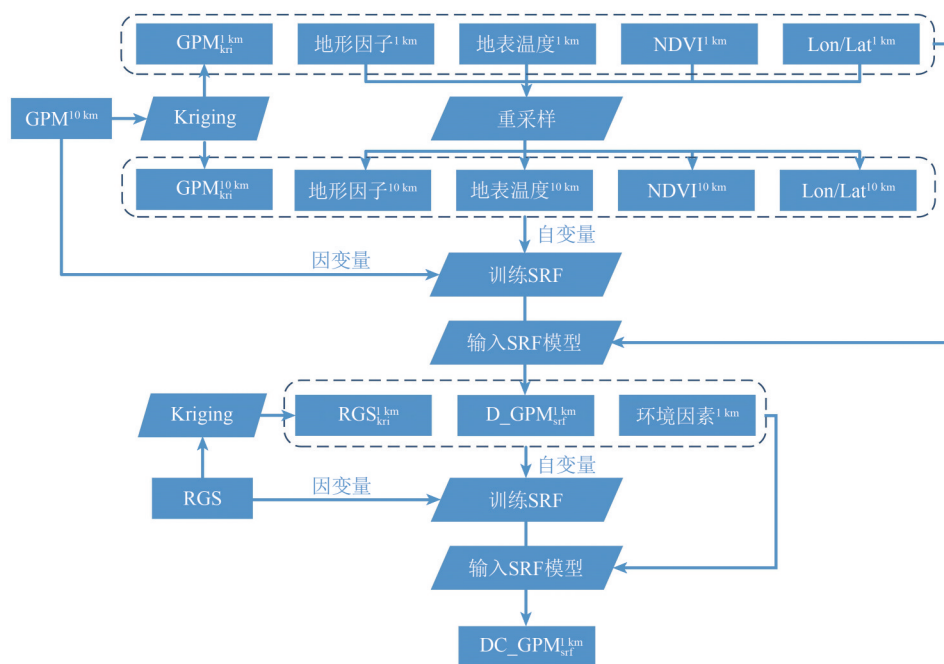


图 2 双阶段降尺度流程图

Fig.2 Two-stage downscaling flowchart

#### 3.1 随机森林 (RF) 模型

随机森林 (RF) 算法是 Breiman 于 2001 年提出的一种基于 bagging 的集成学习方法, 通过构造多种决策树来处理自变量与因变量之间的关系,

可用作数据的分类和回归预测。RF 通过构建大量的树模型, 对多种特征值的重要性进行整合筛选, 充分考虑不同特征值之间的重要性, 选择最优样本特征值进而找到最优解, 并求得所有预测值的

平均值作为最终预估值。与传统回归预测方法相比，RF可以处理复杂多维的特征值，回归预测更准确，采取树状重复放回抽样，避免过拟合，且不必考虑线性回归中的多重共线性问题，稳健性更好，在多种领域得到广泛应用。RF模型的通用公式如下表示：

$$P(s_0) = f(X_i(s_0), i = 1, 2, \dots, k) + \varepsilon \quad (1)$$

式中， $P(s_0)$  为  $s_0$  处的降水预测值， $X_i(s_0)$  ( $i=1, 2, 3, \dots, k$ ) 为  $s_0$  处的自变量， $k$  为自变量个数， $\varepsilon$  为预测误差。

### 3.2 顾及空间相关性的随机森林(SRF)模型

为充分考虑邻近降水之间的空间相关性，本文构建了顾及空间相关性的空间随机森林(SRF)模型。其中，SRF是在RF的基础上引入克里金估计值作为输入变量，其通用公式如下表示：

$$P(s_0) = f(X_i(s_0), P_s(s_0), i = 1, 2, \dots, k) \quad (2)$$

式中， $P_s(s_0)$  为  $s_0$  处的克里金降水估计值。

克里金插值法是一种无偏最优估计值方法(史文娇等, 2012; Chen和Li, 2019; Kim等, 2013)，对于预估点  $s_0$  处的降水值  $P_s(s_0)$  可通过搜索该点周围邻近  $n$  个站点的降水  $P(x)$  经线性加权求得，其计算公式为

$$P_s(s_0) = \sum_{i=1}^n \lambda_i P(s_i) \quad (3)$$

式中， $P_s(s_0)$  为  $s_0$  处的降水估计值， $\lambda_i$  为克里金权重系数， $n$  为周围邻近站点数， $P(s_i)$  为站点  $s_i$  处的降水值。

克里金权重系数  $\lambda_i$  不仅取决于周围邻近降水站点与预估点的距离，还取决于邻近点空间分布，其值是通过求解克里金目标函数实现，即  $P_s(s_0)$  对真实值  $P(s_0)$  的估计为无偏(式(4))且方差最小(式(5))。

$$P(s_0) = E(P_s(s_0)) \quad (4)$$

$$\min_{\lambda_i} \text{Var}(P_s(s_0) - P(s_0)) \quad (5)$$

由于区域性降水分布具有较大的空间相关性(Sekulić等, 2020)，因此，借助克里金插值可充分考虑样本之间的空间相关性。

### 3.3 SRF双阶段降尺度

本研究的双阶段卫星降水产品质量提升方法大致分为SRF降尺度与SRF点面融合两部分，具体步骤如下：

(1) 克里金插值。将10 km分辨率的GPM数据通过克里金插值获得10 km分辨率的  $\text{GPM}_{\text{kri}}^{10\text{km}}$  和1 km分辨率的  $\text{GPM}_{\text{kri}}^{1\text{km}}$ ，对站点实测降水数据(RGS)进行克里金插值获得1 km分辨率  $\text{RGS}_{\text{kri}}^{1\text{km}}$ 。

(2) 将1 km空间分辨率的所有自变量  $X^{1\text{km}}$  (包括NDVI、 $\text{LST}_d$ 、 $\text{LST}_n$ 、 $\text{LST}_{d-n}$ 、DEM、坡度、坡向、地形起伏度以及经纬度)分别使用像元平均重采样至10 km空间分辨率，记为  $X^{10\text{km}}$ 。

(3) 以步骤(2)中重采样自变量  $X^{10\text{km}}$  以及步骤(1)中  $\text{GPM}_{\text{kri}}^{10\text{km}}$  为自变量，原始GPM为因变量建立SRF降尺度模型：

$$P^{\text{GPM}^{10\text{km}}}(s_0) = f_d(X^{10\text{km}}(s_0), \text{GPM}_{\text{kri}}^{10\text{km}}(s_0)) + \varepsilon \quad (6)$$

式中， $P^{\text{GPM}^{10\text{km}}}(s_0)$  为  $s_0$  处的GPM降水值。

(4) 将所有1 km空间分辨率的自变量  $X^{1\text{km}}$  输入到SRF降尺度模型获得1 km空间分辨率的卫星降水  $D\_GPM_{\text{srf}}$ ：

$$D\_GPM_{\text{srf}} = f_d(X^{1\text{km}}, \text{GPM}_{\text{kri}}^{1\text{km}}) \quad (7)$$

(5) 对降尺度数据进行点面融合，即建立以RGS为因变量的校正模型，自变量包括以步骤(4)中的降尺度结果  $D\_GPM_{\text{srf}}$ 、1 km空间分辨率下的自变量  $X^{1\text{km}}$  以及  $\text{RGS}_{\text{kri}}^{1\text{km}}$ ，即：

$$\text{RGS}(s_0) = f_c(X^{1\text{km}}(s_0), D\_GPM_{\text{srf}}(s_0), \text{RGS}_{\text{kri}}^{1\text{km}}(s_0)) + \varepsilon_c \quad (8)$$

式中， $\text{RGS}(s_0)$  为站点  $s_0$  处的实测降水。

(6) SRF模型训练以后，将全部1 km分辨率自变量输入到SRF融合模型中，获取融合校正后的1 km降水数据  $DC\_GPM_{\text{srf}}$ 。

$$DC\_GPM_{\text{srf}} = f_c(X^{1\text{km}}, D\_GPM_{\text{srf}}, \text{RGS}_{\text{kri}}^{1\text{km}}) \quad (9)$$

### 3.4 评价方法

为评价新方法(SRF-SRF)的可行性和高效性，本文选取了7种方法与其进行比较，包括3种传统方法，即GWR、RF、BPNN；3种基于SRF框架下的方法，即GPM经双线性插值降尺度后使用SRF方法进行点面融合方法(Bi-SRF)、对GPM使用SRF降尺度后利用站点实测降水进行地理差异分析校准方法(SRF-GDA)、对GPM年降水使用SRF降尺度后按月比例分解并经SRF校正方法(SRFdis)；最后一种方法为借助克里金直接对站点降水插值(记为Kriging)。

本文采用十折交叉验证方法验证所有方法的性能。选取的精度评价指标包括平均绝对误差

(MAE)、均方根误差 (RMSE)、相关系数 (CC) (Lu 等, 2020)。3种评价指标公式如下:

$$\text{MAE} = \sum_{i=1}^n |\Delta P_i| / n \quad (10)$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^n \Delta P_i^2 / n} \quad (11)$$

$$\text{CC} = \frac{\sum_{i=1}^n (P_{si} - \overline{P_{si}})(P_{oi} - \overline{P_{oi}})}{\sqrt{\sum_{i=1}^n (P_{si} - \overline{P_{si}})^2} \sqrt{\sum_{i=1}^n (P_{oi} - \overline{P_{oi}})^2}} \quad (12)$$

式中,  $P_{oi}$ 、 $P_{si}$  分别代表第  $i$  个站点降水实测值和该站点对应的降水预估数据,  $\Delta P_i = P_{si} - P_{oi}$ ,  $\overline{P_o}$ 、 $\overline{P_s}$  分别代表站点实测降水的算术平均值和该站点对应降水预估数据算术平均值,  $n$  表示气象站点个数。相关系数  $CC$  表示两种数据相关性, 值越接近 1 表示相关程度越好。RMSE 和 MAE 用于评价预估降水与实测数据之间的误差, 值越小表示精度越高。

## 4 结果与分析

采用 SRF-SRF 方法对四川省 GPM 月降水数据进行降尺度与点面融合研究, 并将该方法的结果与经典方法的结果在月、季、年 3 种时间尺度上进行比较。

### 4.1 月尺度分析

表 1 显示了各模型在 2015 年—2019 年全部月份上的计算精度。结果显示, 在所有模型结果中, GWR 与 BPNN 效果最差, 可能是这两种方法没有有效处理降水与环境变量之间的复杂关系。基于站点的 kriging 比 RF 方法好, 且所有考虑空间相关性模型 (包括 SRF-SRF、Bi-SRF、SRF-GDA、SRFdis) 计算结果精度都优于传统的方法, 说明了引入降水之间的空间相关性信息对模型精度提升尤为重要。与原始 GPM 相比, SRF-SRF 的 MAE 和 RMSE 分别降低了 19.51%、16.35%。整体而言, SRF-SRF 精度优于 Bi-SRF、SRF-GDA、SRFdis 方法。

各个方法在月尺度上的 MAE、RMSE、CC 箱线图如图 3 所示。结果显示, BPNN 方法在所有模型中效果最差, MAE、RMSE、CC 等 3 种指标的中位数分别为 22.66 mm、30.48 mm、0.64, 其次为 GWR、RF、Kriging 方法, 这与全部月份精度对比 (表 1) 一致。基于 SRF 的 4 种方法结果精度都优于传统方法, 其中 SRF-SRF 方法的 MAE、RMSE、CC 等 3 种指标的中位数分别为 15.66 mm、21.03 mm、

0.81, 而其他 3 种基于 SRF 方法的 3 个指标分别为 15.83—16.15 mm、21.41—22.27 mm、0.77~0.79, 进一步说明了 SRF-SRF 的高效性。

表 1 2015 年—2019 年全部月份各模型精度对比

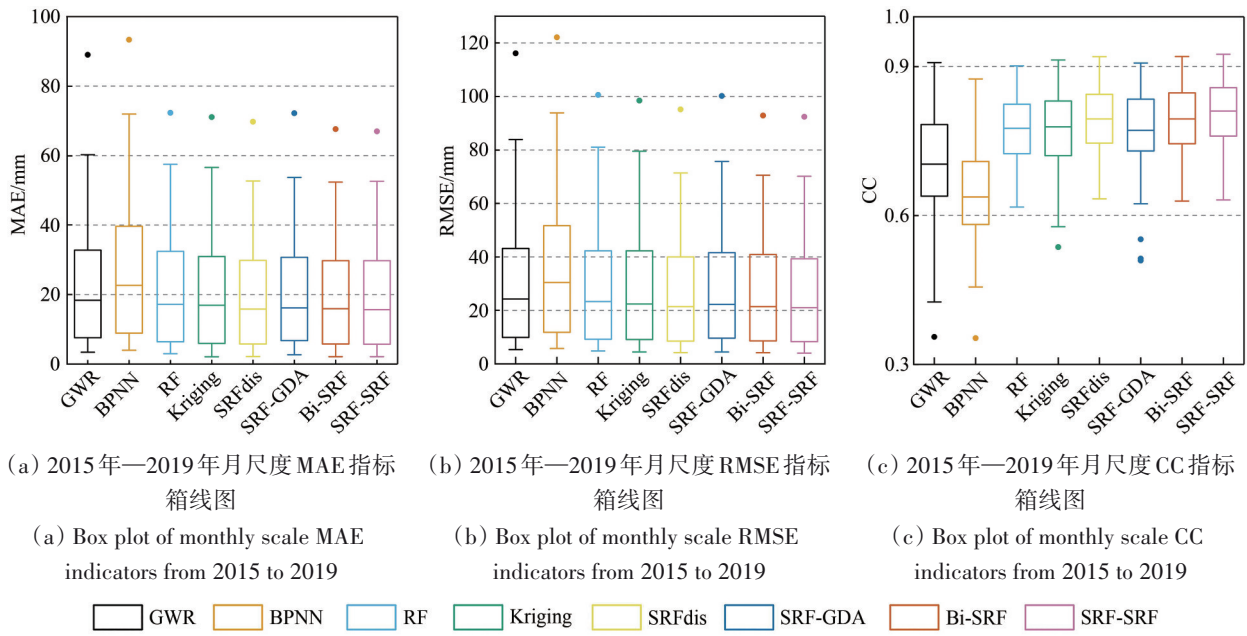
Table 1 Precision comparison of each model in all months from 2015 to 2019

模型	MAE/mm	RMSE/mm	CC
GPM	23.88	38.65	0.913
GWR	22.71	38.77	0.907
BPNN	27.58	45.06	0.877
RF	21.05	36.34	0.919
Kriging	20.52	35.98	0.922
SRFdis	19.58	33.98	0.929
SRF-GDA	20.30	35.15	0.926
Bi-SRF	19.52	33.78	0.930
SRF-SRF	19.22	33.22	0.933

图 4 为 GWR、BPNN、RF、Kriging、SRFdis、SRF-SRF 在月尺度上各站点 RMSE 分布图。由于受地势和气候影响 (图 1), 来自印度洋水汽在四川中部盆地与高原相交地段受阻凝结形成降水, 故雅安与乐山地区降水偏多, 导致其较大的 RMSE。GWR、BPNN、RF 方法在各个站点的 RMSE 均高于基于 SRF 的模型, 尤其在四川盆地降水较多地带。这是因为在盆地区域站点分布较为均匀, 考虑站点之间的相关性可以有效校正该区域的误差。由图 4 (e) 和图 4 (f) 表明, 与 SRFdis 相比, SRF-SRF 方法在四川东北地区与四川盆地地区略好一些。

### 4.2 季尺度分析

表 2 显示了各模型在不同季节上的精度表现。四川地区气候普遍表现为冬暖夏热, 夏季降水多, 春秋降水少, 所以在夏季降水的 GPM 反演降水误差较高。具体而言, GWR 与 BPNN 方法在冬季的表现要优于其他季节, 可能该地区冬季气候较为温暖, 降水与不同环境因子之间相关性较强。3 种传统方法中, RF 优于 GWR、BPNN, 而所有基于 SRF 方法 (除冬季的 SRF-GDA 方法外) 的精度都优于这 3 种方法。SRF-GDA 在冬季的效果较差, 主要是因为冬季降水较少, 且在进行点面融合时, 未考虑站点以外的环境变量。SRF-SRF 方法在这所有模型中方法效果最好, 其中在冬季提升效果明显; 相较于原始 GPM, SRF-SRF 的 MAE 降低了 40.69%, RMSE 降低了 44.24%, CC 提高了 43.36%。



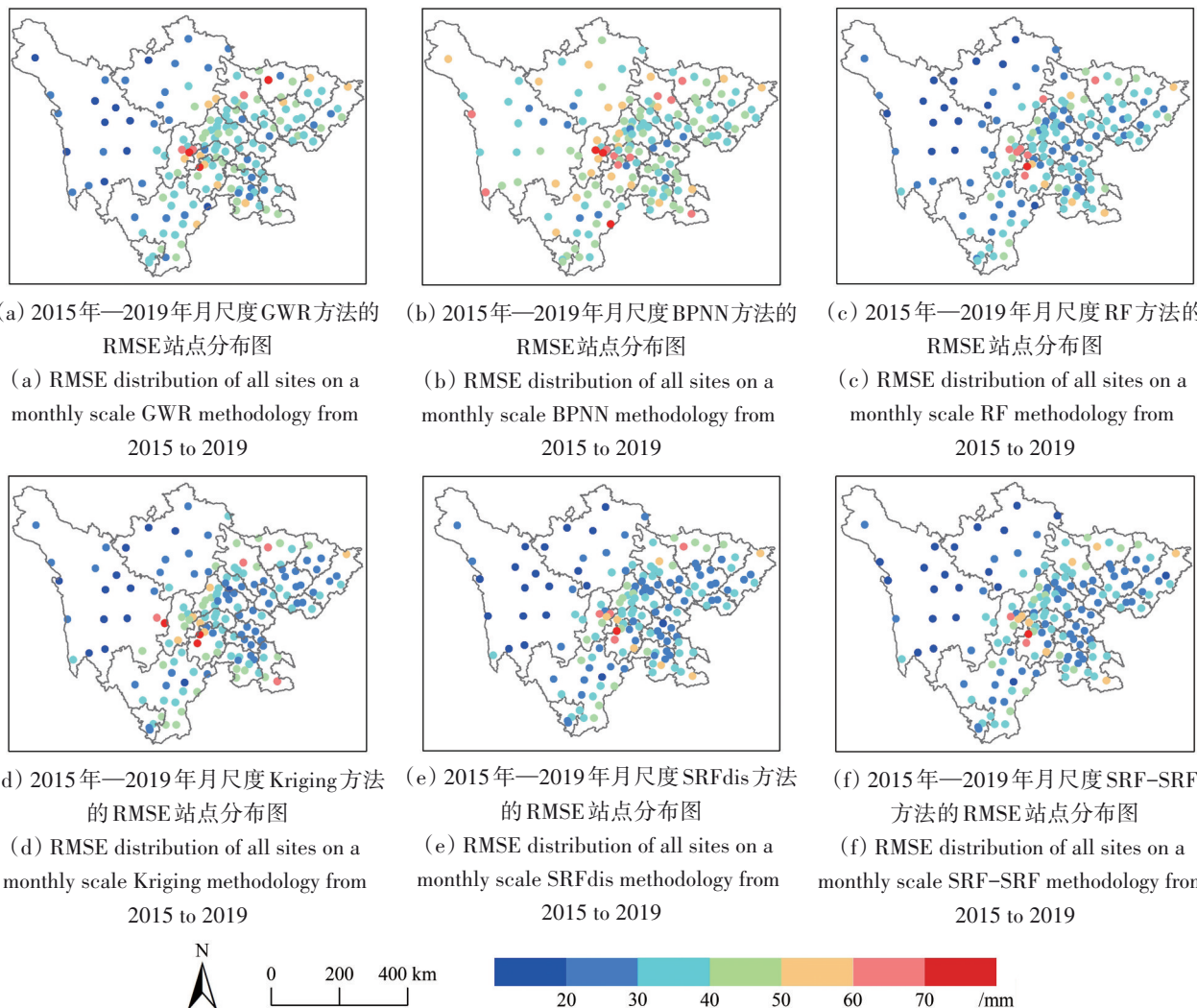
(a) 2015年—2019年月尺度MAE指标箱线图 (b) 2015年—2019年月尺度RMSE指标箱线图 (c) 2015年—2019年月尺度CC指标箱线图

(a) Box plot of monthly scale MAE indicators from 2015 to 2019 (b) Box plot of monthly scale RMSE indicators from 2015 to 2019 (c) Box plot of monthly scale CC indicators from 2015 to 2019

Legend: GWR (white), BPNN (orange), RF (light blue), Kriging (green), SRFdis (yellow), SRF-GDA (blue), Bi-SRF (red), SRF-SRF (purple)

图3 2015年—2019年月尺度精度指标箱线图

Fig. 3 Box plot of monthly scale accuracy indicators from 2015 to 2019



(a) 2015年—2019年月尺度GWR方法的RMSE站点分布图 (b) 2015年—2019年月尺度BPNN方法的RMSE站点分布图 (c) 2015年—2019年月尺度RF方法的RMSE站点分布图

(a) RMSE distribution of all sites on a monthly scale GWR methodology from 2015 to 2019 (b) RMSE distribution of all sites on a monthly scale BPNN methodology from 2015 to 2019 (c) RMSE distribution of all sites on a monthly scale RF methodology from 2015 to 2019

(d) 2015年—2019年月尺度Kriging方法的RMSE站点分布图 (e) 2015年—2019年月尺度SRFdis方法的RMSE站点分布图 (f) 2015年—2019年月尺度SRF-SRF方法的RMSE站点分布图

(d) RMSE distribution of all sites on a monthly scale Kriging methodology from 2015 to 2019 (e) RMSE distribution of all sites on a monthly scale SRFdis methodology from 2015 to 2019 (f) RMSE distribution of all sites on a monthly scale SRF-SRF methodology from 2015 to 2019

Legend: 0, 200, 400 km scale bar. Color scale: 20, 30, 40, 50, 60, 70 /mm

图4 2015年—2019年月尺度RMSE站点分布图

Fig. 4 RMSE distribution of all sites on a monthly scale from 2015 to 2019

表2 2015年—2019年季尺度各模型精度对比  
Table 2 Accuracy comparison of various models on a seasonal scale from 2015 to 2019

季节	模型	MAE/ mm	RMSE/ mm	CC
春季	GPM	19.30	25.97	0.855
	GWR	18.24	25.97	0.841
	BPNN	22.48	30.71	0.793
	RF	16.64	23.67	0.870
	Kriging	16.14	23.38	0.876
	SRFdis	15.48	22.04	0.888
	SRF-GDA	15.93	23.03	0.882
	Bi-SRF	15.52	22.19	0.887
	SRF-SRF	15.36	21.99	0.889
夏季	GPM	47.30	64.46	0.818
	GWR	46.32	65.49	0.795
	BPNN	54.55	74.46	0.745
	RF	43.66	61.83	0.824
	Kriging	43.16	61.07	0.832
	SRFdis	40.63	57.51	0.849
	SRF-GDA	41.77	59.27	0.845
	Bi-SRF	40.44	57.06	0.851
	SRF-SRF	39.92	56.13	0.857
秋季	GPM	21.98	32.19	0.905
	GWR	20.79	31.63	0.902
	BPNN	26.67	39.70	0.864
	RF	19.25	29.48	0.917
	Kriging	18.42	29.49	0.918
	SRFdis	17.96	28.24	0.924
	SRF-GDA	18.53	29.23	0.920
	Bi-SRF	17.89	28.06	0.925
	SRF-SRF	17.51	27.50	0.928
冬季	GPM	6.93	11.28	0.595
	GWR	5.66	8.11	0.735
	BPNN	6.64	9.29	0.688
	RF	4.65	6.83	0.826
	Kriging	4.36	7.01	0.823
	SRFdis	4.26	6.51	0.841
	SRF-GDA	4.97	7.70	0.790
	Bi-SRF	4.25	6.54	0.839
	SRF-SRF	4.11	6.29	0.853

### 4.3 年尺度分析

图5显示了各模型在年尺度上的精度表现。由于四川省从2015年—2018年年降水量逐年增加,且在2018年降水量最多,所以各方法在2018年的MAE、RMSE最大。所有方法中, BPNN最差,其次为GWR,其与月尺度和季尺度表现效果一致。与Kriging与GWR、BPNN相比, RF略好一些;基于SRF模型效果表现均优于传统的方法。整体而言, SRF-SRF精度最优,而SRF-GDA方法在所有基于SRF模型中表现最差,这主要是因为后者在降水校正时未考虑环境变量有关。

### 4.4 降水的空间分布特征

图6显示了各方法在2018年7月的降水空间分布图,其中,该月份为5年内降水最多的月份。从降水分布可以看出,四川省地区降水分布空间差异明显,与地形走势较为接近,即降水多分布于四川中东部地势较低地区,而川西地区由于地势突起降水相对较少。原始GPM(图6(a))降水有效捕捉了川西地区的降水空间分布与降水量,而在四川中部地区却低估了降水,这可能是由于四川中部地形复杂,遥感卫星监测时受四川中部地势和潮湿气候影响,导致反演的卫星降水与实测降水具有一定的误差。经降尺度融合后,各种方法明显改善了中部地区的低估情况,空间分布整体上呈一致性。就空间分辨率而言,原始GPM(图6(a))降水数据分辨率较为粗糙,克里金插值站点结果(图6(b))虽然分辨率提升,但图像比较模糊,细节特征不明显;其他方法降水空间分布细节信息丰富,空间分辨率得到显著提升。GWR方法(图6(c))的降水空间分布存在较多的噪声,即在部分区域呈现断崖式降低,不符合降水连续性分布的特点;RF方法(图6(d))的降水在四川东部与四川南部呈现部分块状分布,但整体上相对于GWR更平滑一些;SRF-SRF方法(图6(e))更能刻画出四川中部地区降水的空间分布情况,且减少了RF预测模型结果的块状分布,降水分布在空间上更具有连续性,符合降水分布的特点。

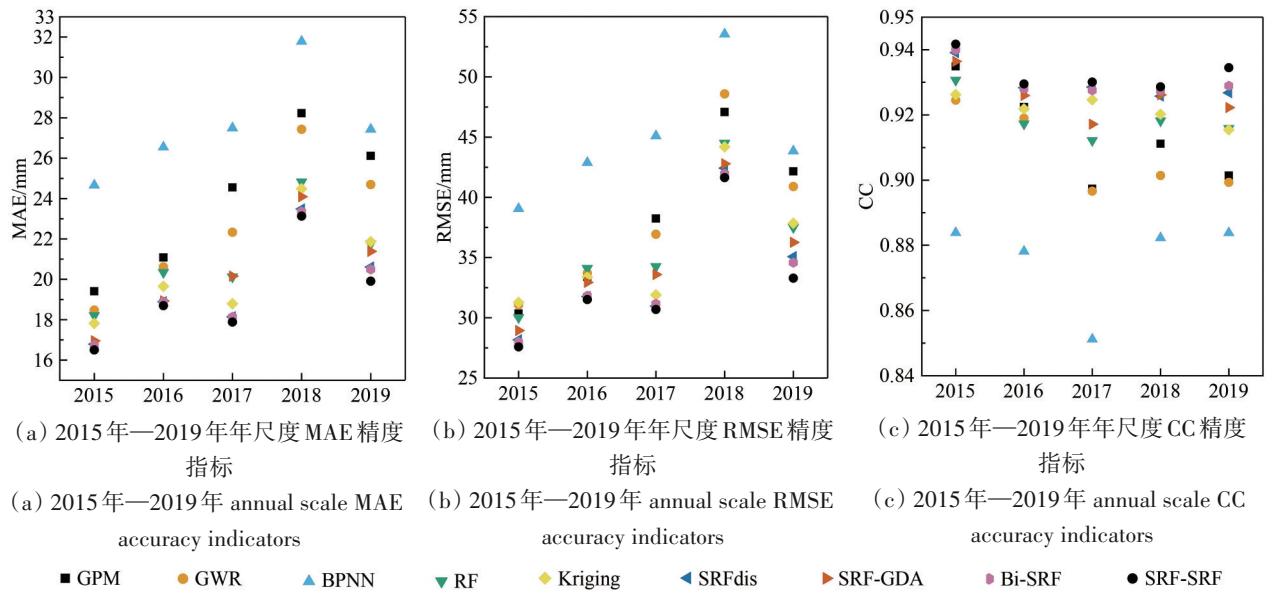


图5 2015年—2019年年尺度精度指标  
 Fig. 5 2015—2019 annual scale accuracy indicators

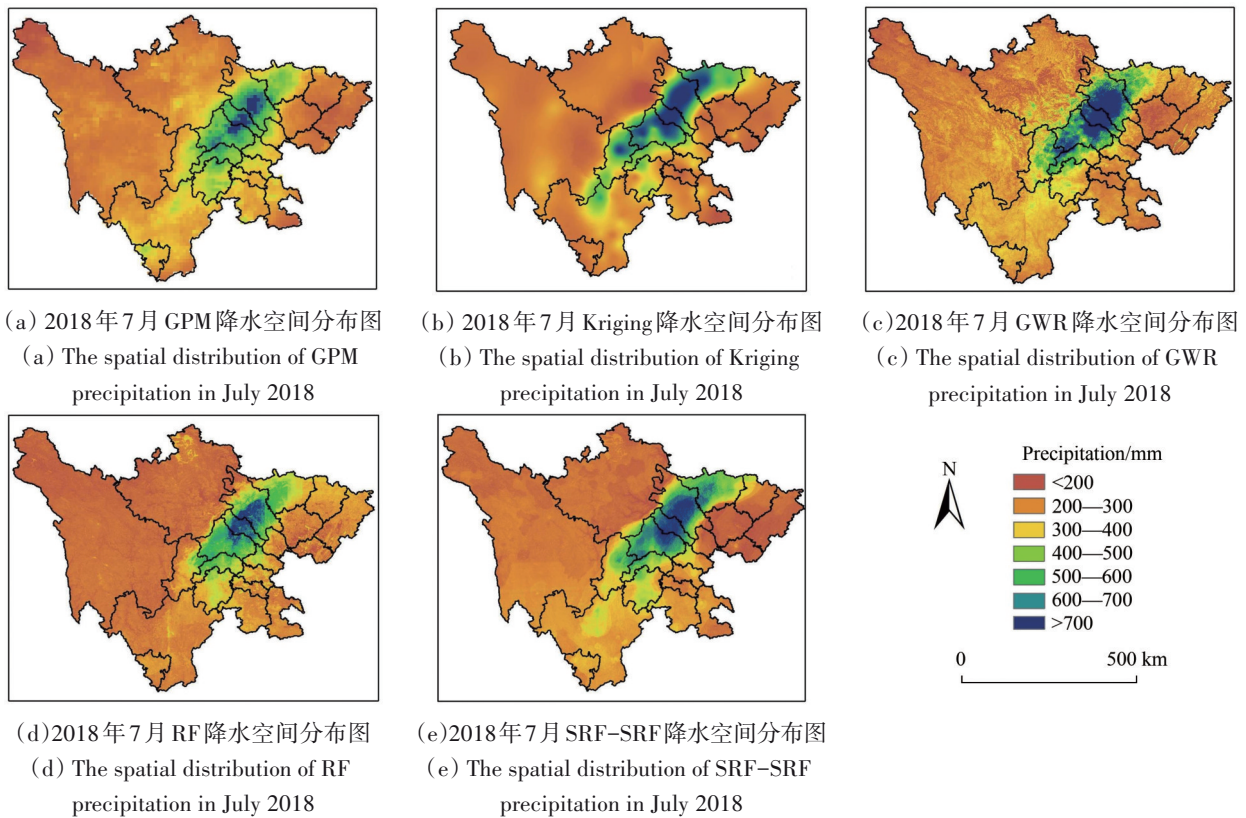


图6 2018年7月降水空间分布图  
 Fig. 6 The spatial distribution of precipitation in July 2018

## 5 讨论

### 5.1 RF模型变量重要性分析

降水是大气与环境相互作用的结果，因此降

水与当地的地形、植被等因素存在着较高的相关性。地理空间位置与地形因素以及植被因素一直以来是研究降水降尺度的常用变量（李净和张晓，2015），然而，在一些高海拔地区，受积雪覆盖等影响，NDVI和地形因素与降水之间的关系往往不



太明显。LST在植被稀少的地区与降水之间存在着显著关系 (Jing 等, 2016), 即降水能影响局部地区的温度变化。因此本方法考虑了空间位置、地形因素、植被因素与地表温度作为降水降尺度的环境变量。基于RF模型的相对变量重要性 (图7) 分析表明, Kriging插值降水的重要性最大, 其次为SRF降尺度值, 说明研究区域站点实测降水之间具有显著的相关性。地表温度的重要性要高于NDVI, 可能研究区域中高海拔地区较多, 温度较低不利于植被生长, 难以有效反映降水和NDVI关系。地形因子中坡向的重要性最低, 其次是坡度。而空间位置对降水的影响也很重要, 这与研究区域所处的气候有很大关系。

### 5.2 NDVI滞后性影响分析

有研究表明, NDVI可能对降水存在3个月以上的滞后性 (胡实 等, 2020; Karbalaye Ghorbanpour 等, 2021)。为验证这种长时间滞后性是否对降尺度融合产生影响, 本研究将年均NDVI融入到年降水降尺度中, 然后按月降水占的比例获得月降水降尺度结果, 最终进行SRF点面融合 (即SRFdis)。图8给出了2015年—2019年月尺度SRF-SRF降水预估值与SRFdis降水预估值的散点分布图。两种方法的离散程度与实测降水的拟合

程度都接近于1:1线, 与原始GPM (表1) 相比均显著提高了GPM降水的准确性。SRF-SRF与SRFdis相比, 前者的MAE和RMSE值比后者分别降低了0.36 mm和0.76 mm。因此, 与年平均NDVI作为环境变量相比, 选择月NDVI值可以有效提高计算结果精度。

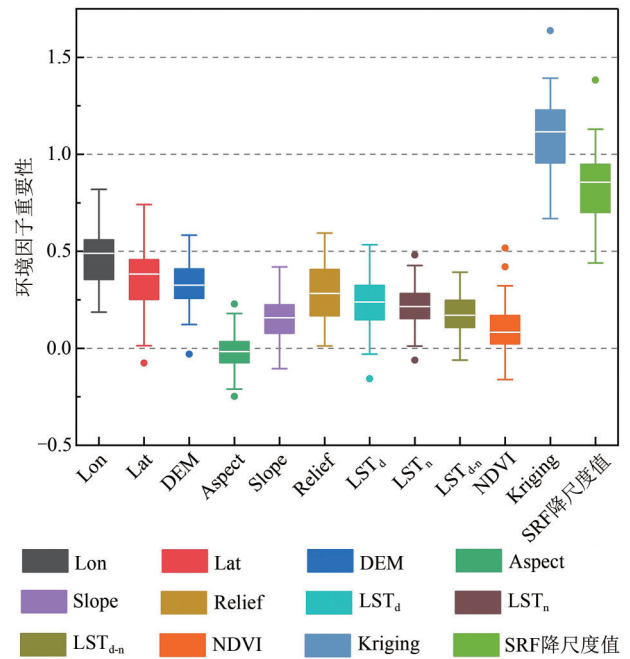
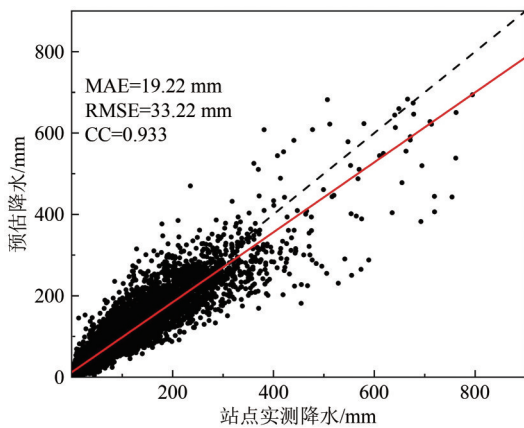
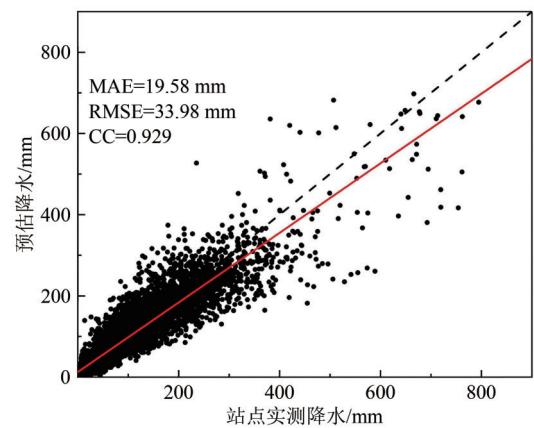


图7 SRF模型中各变量的重要性

Fig. 7 The importance of variables in the SRF model



(a) SRF-SRF降水预估值与实测降水散点分布图  
(a) Scatterplot of SRF-SRF and Measured Precipitation



(b) SRFdis降水预估值与实测降水散点分布图  
(b) Scatterplot of SRFdis and Measured Precipitation

图8 SRF-SRF与SRFdis降水预估值与实测降水散点分布图

Fig. 8 Comparison between SRF-SRF and SRFdis

## 6 结论

为提高GPM IMERG遥感降水产品空间分辨率与准确性, 本文以RF模型为基础, 构造了一种顾

及空间相关性的空间随机森林模型, 主要包括降尺度和点面融合两个阶段。首先使用SRF模型并结合降水之间的空间相关性以及环境变量对GPM IMERG降尺度, 然后以实测降水产品为基准再次

基于SRF模型对降尺度后的结果点面融合, 获得高空间分辨率和高精度降水产品。实验结果表明:

(1) 在复杂地形区域考虑降水空间相关性可有效提升降水产品空间分布细节特征和降水的准确性, 并且通过对RF模型变量重要性分析得出kriging插值信息重要性最大, 说明了研究区域内降水之间的空间相关性不可忽视。该方法解决了研究区域内因降水产品空间分辨率粗糙和准确度低等导致难以开展精细化水文研究等难题, 为降水产品精细化研究提供了技术支持。

(2) 为验证本文方法的有效性, 将该方法实验结果与其他经典方法结果加以对比, 结果表明不同时间尺度上该方法表现效果均优于其他方法, 并且本文方法的预估值与实测降水更具一致性, 说明在降尺度与点面融合两阶段都考虑与降水相关的影响因素可提高结果精度。

后续研究中, 将使用其他高时空分辨率降水驱动因素(如土壤湿度、地表温度、风速等)并融合不同遥感降水产品, 进一步提升遥感降水产品时空分辨率(如逐日、逐小时)和精度。

## 参考文献(References)

- Baez-Villanueva O M, Zambrano-bigiarini M, Beck H E, Mcnamara I, Ribbe L, Nauditt A, Birkel C, Verbist K, Giraldo-Osorio J D and Xuan Thinh N. 2020. RF-MEP: a novel random forest method for merging gridded precipitation products and ground-based measurements. *Remote Sensing of Environment*, 239: 111606 [DOI: 10.1016/j.rse.2019.111606]
- Brodeur Z P and Steinschneider S. 2020. Spatial bias in medium-range forecasts of heavy precipitation in the Sacramento River Basin: implications for water management. *Journal of Hydrometeorology*, 21(7): 1405-1423 [DOI: 10.1175/JHM-D-19-0226.1]
- Chen C F and Li Y Y. 2019. A fast global interpolation method for digital terrain model generation from large LiDAR-derived data. *Remote Sensing*, 11(11): 1324 [DOI: 10.3390/rs11111324]
- Fu Y, Xia J Z, Yuan W P, Xu B, Wu X X, Chen Y and Zhang H C. 2016. Assessment of multiple precipitation products over major river basins of China. *Theoretical and Applied Climatology*, 123(1/2): 11-22 [DOI: 10.1007/s00704-014-1339-0]
- Hu Q F, Yang D W, Li Z, Mishra A K, Wang Y T and Yang H B. 2014. Multi-scale evaluation of six high-resolution satellite monthly rainfall estimates over a humid region in China with dense rain gauges. *International Journal of Remote Sensing*, 35(4): 1272-1294 [DOI: 10.1080/01431161.2013.876118]
- Hu S, Han J, Zhan C S and Liu L M Z. 2020. Spatial downscaling of remotely sensed precipitation in Taihang Mountains. *Geographical Research*, 39(7): 1680-1690 (胡实, 韩建, 占车生, 刘梁美子. 2020. 太行山区遥感卫星反演降雨产品降尺度研究. *地理研究*, 39(7): 1680-1690) [DOI: 10.11821/dlyj020190545]
- Immerzeel W W, Rutten M M and Droogers P. 2009. Spatial downscaling of TRMM precipitation using vegetative response on the Iberian Peninsula. *Remote Sensing of Environment*, 113(2): 362-370 [DOI: 10.1016/j.rse.2008.10.004]
- Islam M A, Yu B F and Cartwright N. 2020. Assessment and comparison of five satellite precipitation products in Australia. *Journal of Hydrology*, 590: 125474 [DOI: 10.1016/j.jhydrol.2020.125474]
- Ji T, Liu R, Yang H, He T R and Wu J F. 2015. Spatial downscaling of precipitation using multi-source remote sensing data: a case study of sichuan-chongqing region. *Journal of Geo-Information Science*, 17(1): 108-117 (嵇涛, 刘睿, 杨华, 何太蓉, 吴建峰. 2015. 多源遥感数据的降水空间降尺度研究——以川渝地区为例. *地球信息科学学报*, 17(1): 108-117) [DOI: 10.3724/SP.J.1047.2015.00108]
- Jia S F, Zhu W B, Lü A F and Yan T T. 2011. A statistical spatial downscaling algorithm of TRMM precipitation based on NDVI and DEM in the Qaidam Basin of China. *Remote Sensing of Environment*, 115(12): 3069-3079 [DOI: 10.1016/j.rse.2011.06.009]
- Jiang S H, Wei L Y, Ren L L, Xu C Y, Zhong F, Wang M H, Zhang L Q, Yuan F and Liu Y. 2021. Utility of integrated IMERG precipitation and GLEAM potential evapotranspiration products for drought monitoring over mainland China. *Atmospheric Research*, 247: 105141 [DOI: 10.1016/j.atmosres.2020.105141]
- Jing W L, Yang Y P, Yue X F and Zhao X D. 2016. A spatial downscaling algorithm for satellite-based precipitation over the tibetan plateau based on NDVI, DEM, and land surface temperature. *Remote Sensing*, 8(8): 655 [DOI: 10.3390/rs8080655]
- Jongjin B, Jongmin P, Dongryeol R and Minha C. 2016. Geospatial blending to improve spatial mapping of precipitation with high spatial resolution by merging satellite-based and ground-based data. *Hydrological Processes*, 30(16): 2789-2803 [DOI: 10.1002/hyp.10786]
- Karbalaye Ghorbanpour A, Hessels T, Moghim S and Afshar A. 2021. Comparison and assessment of spatial downscaling methods for enhancing the accuracy of satellite-based precipitation over Lake Urmia Basin. *Journal of Hydrology*, 596: 126055 [DOI: 10.1016/J.JHYDROL.2021.126055]
- Kim K H, Kim M S, Lee G W, Kang D H and Kwon B H. 2013. The adjustment of radar precipitation estimation based on the kriging method. *Journal of the Korean Earth Science Society*, 34(1): 13-27 [DOI: 10.5467/JKES.2013.34.1.13]
- Lai X and Gong Y F. 2017. Relationship between atmospheric heat source over the Tibetan Plateau and precipitation in the Sichuan-Chongqing region during summer. *Journal of Meteorological Research*, 31(3): 555-566 [DOI: 10.1007/s13351-017-6045-2]
- Li J and Zhang X. 2015. Downscaling method of TRMM satellite precipitation data. *Scientia Geographica Sinica*, 35(9): 1164-1169 (李净, 张晓. 2015. TRMM降水数据的空间降尺度方法研究. *地理科学*, 35(9): 1164-1169) [DOI: 10.13249/j.cnki.sgs.2015.09.1164]
- Lu X Y, Tang G Q, Wang X Q, Liu Y, Jia L H, Xie G H, Li S and Zhang Y X. 2019. Correcting GPM IMERG precipitation data

- over the Tianshan Mountains in China. *Journal of Hydrology*, 575: 1239-1252 [DOI: 10.1016/j.jhydrol.2019.06.019]
- Lu X Y, Tang G Q, Wang X Q, Liu Y, Wei M and Zhang Y X. 2020. The development of a two-step merging and downscaling method for satellite precipitation products. *Remote Sensing*, 12(3): 398 [DOI: 10.3390/rs12030398]
- Lu X Y, Wei M and Wang X Q. 2017. Correction of TRMM monthly precipitation data from 1998 to 2013 in Xinjiang. *Journal of Applied Meteorological Science*, 28(3): 379-384 (卢新玉, 魏鸣, 王秀琴. 2017. TRMM月降水量产品在新疆地区的订正. *应用气象学报*, 28(3): 379-384) [DOI: 10.11898/1001-7313.20170311]
- Ma J H, Qu C, Zhang H X and Xia Y Q. 2013. Spatial downscaling of TRMM precipitation data based on DEM in the upstream of Shiyang River Basin during 2001-2010. *Progress in Geography*, 32(9): 1423-1432 (马金辉, 屈创, 张海筱, 夏燕秋. 2013. 2001-2010年石羊河流域上游TRMM降水资料的降尺度研究. *地理科学进展*, 32(9): 1423-1432) [DOI: 10.11820/dlkxjz.2013.09.012]
- Ma Z Q, He K, Tan X, Xu J T, Fang W Z, He Y and Hong Y. 2018. Comparisons of spatially downscaling TMPA and IMERG over the Tibetan Plateau. *Remote Sensing*, 10(12): 1883 [DOI: 10.3390/rs10121883]
- Markonis Y, Papalexio S M, Martinkova M and Hanel M. 2019. Assessment of water cycle intensification over land using a multi-source global gridded precipitation DataSet. *Journal of Geophysical Research: Atmospheres*, 124(21): 11175-11187 [DOI: 10.1029/2019JD030855]
- Njuki S M, Mannaerts C M and Su Z B. 2020. An improved approach for downscaling coarse-resolution thermal data by minimizing the spatial averaging biases in random forest. *Remote Sensing*, 12(21): 3507 [DOI: 10.3390/rs12213507]
- Pan Y, Shen Y, Yu J J and Zhao P. 2012. Analysis of the combined gauge-satellite hourly precipitation over China based on the OI technique. *Acta Meteorologica Sinica*, 70(6): 1381-1389 (潘咏, 沈艳, 宇婧婧, 赵平. 2012. 基于最优插值方法分析的中国区域地面观测与卫星反演逐时降水融合试验. *气象学报*, 70: 1381-1389) [DOI: 10.11676/qxxb2012.116]
- Renard B, Kavetski D, Leblois E, Thyer M, Kuczera G and Franks S W. 2011. Toward a reliable decomposition of predictive uncertainty in hydrological modeling: characterizing rainfall errors using conditional simulation. *Water Resources Research*, 47(11): W11516 [DOI: 10.1029/2011WR010643]
- Sekulić A, Kilibarda M, Heuvelink G B M, Nikolić M and Bajat B. 2020. Random forest spatial interpolation. *Remote Sensing*, 12(10): 1687 [DOI: 10.3390/rs12101687]
- Sharifi E, Saghafian B and Steinacker R. 2019. Downscaling satellite precipitation estimates with multiple linear regression, artificial neural networks, and spline interpolation techniques. *Journal of Geophysical Research: Atmospheres*, 124(2): 789-805 [DOI: 10.1029/2018JD028795]
- Shen Y, Xiong A Y, Hong Y, Yu J J, Pan Y, Chen Z Q and Saharia M. 2014a. Uncertainty analysis of five satellite-based precipitation products and evaluation of three optimally merged multi-algorithm products over the Tibetan Plateau. *International Journal of Remote Sensing*, 35(19): 6843-6858 [DOI: 10.1080/01431161.2014.960612]
- Shen Y, Zhao P, Pan Y and Yu J J. 2014b. A high spatiotemporal gauge-satellite merged precipitation analysis over China. *Journal of Geophysical Research: Atmospheres*, 119(6): 3063-3075 [DOI: 10.1002/2013jd020686]
- Shi W J, Yue T X, Shi X L and Song W. 2012. Research progress in soil property interpolators and their accuracy. *Journal of Natural Resources*, 27(1): 163-175 (史文娇, 岳天祥, 石晓丽, 宋伟. 2012. 土壤连续属性空间插值方法及其精度的研究进展. *自然资源学报*, 27(1): 163-175) [DOI: 10.11849/zrzyxb.2012.01.017]
- Song F, Hu Q and Qian W H. 2004. Quality control of daily meteorological data in China, 1951-2000: a new dataset. *International Journal of Climatology*, 24(7): 853-870 [DOI: 10.1002/joc.1047]
- Sun R C, Yuan H L, Liu X L and Jiang X M. 2016. Evaluation of the latest satellite-gauge precipitation products and their hydrologic applications over the Huaihe River Basin. *Journal of Hydrology*, 536: 302-319 [DOI: 10.1016/j.jhydrol.2016.02.054]
- Wang Y D, Nan Z T, Chen H and Wu X B. 2016. Correction of CMORPH daily precipitation data over the Qinghai-Tibetan Plateau with K-nearest neighbor model. *Remote Sensing Technology and Application*, 31(3): 607-616 (王玉丹, 南卓铜, 陈浩, 吴小波. 2016. 基于K最近邻模型的青藏高原CMORPH日降水数据的订正研究. *遥感技术与应用*, 31(3): 607-616) [DOI: 10.11873/j.issn.1004-0323.2016.3.0607]
- Wei L Y, Jiang S H, Ren L L, Zhang L Q, Wang M H and Duan Z. 2020. Preliminary utility of the retrospective IMERG precipitation product for large-scale drought monitoring over mainland China. *Remote Sensing*, 12(18): 2993 [DOI: 10.3390/rs12182993]
- Xie P P and Xiong A Y. 2011. A conceptual model for constructing high-resolution gauge-satellite merged precipitation analyses. *Journal of Geophysical Research: Atmospheres*, 116(D21): D21106 [DOI: 10.1029/2011jd016118]
- Yang M X, Liu G D, Chen T, Chen Y and Xia C C. 2020. Evaluation of GPM IMERG precipitation products with the point rain gauge records over Sichuan, China. *Atmospheric Research*, 246: 105101 [DOI: 10.1016/j.atmosres.2020.105101]
- Zhang W, Liu D, Zheng S J, Liu S Y, Loáiciga H A and Li W K. 2020. Regional precipitation model based on geographically and temporally weighted regression kriging. *Remote Sensing*, 12(16): 2547 [DOI: 10.3390/rs12162547]
- Zhou Z T, Guo B, Su Y Z, Chen Z S and Wang J. 2020a. Multidimensional evaluation of the TRMM 3B43V7 satellite-based precipitation product in mainland China from 1998-2016. *PeerJ*, 8: e8615 [DOI: 10.7717/peerj.8615]
- Zhou Z T, Guo B, Xing W X, Zhou J, Xu F L and Xu Y. 2020b. Comprehensive evaluation of latest GPM era IMERG and GSMaP precipitation products over mainland China. *Atmospheric Research*, 246: 105132 [DOI: 10.1016/j.atmosres.2020.105132]

# Improving the quality of remotely sensed precipitation product from GPM satellites by using a spatial random forest

HU Baojian<sup>1</sup>, LI Wei<sup>2</sup>, CHEN Chuanfa<sup>1</sup>, HU Zhanzhan<sup>1</sup>

1. College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, China;

2. Huzhou Zhonghe Survey Planning and Design Co. LTD, Huzhou 313000, China

**Abstract:** Satellite remote-sensing precipitation products are currently the main source for obtaining large-scale and continuous precipitation observations. However, currently available satellite remote-sensing precipitation products have coarse spatial resolution and suffer from certain systematic biases. Thus, this paper aims to downscale the precipitation data and remove its inherent systematic biases.

This paper proposes a two-stage Spatial Random Forest (SRF) method (SRF-SRF) by fully considering the influence of high-resolution environmental variables (including topography, NDVI, surface temperature, latitude, and longitude) on the precipitation and the spatial correlation of neighboring remotely sensed precipitation (stations). Taking the Global Precipitation Measurement Mission (GPM) monthly precipitation data of Sichuan Province from 2015—2019 as an example, its quality is enhanced with the help of SRF-SRF. The calculation results are compared with those of seven existing methods, including Geo-Weighted Regression (GWR), Back-Propagation Neural Network (BPNN), Random Forest (RF), Kriging interpolation of station precipitation (Kriging), Geographic Difference Analysis correction after downscaling by SRF (SRF-GDA), SRF correction after downscaling by bilinear interpolation (Bi-SRF), and annual precipitation downscaled by SRF. Subsequently, the results are scaled by month and corrected using SRF (SRFDis).

Experimental analysis shows the following: (1) At the monthly scale, compared with the original GPM, the mean absolute error (MAE) of SRF-SRF is reduced by 19.51%, and the medium error (RMSE) is reduced by 16.35%. The accuracy is better than those of other methods. At the seasonal scale, SRF-SRF has the smallest error in winter and the largest error in summer, but its calculation accuracy is better than those of other methods. At the annual scale, the four SRF-based methods (including SRF-SRF, SRF-GDA, Bi-SRF, and SRFdis) outperform GWR, BPNN, and RF. The accuracy of SRF-SRF is higher than that of Bi-SRF and SRF-GDA. (2) The spatial-distribution continuity of SRF-SRF precipitation products is better, and the local precipitation details are significantly improved. (3) The spatial correlation of precipitation plays an important role in the improvement in GPM precipitation quality. (4) SRF-SRF based on the monthly scale is better than SRFdis based on the annual scale. This finding indicates that NDVI can be used for precipitation-quality enhancement at the monthly scale in Sichuan province.

This paper proposes a two-stage satellite precipitation product-quality enhancement method that considers spatial correlation. The method takes into account the spatial autocorrelation between precipitation and combines downscaling and calibration while integrating environmental factors. Accordingly, the spatial resolution and accuracy of precipitation products improve. Experimental results show that the new method outperforms the other seven classical methods and is more applicable to the quality improvement of precipitation products in complex terrain.

**Key words:** remote sensing, precipitation, downscaling, point and surface fusion, random forest, GPM, machine learning

**Supported by** Shandong Provincial Natural Science Foundation, China (No. ZR2020YQ26, ZR2019MD007); A Project of Shandong Province Higher Educational Youth Innovation Science and Technology Program (No. 2019KJH007)