

文章编号: 1008 - 2786 - (2019)6 - 923 - 09

DOI:10. 16089/j. cnki. 1008 - 2786. 000482

祁连山 TRMM 降水数据降尺度不同方法比较研究

沈思民, 赵军*, 刘佳茹, 赵彦军

(西北师范大学 地理与环境科学学院, 兰州 730070)

摘 要: 降水数据对研究山区陆面过程、水文模型、生态模型至关重要, 而山区地形复杂降水数据获取困难, 且 TRMM (Tropical Rainfall Measuring Mission) 降水数据空间分辨率难以进行小尺度区域降水研究。本文以祁连山为研究区, 应用 2005—2016 年的 TRMM 降水数据, 从不同时间尺度对 TRMM 降水数据在祁连山的适用性进行分析, 利用多元逐步回归模型、地理加权回归模型、随机森林模型三种降尺度方法, 分别建立了降水与高程、植被指数、纬度、经度、坡度、坡向的关系, 获得 1 km 高分辨率的 TRMM 降尺度数据, 并通过对比三种模型的降尺度结果来选取适用于祁连山的降尺度方法。结果表明: (1) TRMM 降水数据适用于祁连山。TRMM 降水数据和实测数据在不同时间尺度上具有显著的相关性, 年、季、月相关系数分别为 0.88、0.92、0.88。(2) 三种降尺度模型都能有效地获得 1 km 高分辨率的 TRMM 降尺度数据, 其中随机森林模型更适用于获取祁连山 TRMM 降尺度降水数据。(3) 相比于原始 TRMM 降水数据, 随机森林模型降尺度结果整体偏小, 同时具有更高的空间分辨率和更小的偏差。本研究可以为进一步获取西部山地高空间分辨降水数据和开展水文研究提供参考与借鉴。

关键词: TRMM 降水数据; 降尺度; 祁连山; 随机森林

中图分类号: P407, P413

文献标志码: A

降水作为水循环中不可或缺的环节, 是研究气候变化的重要因子, 是地球圈层物质能量交换的重要介质^[1]。山区降水是个复杂的过程, 与地形、高程、植被等环境因素有关, 获取高分辨率降水数据对研究山区陆面过程、水文模型、生态模型至关重要。目前, 山区降水数据的获取方式主要包括地面气象台站观测、地基测雨雷达实测和遥感反演三种途径。由于山地地形起伏较大, 气象站点分布稀疏, 利用实测降水数据插值很难获得高分辨率降水, 无法很好地反映山区降水分布特征^[2]。随着遥感技术的日益成熟和算法的不断完善, 多卫星联合反演日益成为高时空分辨率降水资料的主要获取途径^[3]。

热带测雨任务卫星 (tropical rainfall measuring mission, TRMM) 由美国国家航空航天局 (NASA) 和日本宇宙航空研究开发机构 (JAXA) 共同研制开发, 搭载了全球第一个星载测雨雷达 (precipitation radar, PR), 已获得大量高时空分辨率的降水数据^[4]。目前 TRMM 降水数据主要用于适应性评估^[3-5]、数据校正^[6-7]、干旱研究^[8-9]和降尺度研究四个方面。TRMM 数据空间统计降尺度研究已多有成果发表, 一些学者^[10-12]通过建立降水和不同的因子间的全局回归模型进行降尺度研究, 获得了高空间分辨率降水数据, 为降水降尺度研究提供了参考。虽然, 利用全局回归可以获取高空间分辨率的降水数

收稿日期 (Received date): 2019 - 05 - 07; 改回日期 (Accepted date): 2019 - 11 - 04

基金项目 (Foundation item): 国家自然科学基金项目 (41661084)。[National Nature Science Foundation of China (41661084)]

作者简介 (Biography): 沈思民 (1995 -), 男, 江苏徐州人, 硕士研究生, 主要研究方向: 资源环境遥感。[SHEN Simin (1995 -), male, born in Xuzhou, Jiangsu province, M. Sc candidate, research on remote sensing for natural resources and environment] E-mail: ssmwyy@qq.com

* 通讯作者 (Corresponding author): 赵军 (1963 -), 男, 山西河津人, 教授, 博士, 主要从事环境遥感与 GIS 应用方面的研究。[ZHAO Jun (1963 -), male, born in Hejin, Shanxi province, Ph. D., professor, specialized in environmental remote sensing and GIS application] E-mail: zhaojun@nwnu.edu.cn

1.2 研究数据

本文采用的 TRMM 3B43 降水数据,数据来源于 NASA 网站 (<https://pmm.nasa.gov/data-access>),时间为 2005—2016 年,数据时间分辨率为月尺度,空间分辨率为 $0.25^{\circ} \times 0.25^{\circ}$,覆盖南北纬 50° 范围^[22]。

站点实测数据源于中国气象数据共享网 (<http://data.cma.cn>),时间为 2005—2016 年。因祁连山境内气象站点稀疏,故对祁连山做 80 km 缓冲区,选取 33 个气象站点代表祁连山年均降水情况和验证原始 TRMM 数据在祁连山地区的适用性。选用祁连山境内 8 个气象站点地面实测降水数据验证降尺度结果。地面实测祁连山 2005—2016 年平均降水量为 255.3 mm/a;2005 年平均降水量为 259.4 mm/a,代表了降水量平均年份;2013 年平均降水量为 217.7 mm/a,代表了降水量干旱年份;2016 年平均降水量为 295.5 mm/a,代表了降水量湿润年份。选择三个典型年份进行降尺度结果对比,评价不同气候条件下降尺度结果。

DEM (Digital Elevation Model) 数据采用 STRM 数据,分辨率为 90 m,源于 GSCloud (<http://www.gscloud.cn/>)。利用邻近像元法重采样获得 1 km 和 0.25° 分辨率的坡向和坡度数据。植被指数

(Normalized Difference Vegetation Index, NDVI) 为 1 km 分辨率的月合成数据 MOD13A3,数据来源于 USGS 网站 (<https://earthexplorer.usgs.gov/>)。土地利用数据采用的 MCD12Q1 数据,利用邻近像元法将数据重采样为 1 km 分辨率,下载于 NASA 网站。本文选用 IGBP 的全球植被分类方案,因该方案包含 17 个主要的土地覆盖类型而被广范应用^[23]。土地利用数据用来剔除降水不是主导因素的 NDVI 值,再对空值区域进行克里金插值。

1.3 研究方法

降尺度模型的技术路线如图 2 所示。

(1) 多元逐步回归模型降尺度法

多元逐步回归模型 (Multiple Stepwise Regression) 通过拟合出最优方程,方程中每个因子都对因变量有统计学意义,从而建立最优或合适的回归模型^[24]。本文选取 NDVI、高程、经度、纬度、坡度和坡向六个因子建立多元逐步回归模型。

(2) 地理加权回归模型降尺度法

地理加权回归模型 (Geographical Weighted Regression) 是 Fotheringham 提出的一种局部加权回归模型,可以用来量化空间异质性^[25-26]。本文选取与降水相关性较好的 NDVI 作为因子建立地理加权

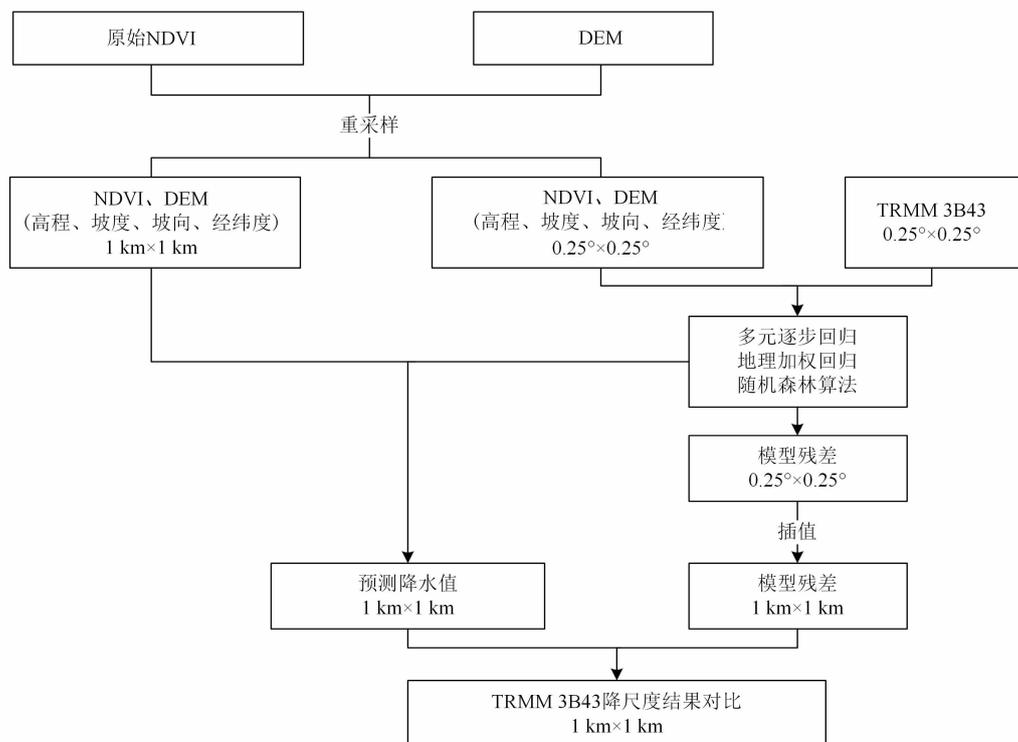


图 2 降尺度模型技术路线

Fig.2 The flow chart of downscaling models

回归模型。

(3) 随机森林模型降尺度法

随机森林模型(Random Forest)是Breiman和Cutler提出的一种基于决策树的机器学习模型^[27],因其具有较高的预测准确率,较优的非线性数据拟合度和较好的离群值容忍度,已被广泛应用于回归预测、遥感反演等领域^[28-29]。本文用袋外数据的模型误判率值(OOB error; out-of-bag error)来确定参数树的数量(ntree)和变量数量(mtry)的值,分别为500、6。

(4) 降尺度结果精度评价

选取相关系数 CC (Correlation Coefficient)、偏差 $BIAS$ 、均方根误差 $RMSE$ (Root Mean Square Error)、决定系数 R^2 、异常值数量这5个指标来评价降尺度结果的精度。

CC 用来反映降尺度结果和站点数据的相关性:

$$CC = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

$BIAS$ 用来表示降尺度结果和站点数据的偏差:

$$BIAS = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n y_i} - 1 \quad (2)$$

$RMSE$ 用来评价降尺度结果和站点数据的误差水平:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (3)$$

式(1)~式(3)中, x_i 为降尺度结果 i 点的值; y_i 为站点数据 i 点的值; \bar{x} 为降尺度数据的平均值; \bar{y} 为站点数据的平均值; n 表示站点数。 R^2 决定系数,也称拟合优度,反映了降尺度结果对实测降水数据

的解释程度。异常值数量,降尺度过程中产生的降水极小值。

2 结果分析

2.1 原始 TRMM 降水数据适应性评价

利用祁连山境内33个站点12年的站点实测数据,对2005—2016年TRMM 3B43降水数据进行多时间尺度的适用性分析(图3)。TRMM降水数据和实测数据在不同时间尺度上具有显著的相关性,年、季、月相关系数分别为0.88、0.92、0.88($P < 0.01$);相应的决定系数 R^2 分别为0.779、0.845、0.776。表明TRMM降水数据在年、季、月尺度上和站点实测数据之间具有显著的相关性和较好的拟合度。同时,相对较大的 $RMSE$ 也说明TRMM降水数据仍存在一定的偏差,相对于实测数据整体偏高。综上所述,TRMM降水数据适用于祁连山。

2.2 降尺度结果对比

利用ArcGIS、ENVI软件和R语言处理,得到降水水平均年(2005年)、干年(2013年)、湿年(2016年)三个典型年份的多元逐步回归模型、地理加权模型和随机森林模型降尺度结果(图4)。为了更好地对比降尺度结果,本文统计了三个典型年份的原始TRMM 3B43降水数据离散程度(图5)。五个评价指标见表1。

由表1可知,三种模型的降尺度结果和实测降水数据的相关性相对较高($CC_{\min} = 0.72$)。干年和湿年的三种降尺度结果均较原始TRMM降水数据拟合度更好, R^2 值有明显提高,其中RF模型 R^2 提高最显著,为0.13;同时, $RMSE$ 均有所减少,其中

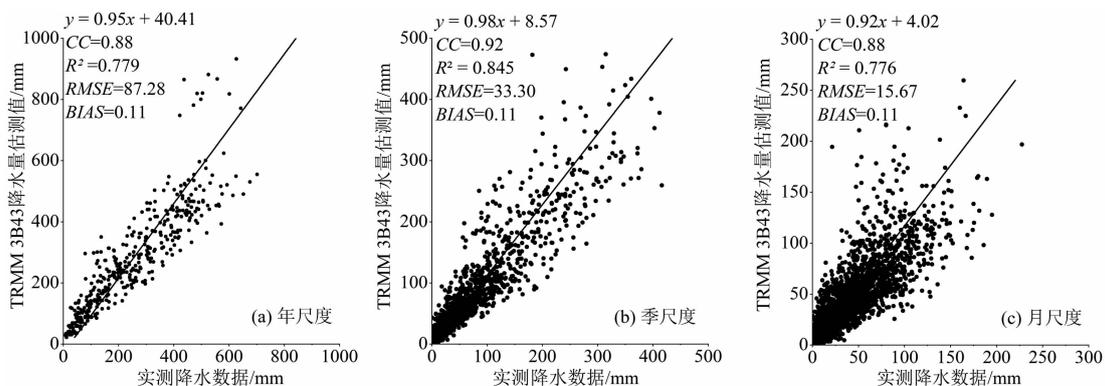


图3 站点降水与 TRMM3B43 数据的相关性

Fig. 3 Correlation between TRMM 3B43 and measured precipitation data

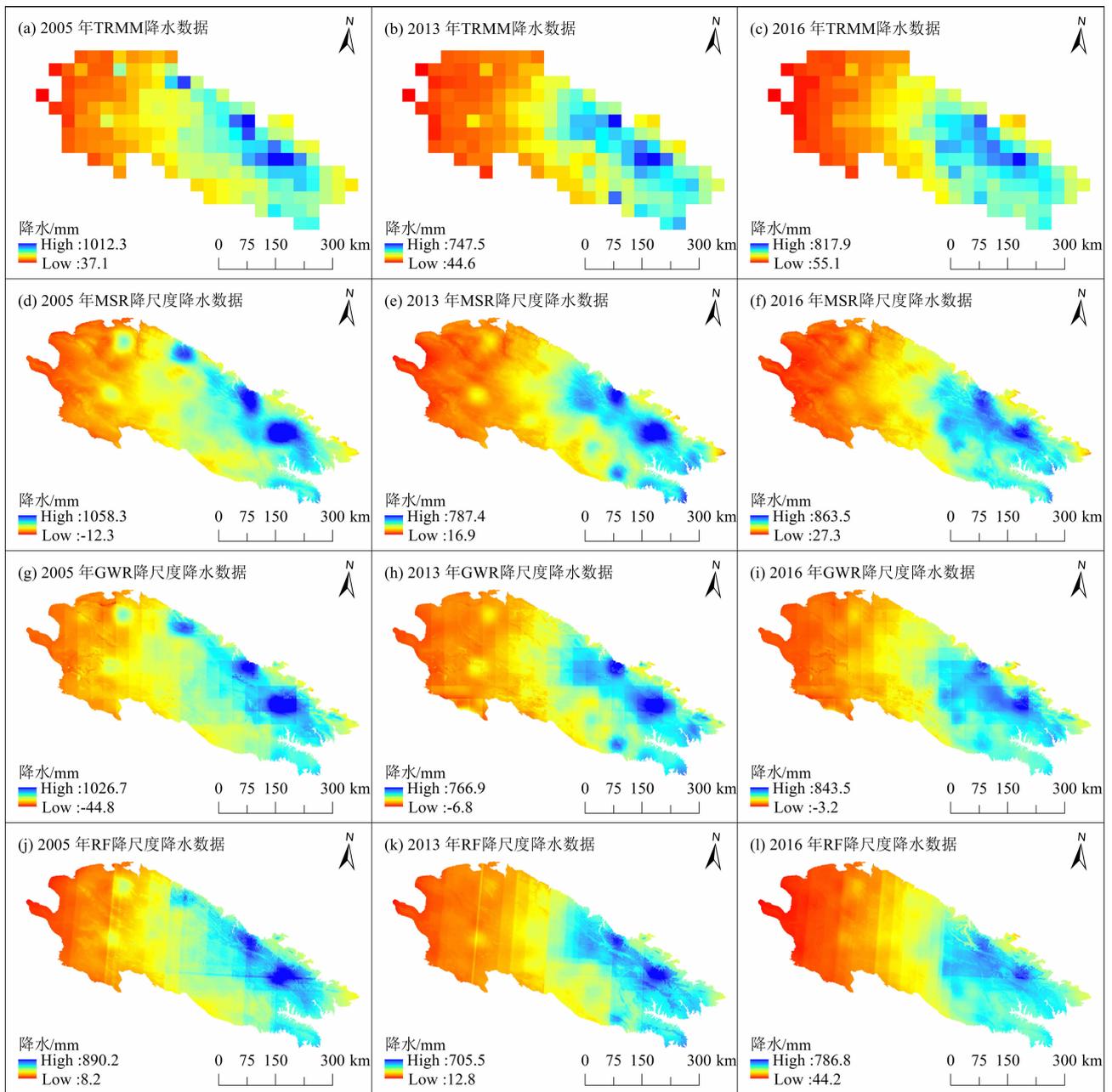


图 4 三个典型年份 TRMM 3B43 数据和降尺度结果

Fig. 4 Downscaled results and TRMM 3B43 data for three typical years

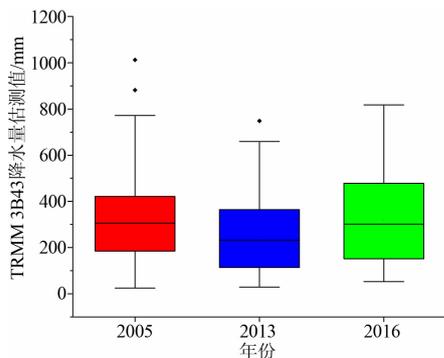


图 5 三个典型年份原始 TRMM 3B43 数据箱线图

Fig. 5 Boxplots of original TRMM 3B43 data for three typical years

RF 模型减少最显著,为 48.27。GWR 模型的 *BIAS* 较原始数据均增加,且 GWR 会产生少数异常值,虽然 GWR 模型可以在局部建立鲁棒性更好的函数,但是祁连山降水受地形差异的影响显著,仅用 NDVI 作为因子,不能很好地拟合降水变化特征。三个典型年份中,湿年和干年的三种降尺度结果整体好于平均年,平均年三种降尺度结果中仅有 RF 模型降尺度结果的 R^2 值有明显提高,*BIAS* 和 *RMSE* 值均减少。

表 1 原始 TRMM 数据和降尺度结果与站点降水数据比较

Tab.1 Comparison of original TRMM precipitation data and downscaled data with measured precipitation data

年份	数据	CC	BIAS	RMSE	R ²	异常值数量
2005 (平均年)	原始 TRMM 数据	0.83	0.10	134.13	0.70	—
	MSR 降尺度数据	0.80	0.09	156.80	0.63	59
	GWR 降尺度数据	0.75	0.19	165.31	0.57	55
	RF 降尺度数据	0.91	0.05	71.16	0.83	—
2013 (干年)	原始 TRMM 数据	0.71	0.05	131.68	0.51	—
	MSR 降尺度数据	0.73	-0.01	115.85	0.54	—
	GWR 降尺度数据	0.72	0.06	119.86	0.52	6
2016 (湿年)	原始 TRMM 数据	0.93	0.07	86.25	0.86	—
	MSR 降尺度数据	0.95	-0.01	59.24	0.90	—
	GWR 降尺度数据	0.93	0.08	76.28	0.87	2
	RF 降尺度数据	0.99	-0.05	37.98	0.97	—

由图 5 可以看出,湿年没有离群值,平均年和干年的原始 TRMM 降水数据中有少数离群值,其中平均年的离群值和上限相差相对较大。可知,GWR 和 MSR 模型不能很好地处理模型中的离群值,当离群值和上限相差较大时,会影响降尺度结果。这就可以很好地解释湿年和干年的三种降尺度结果整体好于平均年。而 RF 模型在生成决策树过程中对离群值不敏感,因而离群值对降尺度结果影响相对较小。

综上所述,三种降尺度模型都获得了较高空间分辨率的 TRMM 降水数据,和站点实测降水值呈现较好的相关性,表明三种方法在祁连山的降尺度效果基本理想,具有一定的可行性。三个模型中,MSR 模型降尺度结果整体好于 GWR 模型,细节表现方面稍好于其他模型,在原始 TRMM 降水数据没有离群值或者离群值和上限相差较少的年份可以选择 MSR 模型。RF 模型对比 MSR、GWR 模型有更好的拟合度,更高的预测准确率,对离群值有更好的容忍度。因此,RF 模型在祁连山 TRMM 降水数据降尺度中有更好的适用性。

2.3 异常值处理

降尺度结果中产生极小异常值的主要原因有:(1)回归模型的预测值和真值之间残差的差值结果会忽略降水的空间异质性。(2)降水受地形因子的影响显著,仅仅用 NDVI 作为因子在降尺度过程中容易产生异常值。(3)GWR 模型和 MSR 模型不能

很好地处理原始 TRMM 数据存在的离群值,会影响回归的预测结果。

对于降尺度结果中降水小于 1 mm 的值作为异常值处理,用 5×5 窗口的临近像元的平均值替换空值。虽然存在一定的误差,但可以在一定程度上消除降尺度结果中的异常值对后续研究的影响。图 6 是经异常值处理后的结果。

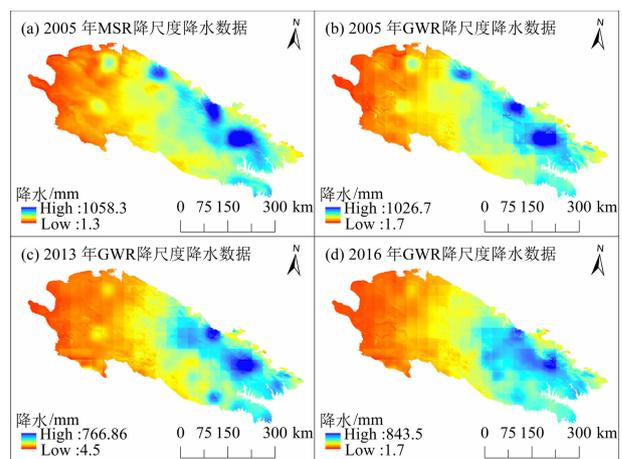


图 6 三个典型年份异常值处理后的降尺度结果

Fig. 6 Downscaled results after handled outliers for the three typical years

3 结论

(1) TRMM 降水数据适用于祁连山地区,但相

对于实测降水数据整体偏高。TRMM 降水数据和实测降水数据,在祁连山地区多时间尺度上均呈现出较高的相关性($CC_{\text{year}} = 0.88$ 、 $CC_{\text{season}} = 0.92$ 、 $CC_{\text{month}} = 0.88$)和较好的拟合度($R_{\text{year}}^2 = 0.779$ 、 $R_{\text{season}}^2 = 0.845$ 、 $R_{\text{month}}^2 = 0.776$)。

(2)三种模型的降尺度结果与实测降水数据之间均呈现显著相关性,可用于获取祁连山 TRMM 降尺度降水数据。相比于另两种模型的降尺度结果,RF 模型降尺度结果具有更显著的相关性、更好的拟合度和更小的偏差,更适用于祁连山 TRMM 降水数据降尺度;在原始 TRMM 降水数据没有离群值或者离群值和上限相差较小的年份也可以选择 MSR 模型降尺度。

(3)RF 模型降尺度结果相对于原始 TRMM 降水数据整体偏小,与站点实测数据相关性和拟合度均较原始 TRMM 降水数据有所提升,偏差则有所减少。表明 RF 模型降尺度结果不仅能获取高空间分辨率的 TRMM 降水数据,而且具有更小的偏差。

参考文献 (References)

- [1] ALLEN M R, INGRAM W J. Constraints on future changes in climate and the hydrologic cycle [J]. *Nature*, 2002, **419**(6903): 224 - 232.
- [2] 叶柏生,杨大庆,丁永建,等. 中国降水观测误差分析及其修正[J]. *地理学报*, 2007, **62**(1): 3 - 13. [YE Baisheng, YANG Daqing, DING Yongjian, et al. A bias-corrected precipitation climatology for China [J]. *Acta Geographica Sinica*, 2007, **62**(1): 3 - 13]
- [3] HUFFMAN G J, BOLVIN D T, NELKIN E J, et al. The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales [J]. *Journal of Hydrometeorology*, 2007, **8**(1): 38 - 55.
- [4] 赵军,刘原峰,朱国锋,等. 热带测雨卫星数据在黑河流域的精度及应用[J]. *水土保持通报*, 2016, **36**(3): 309 - 315. [ZHAO Jun, LIU Yuanfeng, ZHU Guofeng, et al. Accuracy and application of tropical rainfall measuring mission data in Heihe Rvier Basin [J]. *Bulletin of Soil and Water Conservation*, 2016, **36**(3): 309 - 315]
- [5] CAO Yueqian, ZHANG Wu, WANG Wenjing. Evaluation of TRMM 3B43 data over the Yangtze River Delta of China [J]. *Scientific Reports*, 2018, **8**(1): 1 - 12.
- [6] CONDOM T, RAU P, ESPINOZA J C. Correction of TRMM 3B43 monthly precipitation data over the mountainous areas of Peru during the period 1998 - 2007 [J]. *Hydrological Processes*, 2011, **25**(12): 1924 - 1933.
- [7] 孙美平,张海瑜,巩宁刚,等. 基于 TRMM 降水订正数据的祁连山地区最大降水高度带研究[J]. *自然资源学报*, 2019, **34**(3): 646 - 657. [SUN Meiping, ZHANG Haiyu, GONG Ninggang, et al. Study on maximum precipitation height zone in Qilian Mountains area based on TRMM precipitation data [J]. *Journal of Natural Resources*, 2019, **34**(3): 646 - 657]
- [8] 胡鹏飞,李净,王丹,等. 基于 MODIS 和 TRMM 数据的黄土高原农业干旱监测[J]. *干旱区地理*, 2019, **42**(1): 172 - 179. [HU Pengfei, LI Jing, WANG Dan, et al. Monitoring agricultural drought in the Loess Plateau using MODIS and TRMM data [J]. *Arid Land Geography*, 2019, **42**(1): 172 - 179]
- [9] ZHAO Qi, CHEN Qianyun, JIAO Mengyuan, et al. The temporal-spatial characteristics of drought in the Loess Plateau using the remote-sensed TRMM precipitation data from 1998 to 2014 [J]. *Remote Sensing*, 2018, **10**(6): 1 - 20.
- [10] 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.
- [11] 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 Jing, ZHANG Xiao. Downscaling method of TRMM satellite precipitation data [J]. *Scientia Geographica Sinica*, 2015, **35**(9): 1164 - 1169]
- [12] WANG Lei, CHEN Rensheng, HAN Chuntan, et al. An improved spatial-temporal downscaling method for TRMM precipitation datasets in alpine regions: a case study in northwestern China's Qilian Mountains [J]. *Remote Sensing*, 2019, **11**(7): 1 - 23.
- [13] 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 Xiaochan, ZHANG Hongyan, ZHAO Jianjun, 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]
- [14] 曾昭昭,王晓峰,任亮. 基于 GWR 模型的陕西秦巴山区 TRMM 降水数据降尺度研究[J]. *干旱区地理*, 2017, **40**(1): 26 - 36. [ZENG Zhaozhao, WANG Xiaofeng, REN Liang. Spatial downscaling of TRMM rainfall data based on GWR model for Qinling-Daba Mountains in Shaanxi Province [J]. *Arid Land Geography*, 2017, **40**(1): 26 - 36]
- [15] 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 Yelong, TAN Wei, WANG Chao, 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]
- [16] CHEN Cheng, ZHAO Shuhe, ZHENG Duan, et al. An improved spatial downscaling procedure for TRMM 3B43 precipitation product using geographically weighted regression [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote*

- Sensing, 2015, **8**(9, SI):4592–4604.
- [17] ZHANG Yueyuan, LI Yungang, JI Xuan, 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.
- [18] 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 Xuewei, LIU Hailong. Downscaling method of TRMM satellite precipitation data over the Tianshan Mountains [J]. *Journal of Natural Resources*, 2018, **33**(3):478–488]
- [19] 徐彬仁, 魏媛媛. 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. [XU Binren, WEI Yuanyuan. 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]
- [20] 赵军, 黄永生, 师银芳, 等. 2000—2012 年祁连山中段雪线与气候变化关系 [J]. *山地学报*, 2015, **33**(6):683–689. [ZHAO Jun, HUANG Yongsheng, SHI Yinfang, et al. Relationship between snow line change and climate change in the middle of Qilian Mountains during 2000–2012 [J]. *Mountain Research*, 2015, **33**(6):683–689]
- [21] 雪宝顶. 祁连山: 伸向荒漠的湿岛 [J/OL]. *中国国家地理*, 2006, (2) [2019–05–08]. <http://www.dili360.com/cng/article/p5350c3d6d48c591.htm>. [XUE Baoding. Qilian Mountain: Wet island stretching into the desert [J/OL]. *Chinese National Geography*, 2006, (2) [2019–05–08]. <http://www.dili360.com/cng/article/p5350c3d6d48c591.htm>.]
- [22] HUFFMAN G J, BOLVIN D T. TRMM and other data precipitation data set documentation [EB/OL]. (2018–04–26) [2019–05–08]. https://pmm.nasa.gov/sites/default/files/document_files/3B42_3B43_doc_V7_180426.pdf.
- [23] SULLA-MENASHE D, FRIEDL M A. User guide to collection 6 MODIS Land Cover (MCD12Q1 and MCD12C1) product [EB/OL]. (2018–05–18) [2019–05–08]. https://lpdaac.usgs.gov/documents/101/MCD12_User_Guide_V6.pdf.
- [24] 范科科, 段利民, 张强, 等. 基于多种高分辨率卫星数据的 TRMM 降水数据降尺度研究——以内蒙古地区为例 [J]. *地理科学*, 2017, **37**(9):1411–1421. [FAN Keke, DUAN Limin, ZHANG Qiang, et al. Downscaling analysis of TRMM precipitation based on multiple high-resolution satellite data in the Inner Mongolia, China [J]. *Scientia Geographica Sinica*, 2017, **37**(9):1411–1421]
- [25] FOTHERINGHAM A S, BRUNSDON C, CHARLTON M. Geographically weighted regression [M]. New York: Wiley, 2002.
- [26] BRUNSDON C, FOTHERINGHAM A S, CHARLTON M. Geographically weighted regression [J]. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 1998, **47**(3):431–443.
- [27] BREIMAN L. Random forests [J]. *Machine Learning*, 2001, **45**(1):5–32.
- [28] 李欣海. 随机森林模型在分类与回归分析中的应用 [J]. *应用昆虫学报*, 2013, **50**(4):1190–1197. [LI Xinhai. Using “random forest” for classification and regression [J]. *Chinese Journal of Applied Entomology*, 2013, **50**(4):1190–1197]
- [29] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述 [J]. *统计与信息论坛*, 2011, **26**(3):32–38. [FANG Kuangnan, WU Jianbin, ZHU Jianping, et al. A review of technologies on Random Forests [J]. *Statistics & Information Forum*, 2011, **26**(3):32–38]

Comparative Study on Different Downscaling Methods of TRMM Satellite Precipitation Data over the Qilian Mountains, China

SHEN Simin, ZHAO Jun^{*}, LIU Jiaru, ZHAO Yanjun

(College of Geography and Environmental Science, Northwest Normal University, Lanzhou 730070, China)

Abstract: The high-spatial-resolution precipitation data plays a crucial role in the studies of hydrological models, ecological models and land surface processes. However, it is difficult to obtain high-spatial-resolution precipitation data in mountainous areas due to some factors, such as varied topography, sparse rainfall gauges and complicated precipitation environment. Meanwhile, the coarse spatial resolution of the Tropical Rainfall Measuring Mission (TRMM) precipitation data is difficult to apply in small-scale regions. In this paper, TRMM precipitation data at different time scales were used to analyze its applicability in the Qilian Mountains. Based on the TRMM

precipitation data from 2005 to 2016, Multiple Stepwise Regression Model, Geographical Weighted Regression Model and Random Forest Model were used to obtain downscaled precipitation data precipitation data by establishing relationships between precipitation and environmental variables (elevation, vegetation index, latitude, longitude, slope and aspect). TRMM precipitation data with 1 km high spatial resolution in the Qilian Mountains was obtained by comparing the three downscaling models and further to select the most suitable one. Results showed as follows: (1) TRMM precipitation data was well matched with rain gauge data in the Qilian Mountains. TRMM precipitation data and measured data had significant correlation on different time scales. The annual, seasonal and monthly correlation coefficients were 0.88, 0.92 and 0.88, respectively. (2) Three downscaling models can effectively obtain downscaled precipitation data with 1 km high spatial resolution in yearly time scale, the Random Forest Model was the most suitable model for downscaling of TRMM precipitation data in Qilian Mountains. (3) Compared with the original TRMM precipitation data, the downscaled results of the Random Forest Model were relatively small, with higher spatial resolution and smaller deviation. Thus, this study has certain reference significance for obtaining high spatial resolution precipitation data and cucting hydrological research in the arid mountainous areas of Northwest China.

Key words: TRMM precipitation data; downscaling; Qilian Mountains; random forest