

doi: 10.6046/zrzyyg.2021009

引用格式: 杜懿, 王大洋, 王大刚. GPM 卫星降水产品空间降尺度研究——以贵州省为例[J]. 自然资源遥感, 2021, 33(4): 111-120. (Du Y, Wang D Y, Wang D G. Spatial downscaling of GPM precipitation products: A case study of Guizhou Province[J]. Remote Sensing for Natural Resources, 2021, 33(4): 111-120.)

GPM 卫星降水产品空间降尺度研究

——以贵州省为例

杜懿, 王大洋, 王大刚

(中山大学地理科学与规划学院, 广州 510275)

摘要: 为提高 GPM 卫星降水产品的空间分辨率, 扩展其应用范围, 以贵州省为研究区域, 通过建立空间降尺度模型对其进行降尺度研究。首先, 以经度、纬度、高程、坡度、坡向等地形因子为解释变量, 以原始 GPM 卫星降水数据为目标变量, 分别建立多元线性回归、地理加权回归、极限学习机、支持向量机、随机森林回归等降尺度模型; 然后对多年平均时间尺度进行应用与评价; 最后选择效果最佳的模型分别对典型年的年、月降水量进行空间降尺度研究。结果表明, 除随机森林回归模型外的其余 4 种空间降尺度模型均表现良好, 其中以多元线性回归模型表现最为稳定、效果最优; 多元线性回归模型的降尺度结果在观测精度和空间相关性上均有较大程度的提升。该研究可为贵州省提供高分辨率的网格化降水产品, 对区域水文气象研究等工作提供支持。

关键词: 贵州省; GPM; 空间降尺度; 多元线性回归; 地理加权回归; 极限学习机

中图法分类号: TP 79; P 468 **文献标志码:** A **文章编号:** 2097-034X(2021)04-0111-10

0 引言

近年来, 随着计算技术和遥测技术的发展, 气象卫星反演已经成为一种新的降水观测途径, 国际上也先后涌现出多种卫星降水产品, 如 GPCP(global precipitation climatology project), TRMM(tropical rainfall measuring mission), CMORPH(climate prediction center morphing technique), PERSIANN(precipitation estimation from remotely sensed information using artificial neural networks), GSMaP(global satellite mapping of precipitation), GPM(global precipitation measurement mission)等。由于卫星降水产品能够提供连续的降水时空分布信息, 且观测结果受地形和气候条件等的影响较小, 可以很好地弥补地面站点观测的不足, 目前已被广泛地应用于水文、气象、环境等领域的研究。在诸多无资料地区, 卫星反演降水甚至成为区域降水信息获取的唯一手段。然而, 受多种因素限制, 卫星降水产品的空间分辨率往往不高, 如 GPCP 产品的空间分辨率仅为 $1.5^\circ \times 1.5^\circ$; CMORPH, PERSIANN 和 TRMM 等产品的空间分

率也只有 $0.25^\circ \times 0.25^\circ$; GSMaP 和 GPM 等降水产品的空间分辨率相对较高, 达到了 $0.1^\circ \times 0.1^\circ$, 在大尺度区域的研究中总体表现较好, 但在小范围区域和小尺度流域的应用中仍然受到诸多限制, 其空间分辨率仍有待提高。GPM 作为接替 TRMM 的新一代卫星降水产品, 不仅观测范围更广, 时空分辨率也得到了大幅提高, 由于搭载了微波成像仪和双频降水雷达, 有效提高了对弱降水以及固态降水的探测能力^[1]。

目前, 常用的降尺度方法主要分为动力降尺度和统计降尺度两大类。其中统计降尺度方法由于计算量小、建模灵活等优点, 在国内外得到了广泛应用, 是现阶段用来提高卫星降水产品空间分辨率的主要手段^[2]。概括来看, 基于研究区域的地理、地形等特征而建立的统计降尺度模型在实际应用中效果较好, 且理论依据完备。该类降尺度方法的基本假定是目标变量和解释变量之间的相关关系具有空间平稳性, 即二者之间的相关关系与空间尺度大小无关。地形指数由于地理稳定性较高, 对区域特征的描述性较强, 因而常被用作解释变量。

现有研究多以 TRMM 降水产品为研究对

收稿日期: 2021-01-08; 修订日期: 2021-03-17

基金项目: 国家自然科学基金面上项目“基于降水分类的极端降水与气温关系及其未来变化研究”(编号: 51779278)资助。

第一作者: 杜懿(1994-), 男, 博士研究生, 专业方向为水文及水资源。Email: duy19930922@163.com。

通信作者: 王大刚(1975-), 男, 博士, 教授, 主要从事水文气象研究。Email: wangdag@mail.sysu.edu.cn。

象^[3-5],对新一代卫星降水产品 GPM 的降尺度研究却所见不多^[6]。此外,目前研究中用于建立统计降尺度模型的方法较为单调,多为多元线性回归^[7-9]、高次多项式回归^[10]和地理加权回归^[11-13]等,而具有强大函数逼近能力的机器学习方法却很少被用到,仅有的研究也主要集中在对随机森林回归方法的使用上^[14-15]。

鉴于此,本文以 2010—2019 年的 GPM 卫星降水数据为研究对象,以地形复杂且气象站点稀少的贵州省为研究区域,以经度、纬度、高程、坡度、坡向等地形因子为解释变量,综合使用多元线性回归模型、地理加权回归模型、极限学习机模型、支持向量机模型以及随机森林回归模型等来进行空间降尺度研究。

1 研究区概况与数据来源

1.1 研究区概况

贵州省地处中国西南内陆腹地,面积达 17.61 万 km²,地理位置介于 N24°37′~29°13′,E103°36′~109°35′。地形上,贵州省地处云贵高原东部,境内地势西高东低,从中部向东、南、北三面倾斜,平均海拔在 1 100 m 左右,全省地貌可分为高原、山地、丘陵和盆地等 4 种基本类型,其中 92.5% 的面积为山地和丘陵,平原面积极少。气候上,贵州省属亚热带湿润季风气候,降水丰富,雨热同期,全省多年平均温度为 15℃ 左右,多年平均降水量在 1 100~1 300 mm 之间。贵州省主要气象站点及高程分布如图 1 所示。

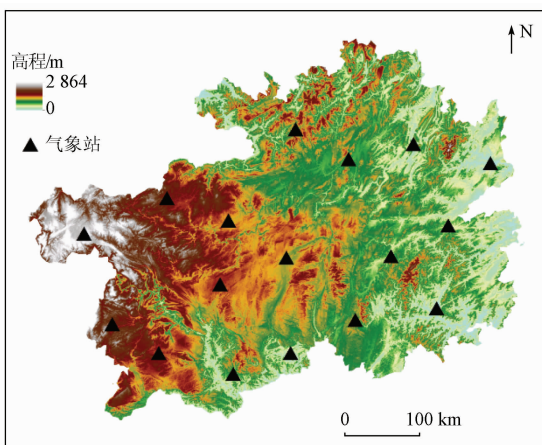


图 1 贵州省主要气象站点及高程分布

Fig. 1 Meteorological stations and elevation distribution in Guizhou Province

1.2 数据来源

根据反演算法的不同,GPM 能够提供 3 种级别

的遥感数据产品,其中三级 IMERG 产品是由校准后的微波所生成的红外降水估计,此外,还融合了地面观测数据,目前已更新至 V06B 版本。IMERG 产品中的 Final Run 质量最高,最适合于科学研究,该产品的空间覆盖范围为 S60°~N60°,空间分辨率为 0.1°×0.1°,时间分辨率为 0.5 h。本文使用的数据长度为 2010 年 1 月—2019 年 12 月,共 10 个完整年,120 个月份。

数字高程模型(digital elevation model, DEM)是目前用来描述区域地形地貌信息的主要手段。本文所用的 DEM 数据来自于地理空间数据云(<http://www.gscloud.cn/>)提供的 GDEM V2 30 m 分辨率原始高程数据。

地面气象站点观测数据来自国家气象科学数据中心(<http://data.cma.cn/>)发布的《中国地面气候资料月值数据集》。本文选用贵州省区域内的 17 个站点的 2010—2019 年间的月降水观测值。贵州省内各气象站点的基本信息如表 1 所列。

表 1 贵州省内气象站点基本信息

Tab. 1 Meteorological station basic information of Guizhou Province

站点	经度/(°)	纬度/(°)	高程/m	年均降水/mm
威宁	104.28	26.87	2 236.2	927.55
盘县	104.62	25.78	1 516.9	1 204.92
桐梓	106.83	28.13	972.0	977.88
毕节	105.28	27.30	1 514.4	897.35
湄潭	107.47	27.77	792.8	1 071.89
思南	108.25	27.95	417.7	1 030.32
铜仁	109.18	27.72	280.8	1 339.56
黔西	106.02	27.03	1 252.5	885.63
安顺	105.92	26.26	1 394.1	1 189.69
贵阳	106.72	26.58	1 074.3	1 158.04
凯里	107.98	26.60	722.6	1 241.95
三穗	108.67	26.97	611.0	1 161.98
兴仁	105.18	25.43	1 379.3	1 167.61
望谟	106.08	25.18	567.0	1 222.87
罗甸	106.77	25.43	441.5	1 170.62
独山	107.55	25.83	1 012.3	1 216.06
榕江	108.53	25.97	287.4	1 259.57

2 研究方法

本文的具体建模步骤如下^[16]:

1) 基于 30 m DEM 数据,以网格为单元,对研究区域内的各个地形因子(高程、坡度、坡向)进行重采样,分别采样成 0.1°×0.1°的低分辨率 D_{LR} 和 0.01°×0.01°的高分辨率 D_{HR} 。

2) 基于 GPM 卫星降水数据,对研究区域内各网

格所对应的降水量 P_{LR} 进行提取,其空间分辨率为 $0.1^\circ \times 0.1^\circ$ 。

3) 在 $0.1^\circ \times 0.1^\circ$ 低分辨率下,建立研究区域内降水量 P_{LR} 与地形因子 D_{LR} 之间的映射关系,即

$$P_{LR} = f(D_{LR})。 \quad (1)$$

4) 以 D_{HR} 为输入,以 f 为映射关系,即可得到 $0.01^\circ \times 0.01^\circ$ 高分辨率下研究区域内各网格的降水量 P_{HR} :

$$P_{HR} = f(D_{HR})。 \quad (2)$$

如此,就把降尺度问题转换成了回归问题,研究的关键在于如何找出最适当的映射关系 f 。回归模型的建模方法众多,本文主要使用了原理简单的多元线性回归、局部信息描述较好的地理加权回归、函数逼近能力优秀的机器学习模型等。

2.1 多元线性回归

多元线性回归原理较为简单,主要是利用统计学方法在多个变量之间建立如下的线性关系:

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_{k-1} X_{k-1} + a_k X_k, \quad (3)$$

式中: X 为解释变量; Y 为目标变量; a 为各解释变量所对应的偏回归系数; k 为解释变量的个数。已有研究表明^[17],叠加残差后的多元回归模型往往能够表现得更好。

2.2 地理加权回归

地理加权回归是一种局部参数估计方法,相较于多元线性回归,地理加权回归假设解释变量与目标变量之间的关系随空间位置的变化而变化,通过估算研究区内每一位置的目标变量与解释变量之间的参数来建立回归模型,不同于线性回归在整个研究区内通用一套回归参数^[18-20]。其原理可用以下公式进行描述:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^n \beta_j(u_i, v_i) x_{ij} + \varepsilon(u_i, v_i), \quad (4)$$

式中: n 为变量个数; y_i 为 i 点处的降水量, $i = 1, 2, \dots, k$; (u_i, v_i) 为 i 点处的地理坐标; $\beta_0(u_i, v_i)$ 为 i 点处的常数项回归参数; $\beta_j(u_i, v_i)$ 为第 j 个变量在 i 点处的回归参数; $\varepsilon(u_i, v_i)$ 为 i 点处的残差。

2.3 极限学习机

极限学习机是 Huang 等^[21-22]提出的一种单隐含层前馈神经网络。极限学习机的出现有效地解决了前馈神经网络学习速度慢的难题,该算法只需要

在网络训练前随机生成输入层和隐含层之间的连接权值和隐含层神经元的阈值,且整个训练过程中无需更新调整。

目前,该模型实现代码已公开发布,下载网址为 <https://www.ntu.edu.sg/home/egbhuang/>,用户可根据需要免费下载使用。该模型用于回归时,主要优选参数仅有一个,即隐含层的神经元个数。鉴于极限学习机运算效率较快,本文直接使用遍历法来优选参数。

2.4 支持向量机

支持向量机是由 Vapnik 等^[23]在统计学习理论的基础上提出的一种新的机器学习方法,其在很大程度上解决了过学习、非线性和维数灾等问题,为小样本机器学习问题建立了一个较好的理论框架^[23-24]。

本文基于 LIBSVM V3.24 开源工具箱来进行支持向量机的建模工作,该工具箱所涉及的参数调节较少,提供了很多默认参数。一般情况下,当采用 RBF 函数作为核函数时,需要优化的参数主要有惩罚系数 c 和 RBF 核函数的自带参数 g , LIBSVM V3.24 工具箱内嵌了交叉验证法来对参数 c 和 g 进行优选,但是考虑到 epsilon-SVR 模式中损失函数值 p 对模型的最终表现也有较大影响,故也将此参数作为待优选参数。由于遗传算法具有强大且高效的全局优化能力,本文以其来对 c, g, p 等参数进行寻优。

2.5 随机森林回归

随机森林是由 Breiman^[25]提出的一种基于决策树的机器学习模型,与神经网络算法相比,其计算量小且精度较高。该算法目前在多种环境下均可实现,如 MATLAB, Python 和 R 等,其中对模型结果影响较大的参数主要有决策树数目 n_{tree} (默认值为 500) 和指定节点中用于二叉树的变量个数 m_{try} (默认值为输入变量个数的二次方根或三分之一)。一般情况下, m_{try} 的取值对模型效果影响不大,需要进行优选的参数就只有 n_{tree} 。由于随机森林模型也有极快的计算效率,故本文直接对 n_{tree} 进行遍历寻优。

2.6 评价指标

本文选用的评价指标主要包括有平均绝对误差 (mean absolute error, MAE)、平均绝对百分比误差 (mean absolute percentage error, MAPE)、均方根误差 (root mean square error, RMSE) 以及皮尔逊相关系数 R , 前 3 个指标用来表征模型的模拟误差,后一个指标用来描述空间一致性。各评价指标的计算公式

如下所示:

$$MAE = \frac{1}{N} \sum_{i=1}^N |P_{i,f} - P_{i,o}|, \quad (5)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|P_{i,f} - P_{i,o}|}{P_{i,o}} \times 100\%, \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_{i,f} - P_{i,o})^2}, \quad (7)$$

$$R = \frac{\sum_{i=1}^N (P_{i,f} - \bar{P}_f)(P_{i,o} - \bar{P}_o)}{\sqrt{\sum_{i=1}^N (P_{i,f} - \bar{P}_f)^2 \sum_{i=1}^N (P_{i,o} - \bar{P}_o)^2}}, \quad (8)$$

式中: N 为研究区域内的网格总单元数; $P_{i,o}$ 为第 i 个格点的降水实测值, mm; $P_{i,f}$ 为第 i 个格点的降水模拟值, mm; \bar{P}_o 为全部格点的实测降水均值, mm; \bar{P}_f 为全部格点的降水模拟均值, mm。

3 研究结果

3.1 多年平均尺度

对 2010—2019 年间贵州省的多年平均、多年季节平均等时间尺度的降水量分别建立基于多元线性回归、地理加权回归、极限学习机、支持向量机、随机森林回归等降尺度模型, 并将各情景下的降尺度前、后的降水量与相应的实测站点的降水量进行比较。

基于研究区域范围, 在 $0.1^\circ \times 0.1^\circ$ 分辨率下共提取出 1 991 组有效样本, 每组样本的输入变量包括有网格的经度、纬度、高程、坡度、坡向等 5 个地形因子, 输出变量则为网格在指定时间尺度下的 GPM 降水量。在建立多元线性回归模型和地理加权回归模型时, 以全部 1 991 组样本计算模型, 再利用该模型对新的输入样本进行模拟, 最终得到 $0.01^\circ \times 0.01^\circ$ 分辨率下的 GPM 降尺度结果。而在机器学习模型的建模过程中, 通常需要先对模型的可行性进行验证, 进而才能对新的输入样本进行模拟。根据经验, 按照 4:1 的原则进行样本划分, 各时间尺度下的机器学习模型在建模时均取前 1 600 组样本组成训练集, 剩余 391 组样本组成验证集。模型有效性以 MAPE 指标来进行评价, 一般当 $MAPE \leq 20\%$ 时, 说明模型表现较好, 可以实际应用。

表 2 为 3 种机器学习模型 MAPE 结果, 可以看出, 除随机森林回归模型在多年冬季的验证期内模拟误差过大, 其余各机器学习模型在各时间尺度下均建模成功, 且模拟精度较高。

表 2 机器学习模型的 MAPE 计算结果

Tab. 2 The MAPE of machine learning models (%)

时间尺度	极限学习机		支持向量机		随机森林	
	训练期	验证期	训练期	验证期	训练期	验证期
多年平均	1.30	3.95	3.54	10.17	0.94	9.43
多年春季	2.28	3.78	8.50	6.46	1.87	14.07
多年夏季	1.61	7.37	14.10	15.33	0.92	4.47
多年秋季	2.23	10.64	8.12	8.72	1.14	6.49
多年冬季	6.71	9.09	6.63	11.53	4.91	28.36

表 3—7 给出各降尺度模型在多年平均、多年季节时间尺度下的降尺度效果评价, 其中 $0.1^\circ \times 0.1^\circ$ 分辨率表示降尺度前的 GPM 数据, $0.01^\circ \times 0.01^\circ$ 分辨率表示降尺度后的 GPM 数据。

表 3 各时间尺度下多元线性回归模型评价指标计算结果

Tab. 3 The results of multiple linear regression models under various time scales

时间尺度	分辨率	MAE/ mm	MAPE/ %	RMSE/ mm	R
多年平均	$0.1^\circ \times 0.1^\circ$	80.07	7.17	102.70	0.83
	$0.01^\circ \times 0.01^\circ$	71.07	6.39	94.37	0.88
多年春季	$0.1^\circ \times 0.1^\circ$	25.41	8.71	33.69	0.97
	$0.01^\circ \times 0.01^\circ$	24.71	8.36	33.39	0.97
多年夏季	$0.1^\circ \times 0.1^\circ$	49.50	9.60	57.22	0.79
	$0.01^\circ \times 0.01^\circ$	44.89	8.89	53.49	0.85
多年秋季	$0.1^\circ \times 0.1^\circ$	17.37	7.47	22.13	0.61
	$0.01^\circ \times 0.01^\circ$	15.92	6.91	20.24	0.70
多年冬季	$0.1^\circ \times 0.1^\circ$	10.30	16.66	12.21	0.90
	$0.01^\circ \times 0.01^\circ$	9.81	14.91	11.94	0.89

表 4 各时间尺度下地理加权回归模型评价指标计算结果

Tab. 4 The results of geographical weighted regression models under various time scales

时间尺度	分辨率	MAE/ mm	MAPE/ %	RMSE/ mm	R
多年平均	$0.1^\circ \times 0.1^\circ$	80.07	7.17	102.70	0.83
	$0.01^\circ \times 0.01^\circ$	75.12	6.73	96.89	0.85
多年春季	$0.1^\circ \times 0.1^\circ$	25.41	8.71	33.69	0.97
	$0.01^\circ \times 0.01^\circ$	26.58	8.95	34.89	0.97
多年夏季	$0.1^\circ \times 0.1^\circ$	49.50	9.60	57.22	0.79
	$0.01^\circ \times 0.01^\circ$	46.28	9.07	54.46	0.82
多年秋季	$0.1^\circ \times 0.1^\circ$	17.37	7.47	22.13	0.61
	$0.01^\circ \times 0.01^\circ$	16.21	6.95	20.64	0.66
多年冬季	$0.1^\circ \times 0.1^\circ$	10.30	16.66	12.21	0.90
	$0.01^\circ \times 0.01^\circ$	9.71	16.12	11.09	0.92

表 5 各时间尺度下极限学习机模型评价指标计算结果

Tab. 5 The results of extreme learning machine models under various time scales

时间尺度	分辨率	MAE/ mm	MAPE/ %	RMSE/ mm	R
多年平均	$0.1^\circ \times 0.1^\circ$	80.07	7.17	102.70	0.83
	$0.01^\circ \times 0.01^\circ$	76.41	6.92	94.01	0.86
多年春季	$0.1^\circ \times 0.1^\circ$	25.41	8.71	33.69	0.97
	$0.01^\circ \times 0.01^\circ$	22.88	7.85	31.10	0.96

(续表)

时间尺度	分辨率	MAE/ mm	MAPE/ %	RMSE/ mm	R
多年夏季	0.1° × 0.1°	49.50	9.60	57.22	0.79
	0.01° × 0.01°	46.66	9.29	57.14	0.82
多年秋季	0.1° × 0.1°	17.37	7.47	22.13	0.61
	0.01° × 0.01°	18.21	7.67	22.91	0.57
多年冬季	0.1° × 0.1°	10.30	16.66	12.21	0.90
	0.01° × 0.01°	10.45	16.82	13.08	0.88

表 6 各时间尺度下支持向量机模型评价指标计算结果

Tab.6 The results of support vector machine models under various time scales

时间尺度	分辨率	MAE/ mm	MAPE/ %	RMSE/ mm	R
多年平均	0.1° × 0.1°	80.07	7.17	102.70	0.83
	0.01° × 0.01°	73.63	6.71	87.01	0.79
多年春季	0.1° × 0.1°	25.41	8.71	33.69	0.97
	0.01° × 0.01°	25.41	9.57	31.55	0.92
多年夏季	0.1° × 0.1°	49.50	9.60	57.22	0.79
	0.01° × 0.01°	43.97	8.70	53.86	0.80
多年秋季	0.1° × 0.1°	17.37	7.47	22.13	0.61
	0.01° × 0.01°	14.55	6.25	19.24	0.65
多年冬季	0.1° × 0.1°	10.30	16.66	12.21	0.90
	0.01° × 0.01°	9.21	14.32	11.30	0.89

表 7 各时间尺度下随机森林回归模型评价指标计算结果

Tab.7 The results of random forest regression models under various time scales

时间尺度	分辨率	MAE/ mm	MAPE/ %	RMSE/ mm	R
多年平均	0.1° × 0.1°	80.07	7.17	102.70	0.83
	0.01° × 0.01°	68.13	6.14	93.31	0.71
多年春季	0.1° × 0.1°	25.41	8.71	33.69	0.97
	0.01° × 0.01°	19.16	5.99	31.61	0.89
多年夏季	0.1° × 0.1°	49.50	9.60	57.22	0.79
	0.01° × 0.01°	44.02	8.80	52.45	0.82
多年秋季	0.1° × 0.1°	17.37	7.47	22.13	0.61
	0.01° × 0.01°	17.77	7.53	22.45	0.45
多年冬季	0.1° × 0.1°	10.30	16.66	12.21	0.90
	0.01° × 0.01°	15.82	20.83	21.65	0.73

1) 多元线性回归模型。相较于降尺度前,降尺度后的结果不仅空间分辨率有了极大提高,观测精度也有较大提升。在多年平均和多年季节时间尺度下,降尺度后的高分辨率数据不仅误差减小明显,与站点实测数据的相关性也有了明显提高。在多年冬季时间尺度上,降尺度后的高分辨率数据与站点实测数据的相关性虽然没有提高,但基本稳定。

2) 地理加权回归模型。从观测误差上来看,除了多年春季时间尺度基本保持不变,其余 4 种时间尺度下的观测误差均有明显减小。从与站点

实测数据的相关性上来看,相较于多元线性回归模型,地理加权回归模型在全部时间尺度下的相关性均表现出明显的提高。

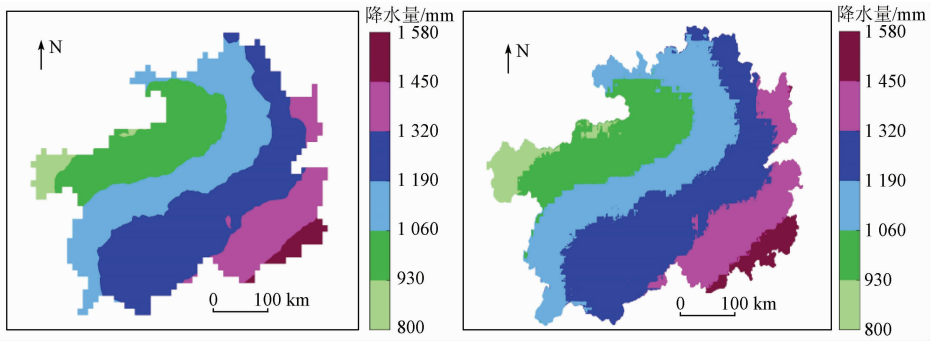
3) 极限学习机模型。基于极限学习机的降尺度模型也基本表现不错,但相较于多元线性回归和地理加权回归模型仍有一定差距。多年平均和多年夏季时间尺度下,降尺度后的数据无论是观测精度还是与站点实测降水的相关性,均有所提升;多年春季时间尺度下,虽然观测精度有所提高,但相关性有所减小;多年秋季和多年冬季时间尺度下,不仅观测精度有所下降,与站点实测数据的相关性也有所减小。

4) 支持向量机模型。基于支持向量机的降尺度模型整体表现较好,要优于极限学习机模型。在多年夏季和多年秋季时间尺度下,降尺度后的数据在精度和相关性上均有明显提升;多年平均和多年冬季时间尺度下,降尺度后的数据在观测精度上有较大的提高,但与站点实测数据的相关性却有所减小;多年春季时间尺度下,降尺度后的数据在观测精度和相关性两方面均存在微弱下降。

5) 随机森林回归模型。基于随机森林回归的降尺度模型表现不不理想。在多年夏季时间尺度下,降尺度后的数据在观测精度和相关性上均有明显提高;在多年平均、多年春季时间尺度下,虽然观测精度有一定程度的提高,但与站点实测数据的相关性却有所减弱;多年秋季和多年冬季时间尺度下,不仅观测误差上升,与站点实测数据的相关性也明显减小,尤其是多年冬季时间尺度的表现最差。造成这一结果的原因是在建模过程中模型未通过可行性检验,导致模型失效。

综合比较以上 5 种降尺度模型,可以发现多元线性回归模型、地理加权回归模型、极限学习机模型、支持向量机模型表现较好。进一步对比发现,多元线性回归模型和地理加权回归模型的降尺度效果整体上还是要优于极限学习机模型和支持向量机模型,可能的原因是在建模过程中前两个模型利用了全部样本信息(1 991 组),而后两种机器学习模型由于需要额外划出一部分样本来验证模型的可行性,造成了学习样本数(1 600 组)减少,样本学习信息不足,进而影响了降尺度结果。整体来看,多元线性回归模型原理最为简单、计算最为方便、效果也最为稳定。

图 2—6 为多元线性回归模型在多年平均和多年各季时间尺度下的降尺度前、后效果对比。

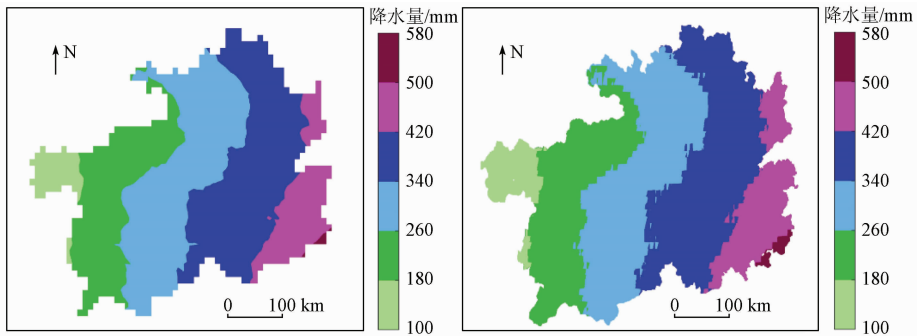


(a) 降尺度前

(b) 降尺度后

图 2 多年平均年降水量空间分布

Fig. 2 Spatial distribution of average annual precipitation

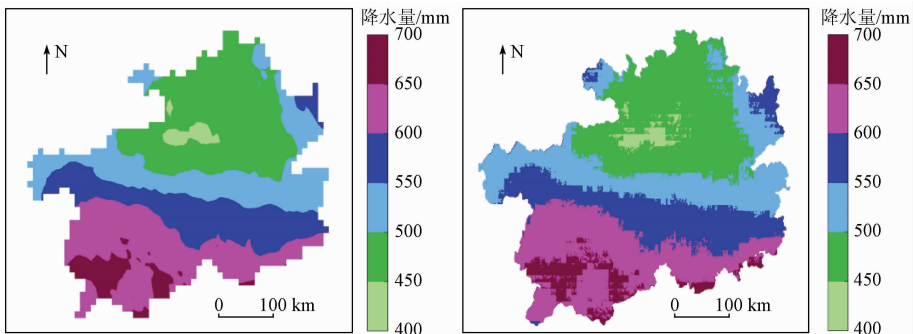


(a) 降尺度前

(b) 降尺度后

图 3 多年平均春季降水量

Fig. 3 Spatial distribution of average spring precipitation

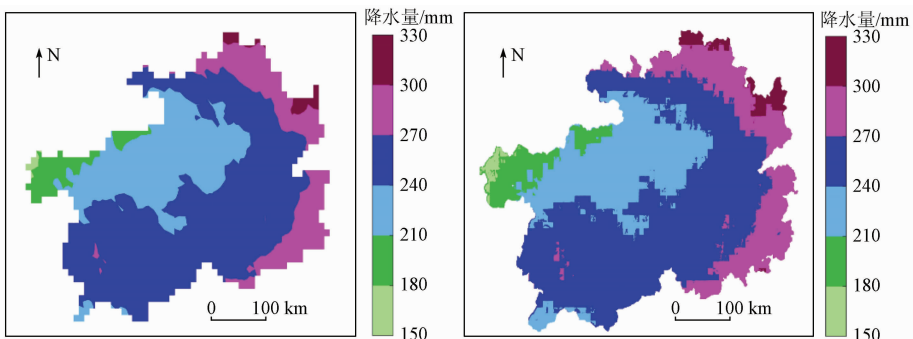


(a) 降尺度前

(b) 降尺度后

图 4 多年平均夏季降水量

Fig. 4 Spatial distribution of average summer precipitation

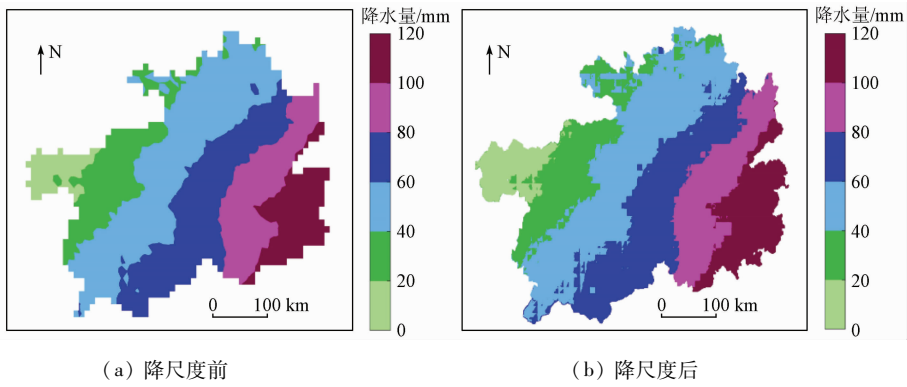


(a) 降尺度前

(b) 降尺度后

图 5 多年平均秋季降水量

Fig. 5 Spatial distribution of average autumn precipitation



(a) 降尺度前 (b) 降尺度后

图6 多年平均冬季降水量

Fig. 6 Spatial distribution of average winter precipitation

3.2 年尺度

本文使用的 GPM 降水数据的时间范围为 2010 年到 2019 年,采用算术平均法来计算研究区域的面平均降水量。图 7 为研究区域在 2010—2019 年期间的逐年降水量变化过程。可以明显看出,2011 年为典型干旱年,2014 年为典型湿润年。

采用多元线性回归模型分别对干、湿典型年的年降水量进行空间降尺度研究。表 8 为基于多元线性回归模型的降尺度结果评价。

表 8 干、湿典型年降水量降尺度效果评价

Tab. 8 Evaluation of downscaling results of typical annual precipitation of dry and wet

时间尺度	分辨率	MAE/ mm	MAPE/ %	RMSE/ mm	R
干旱典型年	0.1° × 0.1°	93.12	11.62	103.72	0.72
	0.01° × 0.01°	89.82	11.23	100.03	0.73
湿润典型年	0.1° × 0.1°	145.07	11.22	189.44	0.50
	0.01° × 0.01°	135.84	10.71	172.37	0.62

由表 8 可见,基于多元线性回归的降尺度模型在干、湿典型年中均表现较好。降尺度后的 GPM 数据,无论是在观测精度上还是在与站点实测数据的相关性上,均得到了不同程度的提高。此外,相较于典型干旱年,典型湿润年的降尺度效果更优。

图 8—9 为研究区域内典型干、湿年降尺度前、后的年降水量空间分布情况。

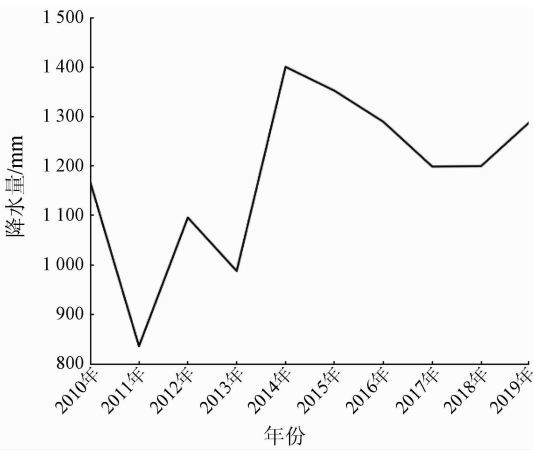
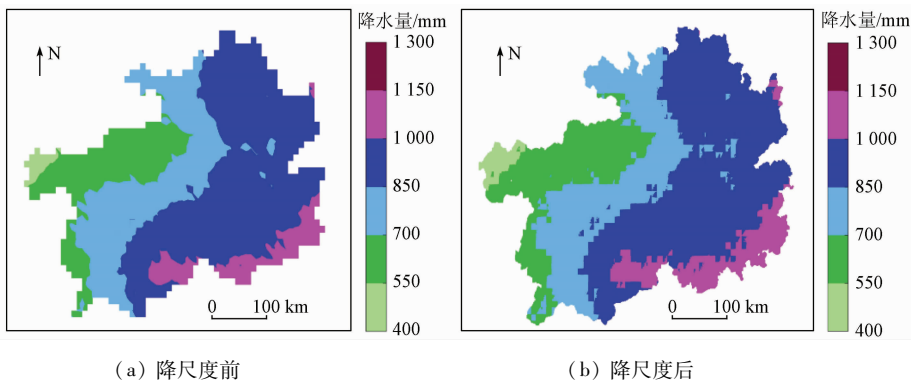


图7 研究区逐年降水量变化过程

Fig. 7 The change process of precipitation in the Guizhou Province from 2010 to 2019



(a) 降尺度前 (b) 降尺度后

图8 典型干旱年年降水量

Fig. 8 Annual precipitation in a typical drought year

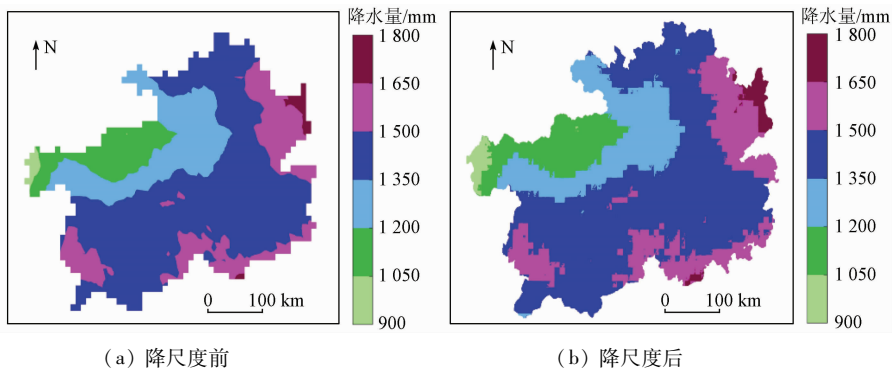


图 9 典型湿润年年降水量

Fig. 9 Annual precipitation in a typical wet year

3.3 月尺度

为了进一步验证多元线性回归降尺度模型在月尺度上的应用效果,分别对研究区域典型干、湿年的

月降水量进行降尺度处理。并通过观察降尺度前后降水的年内分配情况来检验模型的表现效果,结果如图 10 和图 11 所示。

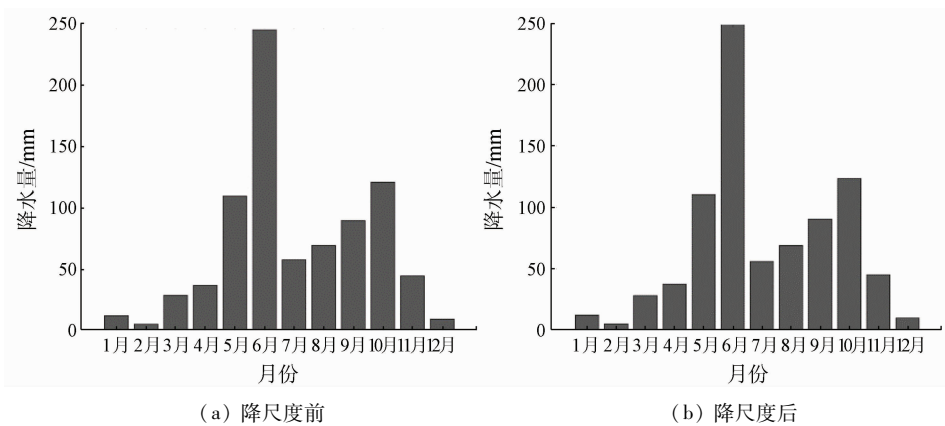


图 10 典型干旱年降水量月程分配

Fig. 10 Monthly distribution of precipitation in a typical dry year

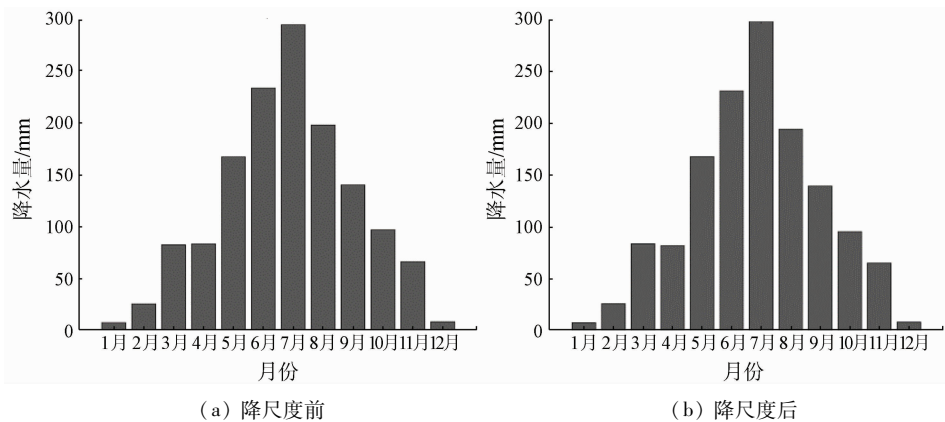


图 11 典型湿润年降水量月程分配

Fig. 11 Monthly distribution of precipitation in a typical wet year

从图 10 和图 11 中可以看出,两种典型年下,降尺度后的月降水量无论是在量值上还是在月程分配上均和降尺度前保持了较高的一致性,可以充分地说明基于多元线性回归的降尺度模型在贵州省区域有着相当好的适用性与可靠性。

降水数据为研究对象,借助经度、纬度、高程、坡度和坡向等地形因子来建立空间降尺度模型,主要取得以下结论:

1) 在多年时间尺度的降尺度研究中,多元线性回归、地理加权回归、极限学习机、支持向量机等模型均表现较好,其中以多元线性回归模型表现最好。鉴于多元线性回归模型原理简单、建模方便,且在所有模型中对观测精度的提高最为明显,故将其选为贵州省区域 GPM 卫星降水数据的最佳空间降尺度

4 结论

本文以 2010—2019 年贵州省区域的 GPM 卫星

模型。

2) 基于多元线性回归模型, 分别对研究区典型干、湿年的年降水量进行降尺度研究, 相较于降尺度前, 降尺度后的降水数据除了具有更高的分辨率外, 还具有更高的观测精度以及更好的相关性。典型干旱年下, 平均绝对百分比误差从 11.62% 下降到了 11.23%, 与站点实测数据的相关系数从 0.72 上升到了 0.73; 典型湿润年下, 平均绝对百分比误差从 11.22% 下降到了 10.71%, 与站点实测数据的相关系数从 0.50 上升到了 0.62。

3) 为了进一步检验多元线性回归降尺度模型在月时间尺度上的应用效果, 分别对典型干、湿年的月降水量进行了降尺度处理, 效果仍然令人满意。在提高降水数据空间分辨率的同时, 无论是在量值上还是在月程分配上, 均保持着较高的一致性。

机器学习模型在本次研究中的表现并未优于多元线性回归模型和地理加权回归模型, 可能的原因是研究区域较小, 可利用样本数较少, 模型无法得到充分的训练, 导致学习能力有限。随着研究区域的扩大、样本数的增多, 在经过大数据训练后的机器学习模型相信能获得更好的降尺度效果。

本文在建立降尺度模型的过程中仅使用了地形因子变量, 虽然在各个时间尺度上均能得到较好的应用效果, 但并不表明没有进一步提升的空间, 如引入植被指数、地表温度、土壤含水量等变量对降尺度效果是否会有改进, 仍需要继续开展研究。

参考文献 (References):

- [1] 史岚, 何其全, 杨娇, 等. 闽浙赣地区 GPM IMERG 降水产品降尺度建模与比较分析[J]. 地球信息科学学报, 2019, 21(10): 1642-1652.
Shi L, He Q Q, Yang J, et al. Downscaling modeling of the GPM IMERG precipitation product and comparative analysis in the Fujian-Zhejiang-Jiangxi region[J]. Journal of Geo-Information Science, 2019, 21(10): 1642-1652.
- [2] 唐国强, 万玮, 曾子悦, 等. 全球降水测量(GPM)计划及其最新进展综述[J]. 遥感技术与应用, 2015, 30(4): 607-615.
Tang G Q, Wan W, Zeng Z Y, et al. An overview of the global precipitation measurement (GPM) mission and its latest development[J]. Remote Sensing Technology and Application, 2015, 30(4): 607-615.
- [3] Immerzeel W W, Rutten M M, Droogers P. Spatial downscaling of TRMM precipitation using vegetative response on the Iberian Peninsula[J]. Remote Sensing of Environment, 2009, 113(2): 362-370.
- [4] Jia S F, Zhu W B, Lyu A F, et al. A statistical spatial downscaling algorithm of TRMM precipitation based on NDVI and DEM in the Qaidam basin of China[J]. Remote sensing of Environment, 2011, 115(12): 3069-3079.
- [5] Zhang Y Y, Li Y G, Ji X, et al. Fine-resolution precipitation mapping in a mountainous watershed: Geostatistical downscaling of TRMM products based on environmental variables[J]. Remote Sensing, 2018, 10(1): 1-27.
- [6] 胡实, 韩建, 占车生, 等. 太行山区遥感卫星反演降雨产品降尺度研究[J]. 地理研究, 2020, 39(7): 1680-1690.
Hu S, Han J, Zhan C S, et al. Spatial downscaling of remotely sensed precipitation in Taihang Mountains[J]. Geographical Research, 2020, 39(7): 1680-1690.
- [7] 马金辉, 屈创, 张海筱, 等. 2001—2010年石羊河流域上游 TRMM 降水资料的降尺度研究[J]. 地理科学进展, 2013, 32(9): 1423-1432.
Ma J H, Qu C, Zhang H X, et al. Spatial downscaling of TRMM precipitation data based on DEM in the upstream of Shiyang River basin during 2001—2010[J]. Progress in Geography, 2013, 32(9): 1423-1432.
- [8] 蔡明勇, 吕洋, 杨胜天, 等. 雅鲁藏布江流域 TRMM 降水数据降尺度研究[J]. 北京师范大学学报(自然科学版), 2017, 53(1): 111-119, 2.
Cai M Y, Lv Y, Yang S T, et al. TRMM precipitation downscaling in the data scarce Yarlung Zangbo River basin[J]. Journal of Beijing Normal University (Natural Science), 2017, 53(1): 111-119, 2.
- [9] 吴健, 周秋文, 韦小茶, 等. TRMM 降水产品在喀斯特地区的空间降尺度研究——以贵州省为例[J]. 人民珠江, 2019, 40(6): 20-26, 53.
Wu J, Zhou Q W, Wei X C, et al. Spatial downscaling research of TRMM precipitation products in karst region: A case study of Guizhou province[J]. Pearl River, 2019, 40(6): 20-26, 53.
- [10] 樊东, 薛华柱, 董国涛, 等. 基于二次多项式回归模型的黑河流域 TRMM 数据降尺度研究[J]. 水土保持研究, 2017, 24(2): 146-151.
Fan D, Xue H Z, Dong G T, et al. Downscaling study on TRMM 3B43 data of the Heihe river basin based on quadratic polynomial regression model[J]. Research of Soil and Water Conservation, 2017, 24(2): 146-151.
- [11] 刘小婵, 张洪岩, 赵建军, 等. 东北地区 TRMM 数据降尺度的 GWR 模型分析[J]. 地球信息科学学报, 2015, 17(9): 1055-1062.
Liu X C, Zhang H Y, Zhao J J, et al. Spatial downscaling of TRMM precipitation data based on GWR model in Northeast China[J]. Journal of Geo-Information Science, 2015, 17(9): 1055-1062.
- [12] 曾业隆, 谭伟, 王超, 等. 基于 GWR 模型的贵州喀斯特山区 TRMM 3B43 降水资料降尺度分析[J]. 干旱气象, 2018, 36(3): 405-414.
Zeng Y L, Tan W, Wang C, et al. Spatial downscaling of TRMM 3B43 precipitation data based on GWR model in karst mountainous area of Guizhou Province[J]. Journal of Arid Meteorology, 2018, 36(3): 405-414.
- [13] 熊俊楠, 李伟, 刘志奇, 等. 基于 GWR 模型的青藏高原地区 TRMM 数据降尺度研究[J]. 国土资源遥感, 2019, 31(4): 88-95. doi: 10.6046/gtzyyg.2019.04.12.
Xiong J N, Li W, Liu Z Q, et al. Research on downscaling of TRMM data in the Tibetan Plateau based on GWR model[J]. Remote Sensing for Land and Resources, 2019, 31(4): 88-95. doi:

- 10.6046/gtzyyg.2019.04.12.
- [14] 徐彬仁,魏瑗瑗.基于随机森林算法对青藏高原 TRMM 降水数据进行空间统计降尺度研究[J].国土资源遥感,2018,30(3):181-188. doi:10.6046/gtzyyg.2018.03.25.
- Xu B R, Wei Y Y. Spatial statistics of TRMM precipitation in the Tibetan Plateau using random forest algorithm[J]. Remote Sensing for Land and Resources, 2018, 30(3):181-188. doi:10.6046/gtzyyg.2018.03.25.
- [15] 沈思民,赵军,刘佳茹,等.祁连山 TRMM 降水数据降尺度不同方法比较研究[J].山地学报,2019,37(6):923-931.
- Shen S M, Zhao J, Liu J R, et al. Comparative study on different downscaling methods of TRMM satellite precipitation data over the Qilian Mountains, China[J]. Mountain Research, 2019, 37(6):923-931.
- [16] 范雪薇,刘海隆.天山山区 TRMM 降水数据的空间降尺度研究[J].自然资源学报,2018,33(3):478-488.
- Fan X W, Liu H L. Downscaling method of TRMM satellite precipitation data over the Tianshan Mountains[J]. Journal of Natural Resources, 2018, 33(3):478-488.
- [17] 李净,张晓. TRMM 降水数据的空间降尺度方法研究[J].地理科学,2015,35(9):1164-1169.
- Li J, Zhang X. Downscaling method of TRMM satellite precipitation data[J]. Scientia Geographica Sinica, 2015, 35(9):1164-1169.
- [18] Brunson C, Fotheringham A S, Charlton M E. Geographically weighted regression: A method for exploring spatial nonstationarity[J]. Geographical Analysis, 1999, 28(4):281-298.
- [19] 雷蕾,程志刚,冯冬蕾,等.近16年秦巴山区 TRMM 降水资料的降尺度研究[J].成都信息工程大学学报,2018,33(2):190-196.
- Lei L, Cheng Z G, Feng D L, et al. Statistical downscaling of TRMM precipitation data in Qinling-Daba Mountains area in recent 16 years[J]. Journal of Chengdu University of Information Technology, 2018, 33(2):190-196.
- [20] 李豪,雷苑.复杂地形下 TRMM 降水数据的降尺度研究:四川省为例[J].中国农业气象,2019,40(10):607-619.
- Li H, Lei Y. Spatial downscaling of TRMM precipitation data in areas of complex terrain: A case study in Sichuan Province[J]. Chinese Journal of Agrometeorology, 2019, 40(10):607-619.
- [21] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and applications[J]. Neurocomputing, 2006, 70(1):489-501.
- [22] Huang G B, Chen L. Letters: Convex incremental extreme learning machine[J]. Neurocomputing, 2007, 70(16):3056-3062.
- [23] Cortes C, Vapnik V. Support-vector networks[J]. Machine Learning, 1995(20):273-297.
- [24] Vapnik V. The Nature of statistical learning theory[M]. New York: Springer Verlag, 1995.
- [25] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1):5-32.

Spatial downscaling of GPM precipitation products: A case study of Guizhou Province

DU Yi, WANG Dayang, WANG Dagang

(School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China)

Abstract: To improve the spatial resolution and expand the application scopes of GPM precipitation products, the downscaling study of GPM precipitation products was conducted based on the precipitation data of Guizhou Province by establishing multiple spatial downscaling models. Firstly, with the topographic factors including longitude, latitude, elevation, slope, and aspect as explanatory variables and the original GPM precipitation data as target variables, multiple downscaling models were established based on the methods of multivariate linear regression, geographically weighted regression, extreme learning machine, support vector machine, and random forest regression. Then multiyear average precipitation data were applied and assessed, and the optimal model was selected to conduct the spatial downscaling study of the annual and monthly precipitation amount in typical years in Guizhou Province. According to the results, the downscaling models except for the random forest regression model all performed well. Most especially, the multivariate linear regression model performed the most stably and effectively and yielded the highly improved downscaling results in terms of observation accuracy and spatial correlation. This study will provide a set of high-resolution gridded precipitation products for Guizhou Province and provide support for regional hydrometeorological research.

Keywords: Guizhou Province; GPM; spatial downscaling; multivariate linear regression; geographically weighted regression; extreme learning machine

(责任编辑:张仙)