

高分辨率青藏高原历史月降水数据重建

徐明, 石玉立, 王彬

(南京信息工程大学地理与遥感学院, 南京 210044)

摘要: 青藏高原对全球气候研究具有重要意义,而降水数据对水文、气象和生态等领域的研究也至关重要,且随着研究内容和尺度的变化,对高时空分辨率的历史降水数据的需求越发迫切。本文基于TRMM 3B43降水数据,采用随机森林算法,引入归一化植被指数(AVHRR NDVI)、高程(SRTM DEM)、坡度、坡向、经度、纬度6个地理因子,建立历史降水重建模型,获得1982–1997年分辨率为0.0833°的青藏高原年降水数据,然后根据比例系数法计算出月降水数据。为提高精度,利用站点数据对月降水数据进行校正。结果表明,该方法能简单有效地获得高时空分辨率的历史降水数据,决定系数 R^2 大部分在0.4~0.9之间,平均值为0.6767,其中夏季效果最好,冬季效果最差;均方根误差RMSE和平均绝对误差MAE均在50 mm以下, RMSE均值为22.66 mm, MAE均值为15.97 mm;偏差Bias较小,基本在0.0~0.1之间。

关键词: 随机森林;历史降水;重建;TRMM 3B43;青藏高原

1 引言

降水是重要的气象要素之一,对于全球气候变化、水循环、生态环境和农业生产都具有重要意义。观测降水的传统方式是在站点内使用雨量计进行观测,但由于青藏高原地形复杂,拥有众多高海拔的山脉,站点观测数据极为有限,并且分布不均匀(Chen et al, 2015; 徐祥德等, 2015)。随着遥感科技的发展,出现了一系列卫星观测的降水产品,如长时间序列降水数据GPCP(Global Precipitation Climatology Project)、CMAP(Climate Prediction Center Merged Analysis of Precipitation),时间分辨率高、时间序列短的降水数据TMPA(TRMM Multi-satellite Precipitation Analysis)、CMORPH(CPC MORPHing technique)、PERSIANN(Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks),高时空分辨率数据GSMaP(Global Satellite Mapping of Precipitation)、UBham

(University of Birmingham)等。然而,长时间序列的降水数据分辨率较低,高时空分辨率的降水数据时间跨度较短,进行区域性气候、水文过程等的长时间序列趋势性分析,需要高分辨率的历史降水产品。当前降水产品空间分辨率和时间上的限制,使其不能满足相应的研究需求,所以需要进行历史降水数据的重建。与其他遥感降水产品相比,TRMM (Tropical Rainfall Measuring Mission)降水数据的准确性和空间分辨率较高(黄聪敏等, 2014)。因此,本文采用TRMM降水数据对青藏高原地区历史降水数据进行重建。

对于青藏高原地区历史降水数据重建国内外已经有了一些研究,通常选择与降水相关的要素作为因子构建模型。何红艳等(2005)考虑到青藏高原独特的地理特征,利用站点和海拔数据混合差值得到青藏高原降水图;Jia等(2011)利用NDVI和海拔建立多元线性回归模型;候光良等(2012)考虑地形因素对降水的影响,选取经度、纬度、海拔、坡度和

收稿日期:2017-09-06;修订日期:2018-02-10。

基金项目:国家自然科学基金项目(41471312) [Foundation: National Natural Science Foundation of China, No. 41471312]。

作者简介:徐明(1993-),女,黑龙江齐齐哈尔人,硕士研究生,主要从事青藏高原降水研究, E-mail: xumingnuist@163.com。

通讯作者:石玉立(1973-),男,河南新乡人,博士,教授,主要从事降水、机器学习等研究, E-mail: yishi.nuist@gmail.com。

引用格式:徐明, 石玉立, 王彬. 2018. 高分辨率青藏高原历史月降水数据重建[J]. 地理科学进展, 37(7): 923-932. [Xu M, Shi Y L, Wang B. 2018. Reconstruction of high resolution monthly precipitation data of the Tibetan Plateau[J]. Progress in Geography, 37(7): 923-932.]. DOI: 10.18306/dlkxjz.2018.07.006

坡向建立分区模拟方程。选择青藏高原月降水模拟的因子时,由于海拔与降水相关性较大,高程越高降水越多,因此将海拔作为构建模型的因子之一(Jia et al, 2011)。地理位置对于降水也有影响,可将经纬度作为模型因子。青藏高原地区山脉众多,在迎风坡,容易形成地形雨,坡度和坡向因子可体现该地区山脉的分布(傅抱璞, 1992)。基于以上研究,本文选取坡度、坡向、高程、NDVI和经纬度作为模型因子,进行历史降水数据重建。

2 研究区域与数据

2.1 研究区域

青藏高原位于中国西南部,总面积约 250 万 km²,经度范围为 73°18'52"E~104°46'59"E,纬度范围为 26°00'12"N~39°46'50"N(图 1)。该地区拥有多种植被覆盖类型,南部和东部边缘的混交林和稀疏灌丛,高原腹地大部分为草地,在高原北部则被稀疏植被所覆盖。青藏高原海拔在 1000~7000 m 之间,平均海拔 4000 m 以上,高原西南部地区海拔较高,基本在 5000 m 以上,自西南部向东北部海拔逐渐降低,南部边缘及东部边缘地区海拔低于其他地区。青藏高原地区地貌复杂,有众多山脉,如西部的昆仑山脉,南部的喜马拉雅山脉,东北部的祁连山脉,以及位于青藏高原东部的横断山脉,同时还具有宽谷、盆地和高原面等地貌类型,内外均有水系分布。由于青藏高原的海拔高,且拥有众多山脉,对于内陆气候起到了屏障作用,因此对中国乃至全球的气候具有重要的影响(吴国雄, 2004; Shi

et al, 2015)。

青藏高原的水汽主要来源于印度洋的孟加拉湾和阿拉伯海。孟加拉湾的暖湿气流沿布拉马普特拉河、雅鲁藏布江、横断山三江流域向高原输送水汽。青藏高原东南地区地势低平,水汽在进入雅鲁藏布江大拐弯后以“湿舌”状深入高原腹地。阿拉伯海的水汽输送在不同季节有差异,夏季自玛法木错湖以西进入,冬季从帕米尔高原缓慢东移,在青藏高原南部带来边缘降水。阿拉伯海的水汽大部分被喜马拉雅山、喀喇昆仑山、昆仑山所截留,因此并未深入高原腹地(林振耀等, 1990)。

2.2 数据介绍

(1) SRTM(Shuttle Radar Topography Mission)数据。SRTM 由美国太空总署(NASA)和国防部国家测绘局(NIMA)联合测量,数据范围在 60°N~60°S 之间。SRTM 在美国地区为 30 m 分辨率,在其他地区数据分辨率为 90 m。该数据下载于 <http://www.gscloud.cn/>。本文从 DEM 数据提取坡度、坡向及经纬度信息,并且将其分别重采样到 0.25°和 0.0833°分辨率,用于随机森林模型构建。

(2) AVHRR NDVI 3g 数据。NDVI(Normalized Difference Vegetation Index)具有时间序列长、覆盖范围广、时空可比、较强地植被动态变化表征能力(Beck et al, 2011)等特点,被证明是描述植被生长动态变化的数据集之一(杜加强等, 2016)。NDVI 3g 数据的空间分辨率为 0.0833°,时间分辨率为 15 天。本文使用 1982-2012 年的 NDVI 数据,获取青藏高原的植被指数用于随机森林模型构建。采用最大合成算法(MVC)进一步去除云的影响,计算每个像元的月最大值,获得每月最大 NDVI 值。年 NDVI 值是 12 个月 NDVI 的平均值,如式(1)所示:

$$NDVI_{\text{annual},i} = \frac{\sum_{j=1}^{12} NDVI_{\text{monthly},j}}{12} \quad (1)$$

式中: $NDVI_{\text{annual},i}$ 为第 i 年的 NDVI 值; $NDVI_{\text{monthly},j}$ 为第 i 年第 j 月的 NDVI 值。

(3) TRMM(The Tropical Rainfall Measuring Mission)3B43 数据。热带降雨测量计划由美国国家航空航天局(NASA)与日本宇宙航空研究开发机构(JAXA)联合开展,用于研究降雨对天气和气候的影响(Prakash et al, 2015)。TRMM 3B43 数据覆盖范围为纬度 50°N~50°S,经度 180°W~180°E,分辨率为 0.25°,数据收集时间为 1997 年 11 月 27 日~2015 年 4 月 15 日,本文使用 1998~2012 年 TRMM 数据获取青藏高原地区降水。该数据下载于 <http://trmm>。

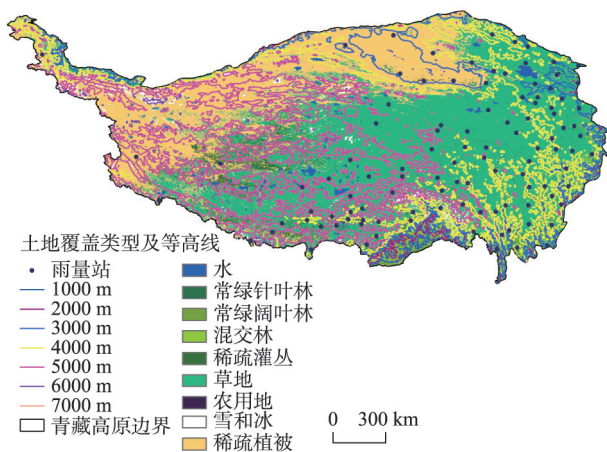


图 1 青藏高原高程及雨量站分布图

Fig.1 Elevation of the Tibetan Plateau and distribution of rainfall stations

gsfc.nasa.gov/data_dir/ProductStatus.html。

(4) 站点实测数据。从中国气象数据共享服务系统(<http://data.cma.gov.cn/>)获得 1982–2012 年站点实测数据,包括中国地面气候资料月值数据集以及中国地面气候资料年值数据集。青藏高原地区站点个数约为 97 个,由于地理位置和环境因素,在不同年份的站点数量稍有变化。该数据用于验证历史降水重建的结果,以及模型模拟降水的校正。

3 研究方法

降水是一个复杂的气象要素,与各个因素相互作用。在降水模拟过程中,要选择与降水相关性较高的因子作为输入变量,并且所选因子应具有相应的高时空分辨率遥感观测数据(Immerzeel et al, 2009; 刘永和等, 2011)。

青藏高原属于干旱半干旱地区,降水对于植被生长有重要影响,两者有显著的正相关关系。因此,将 NDVI 作为模拟反演降水的一个重要因子(Duan et al, 2013; Pang et al, 2017)。同时,青藏高原海拔较高,拥有众多山脉,因此地形与降水关系密切(Jia et al, 2011; Xu et al, 2015),傅抱璞等(1992)曾通过半经验理论建立的降水随地形和海拔高度变化的数学模型得出,坡度对降水的影响存在显著的相关性。反观青藏高原的地形与降水特征(图 1),由于青藏高原受东南方向季风影响显著,降水存在着西北—东南阶梯特征,因此降水与地理位置(如经度和纬度)关系同样密切。综上所述,本文采用了 NDVI、海拔、坡度、坡向、经度、纬度 6 个地理因子建立模型。

3.1 模型构建

本文采用降水作为因变量,6 个地理因子作为自变量,构建随机森林模型,具体步骤如下(图 2):

(1) 数据预处理:SRTM 数据获取高程、坡度、坡向、经度、纬度;NDVI 3g 数据根据最大合成算法(MVC)计算 NDVI 月值,再根据某一年 12 个月的 NDVI 月值取平均得到 NDVI 年值。

(2) 将经过数据预处理的高程、坡度、坡向、经度、纬度、归一化植被指数数据,利用最邻近法重采样到 0.25° ,使其与原始的 TRMM 3B43 降水数据具有相同分辨率。

(3) 选取 1999–2012 年的数据,将高程、坡度、坡向、经度、纬度、归一化植被指数作为自变量,TRMM 3B43 降水数据作为因变量,构建随机森林模型。

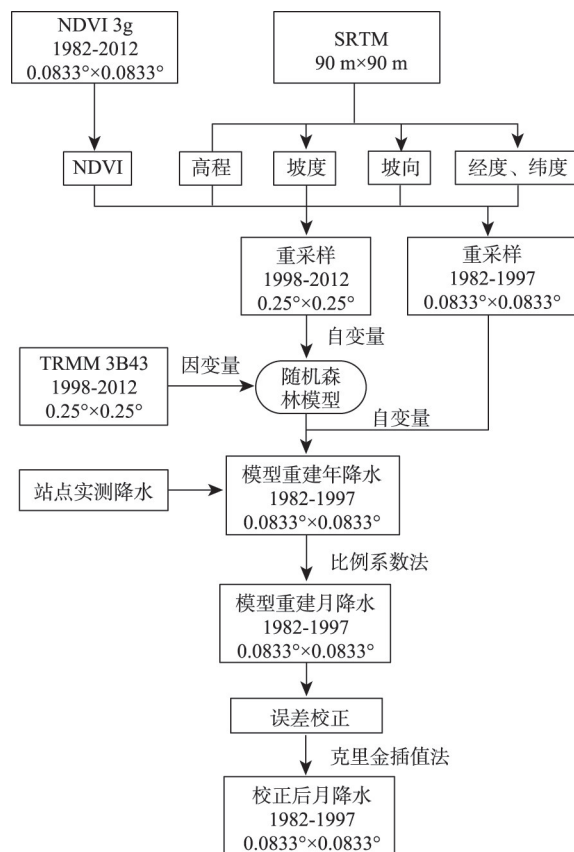


图 2 1982-1997 重建月降水流程图

Fig.2 Flow chart for the reconstruction of monthly precipitation, 1982-1997

(4) 将高程、坡度、坡向、经度、纬度、归一化植被指数数据利用最邻近法重采样到 0.0833° 的分辨率,输入已构建的随机森林模型中,模拟得到 0.0833° 分辨率 1982–1997 历史年降水数据。

(5) 利用比例系数法(Duan et al, 2013),根据 TRMM 3B43 数据确定月降水在该年降水中所占比例。利用年降水值乘以月降水所占比例,即可得到月降水值。1982–1997 年间月降水所占年降水的比例是根据 1998–2012 年的各个月降水取平均值得到的。

(6) 模拟降水校准:计算月站点数据与模拟月降水差值,利用克里金插值法计算青藏高原地区差值图。用模拟月降水减去差值图,对模拟降水进行校正。

3.2 随机森林算法

随机森林是一个具有分类响应变量的分类器,根据预测变量中包含的信息对观测值进行分类。随机森林是通过一系列决策树构建的,通常将随机森林看作决策树分类器的集合。

$$\{f(x, \theta_k), k = 1, \dots\} \quad (2)$$

式中: x 为输入变量; k 是对所给出的树的索引; θ_k 是第 k 棵树的随机变量, θ_k 与之前的随机变量 $\theta_1, \dots, \theta_{k-1}$ 没有相关关系。袋装方法是从训练数据中可重复的随机选取观测值,而随机森林是从每一个潜在分支中不可重复抽取预测因子的子集。在这2种情况中, θ_k 是因子的集合。这2种采样程序的因子都可以用 θ_k 来表示(Maronna, 2011)。

每一个观测值输入分类器后都会给定一个预测值,这个值取决于变量 x 。在分类回归树算法中,观测值被分为哪一类是与观测值的终点相关的。在随机森林中,观测值的分类是通过一系列决策树决定的,于随机森林是通过第 k 个分类器确定观测值类别,最终通过所有决策树来决定观测值类别,有效提高了模拟进度(明均仁等, 2012; 李欣海, 2013)。

随机森林模型是一种机器学习模型,可用于回归。随机森林算法是先将多维数据输入随机森林模型,模型有放入的随机采样选出 n 个样本,随机选择 m 个特征,利用这些特征对选出的样本建立决策树,重复以上步骤多次,生成一系列决策树,从而形成随机森林。本文是将降水与高程、坡度、坡向、经度、纬度、归一化植被指数数据输入随机森林模型,构建降水与其他相关因子之间的关系。随机森林模型对数据集的适应性强,可避免出现过拟合的现象,同时也具有较好的抗噪声能力。

3.3 精度评价

(1) 决定系数 R^2 : 对所建立的模型而言,表示站点实测数据与模型重建数据的相关密切程度,范围是 $[0, 1]$ 。当 R^2 越接近1时,模型模拟降水效果越好,反之,则效果越差。

$$R^2 = \left(\frac{\sum_{i=1}^n (P_{\text{station},i} - \bar{P}_{\text{station}}) \sum_{i=1}^n (P_{\text{model},i} - \bar{P}_{\text{model}})}{\sqrt{\sum_{i=1}^n (P_{\text{station},i} - \bar{P}_{\text{station}})^2} \sqrt{\sum_{i=1}^n (P_{\text{model},i} - \bar{P}_{\text{model}})^2}} \right)^2 \quad (3)$$

(2) 均方根误差 $RMSE$: 站点实测降水值与模型模拟降水值偏差的平方与观测站点个数比值的平方根,表明观测值与其模拟值之间的偏差。

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_{\text{station},i} - P_{\text{model},i})^2}{n}} \quad (4)$$

(3) 平均绝对误差 MAE : 所有站点实测值与模型模拟降水值的偏差的绝对值的平均,反应模型模拟误差情况。

$$MAE = \frac{\sum_{i=1}^n |P_{\text{station},i} - P_{\text{model},i}|}{n} \quad (5)$$

(4) 偏差 $Bias$: 随机森林模型模拟历史降水与站点实测值之间的差距。

$$Bias = \frac{\sum_{i=1}^n P_{\text{station},i}}{\sum_{i=1}^n P_{\text{model},i}} - 1 \quad (6)$$

式中: $P_{\text{station},i}$ 为第 i 个站点实测降水量; \bar{P}_{station} 为站点实测降水平均值; $P_{\text{model},i}$ 为模型模拟第 i 个站点的降水量; \bar{P}_{model} 为模型模拟降水平均值; n 为验证模型的地面观测站点的个数。

4 结果分析

4.1 重建年降水结果

利用1998–2012年6个因子构建随机森林模型,从而得到1982–1997年的年降水数据,如图3所示。从图3可以看出,不同年份之间的年降水差异不大,降水的年际变化较小,年总降水量均在6–3192 mm之间。青藏高原地区年降水西部地区降水量较少,自西北向东南逐渐增加。东南部地区降水较多,均在600 mm以上,在东南部边缘由于受到孟加拉湾水汽的影响年降水量高达3000 mm以上。孟加拉湾的水汽沿着布拉马普特拉河进入雅鲁藏布江大拐弯从而影响东南部降水。青藏高原西南部边缘降水,受阿拉伯海水汽影响。因水汽被喜马拉雅山脉、喀喇昆仑山脉阻隔,形成地形雨,年降水量在700 mm左右。西南地区由于山体的阻隔,地形不象东南部平坦,因而降水较东南部地区少,在400 mm以上。在青藏高原东北部地区有一小部分地区降水值高于周围地区,是该地区为青海湖,因此降水偏高。

4.2 重建月降水结果

用同样方法反演得到1982–1997年青藏高原月降水数据。由于各个年份在同一月份的模型模拟精度有所波动,为更精确地显示体现青藏高原各个月份的降水分布情况,本文选取1982–1997年各月降水反演精度最高时所对应的年份,得到青藏高原月降水图,如图4所示,个别年份的月降水精度均低于其他年份,因此在图4中没有显示。根据重建获取的0.1°分辨率的数据,可以获取青藏高原降水在各个地区的降水分布规律。。总体而言,青藏高原东南部地区降水量较多,向西北部逐渐减少,呈现阶梯状分布。在东南部边缘地区,由于受到孟加

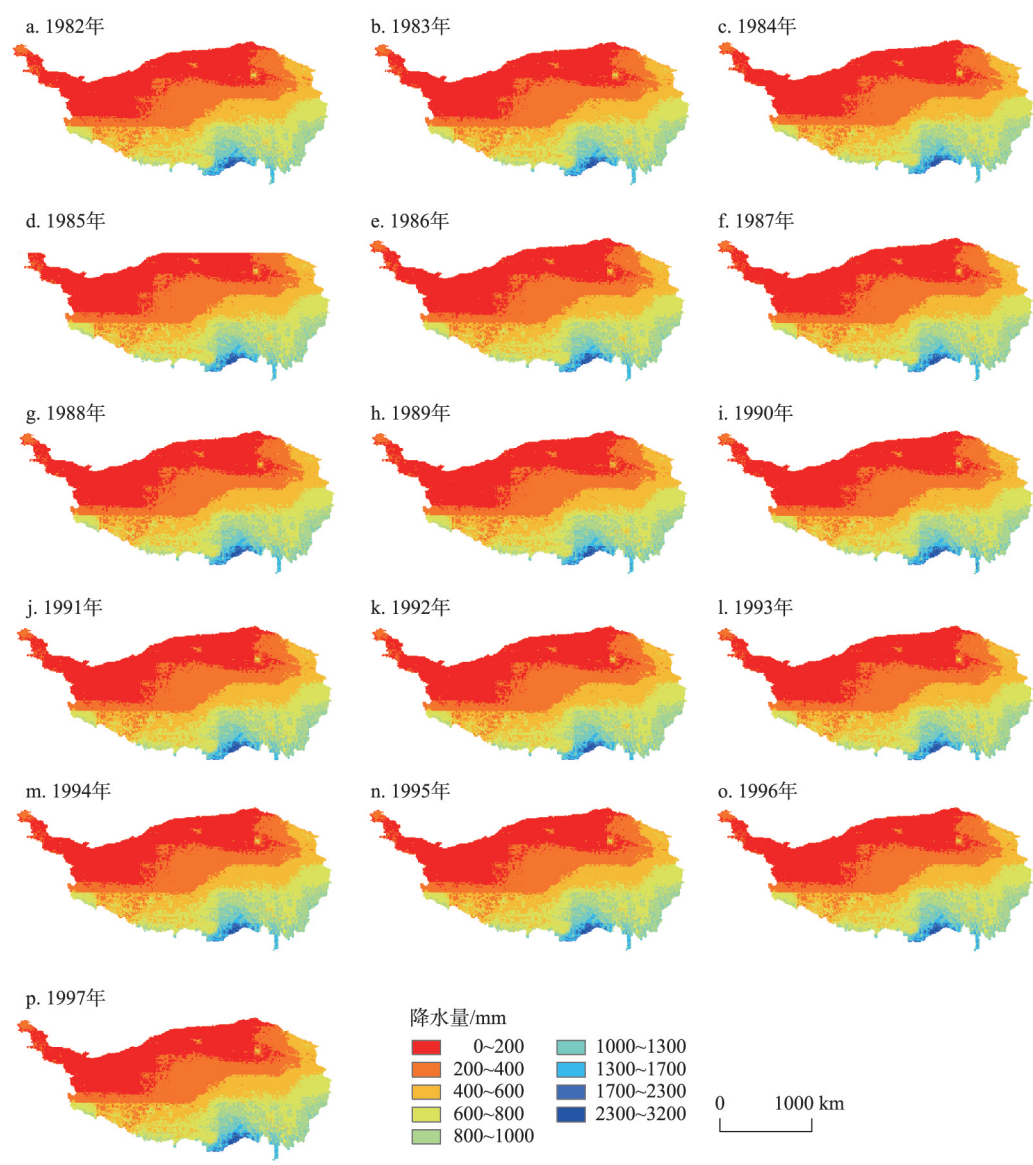


图3 1982-1997年青藏高原年模拟降水量图
Fig.3 Annual precipitation on the Tibetan Plateau, 1982-1997

拉湾湿暖气流的影响,降水量较多,形成一个向北伸展的降水充沛的“舌状地带”,雅鲁藏布江大拐弯区域的月降水量最高可达600 mm以上(齐文文等, 2013)。东北部有少部分地区降水量较多,是由于祁连山脉的原因,整体呈现自东向西降水量逐渐减少的趋势(陈少勇等, 2007)。青藏高原的西北部大部分地区降水量较少,其原因在于:一方面,是由于喜马拉雅山、昆仑山等山脉对于水汽的阻截作用;另一方面,位于青藏高原的柴达木盆地海拔较低,距离水源较远。

根据重建数据可知,青藏高原月降水量有季节

性变化,从4月份开始,高原雨季从东南部开始逐渐向西移动,春季3月、4月、5月份降水量在100~200 mm之间,而夏季的6月、7月、8月份由于大部分地区进入雨季,降水明显增多,最大降水量在300~600 mm;夏季过后,秋季9月、10月、11月降水量与夏季相比明显的减少;青藏高原东南部地区雨季开始早、结束晚,降水量大(齐文文等, 2013)。而在冬季12月、1月、2月,青藏高原大部分地区几乎没有降水,只有西南部地区和东南少部分地区有少量降水。

为了比较不同年份在同一月份的降水模拟效

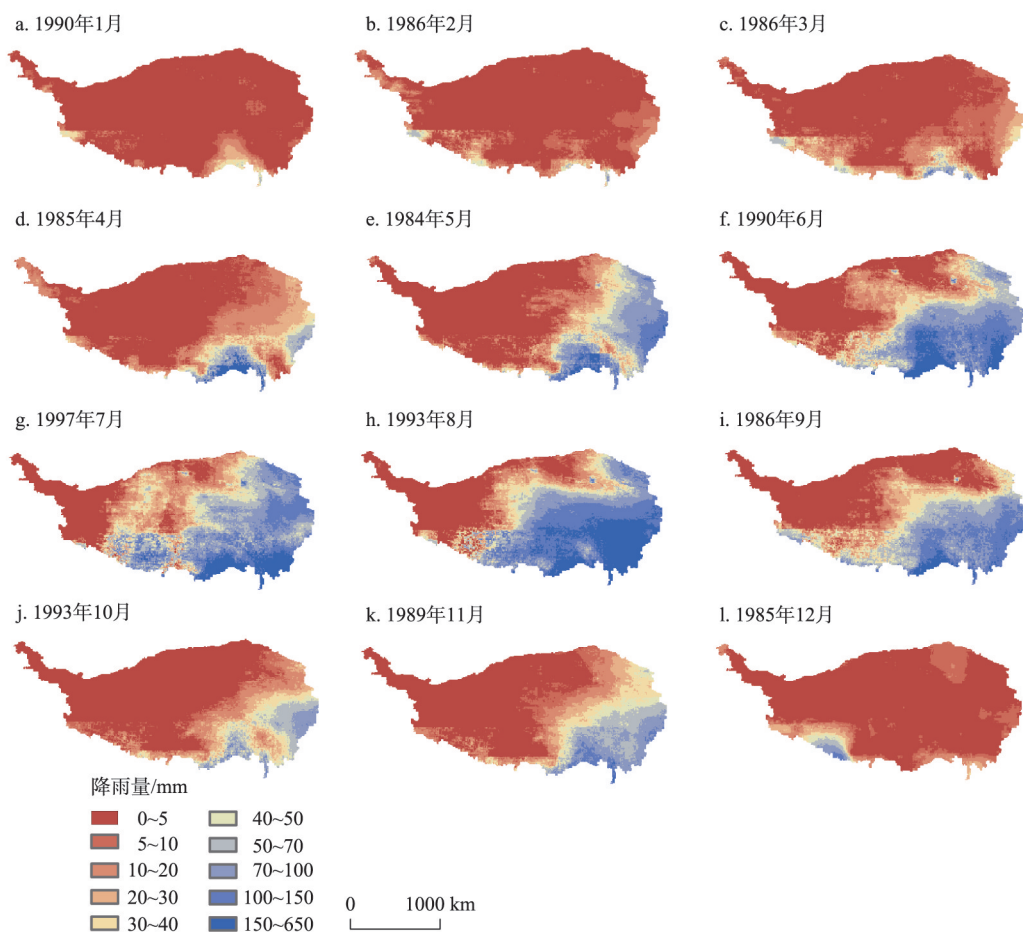


图4 1982-1997年青藏高原各月最高精度月降水图

Fig.4 Highest precision monthly precipitation of the Tibetan Plateau, 1982-1997

果,基于历史重建1982-1997年间1-12月各个月份的降水模拟结果,其中4月份的降水在1982-1997年间模拟的效果最好,因此选取了1982-1997年4月份月降水进行分析,如图5所示。月降水与年降水具有一致性,仍具有阶梯状分布特征,从东南部向西北部降水逐渐增多。其中1990年和1992年4月份月降水存在异常情况。1990年中部地区降水量偏高,1992年西部地区降水量偏高,这2个年份的模型模拟降水精度较低。

4.3 月降水精度评价

本文采用决定系数 R^2 、均方根误差 $RMSE$ 、平均绝对误差 MAE 、偏差 $Bias$ 对月降水的精度进行评价,如图6所示。月降水精度总体较高,但由于在高原西部地区气象站点和观测数据不足,重建结果在西部地区例如羌塘高原等,其精度评价受到限制。从图6a-6d可以看出, R^2 、 $Bias$ 值在校准前后变化不大,而 $RMSE$ 、 MAE 的值在校准前后变化较大,尤其

是在6月、7月、8月份校正后明显降低。总体上,计算站点数据与模型模拟月降水的差值,用克里金插值校准,有利于提高模型模拟月降水精度。图6a中可以看出,月降水数据降尺度的 R^2 变化较大,模型模拟效果具有明显的季节性,在春季3月、4月、5月 R^2 值较高,如1986年3月可高达0.97,模拟月降水精度较高,主要是由于春季植被生长与降水的相关性较好,模型模拟效果较好。而在冬季12月、1月、2月 R^2 的值较低,其中12月在各年份中月降水模拟精度最低,主要是由于在冬季植被生长与降水相关性较差。在6-11月, R^2 较稳定,通常在0.4~0.5之间,主要是由于夏季和秋季植被生长与降水有一定相关性,因此夏季 R^2 介于春季与冬季 R^2 值之间。图6b-6d总体上与 R^2 的趋势相同,具有明显的季节性变化,而随机森林模型中的高程、坡度、坡向、经度、纬度因子不受季节影响,而植被生长与降水的相关性具有季节性变化,因此降水模拟的精度会受到植

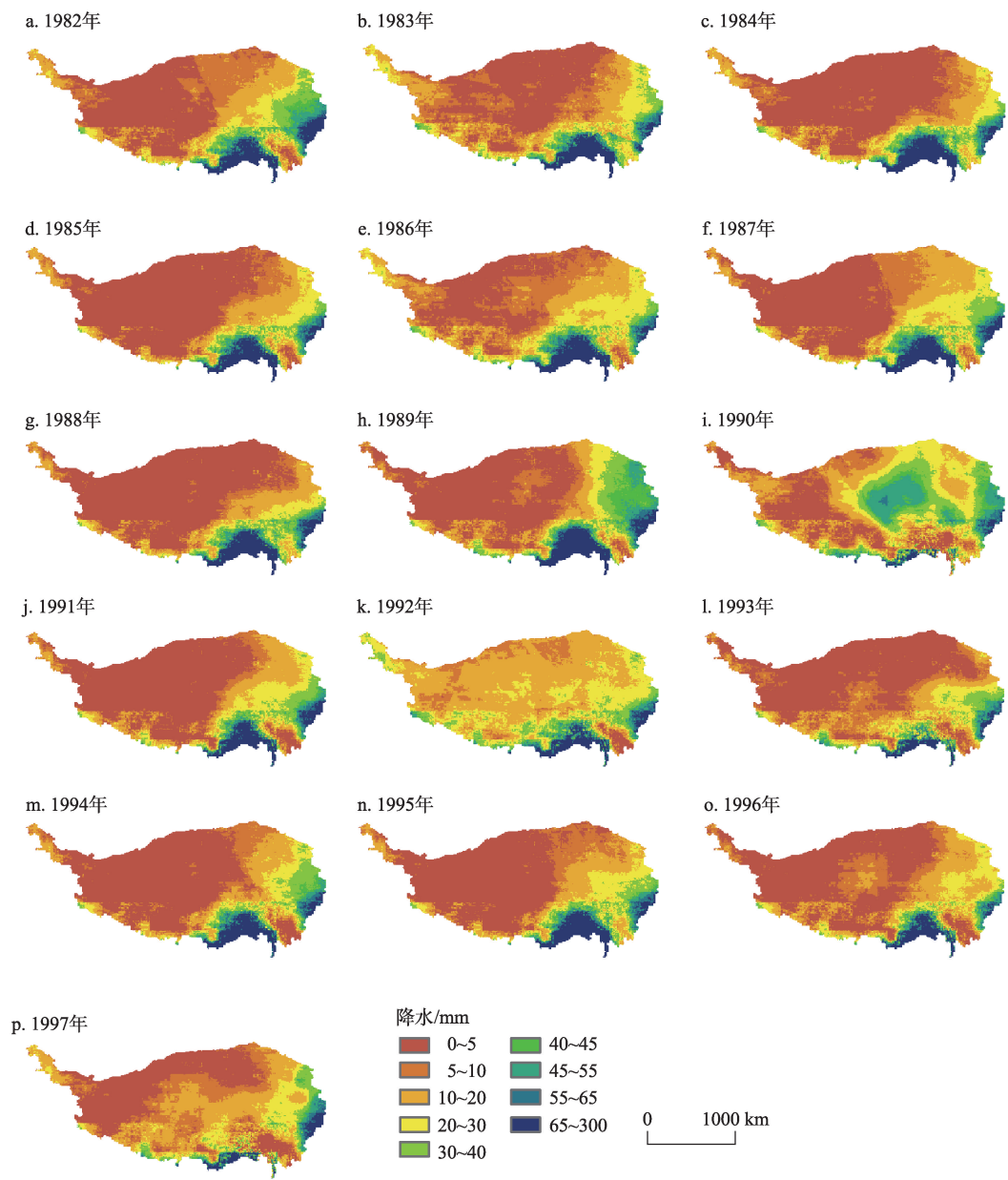


图5 1982-1997年4月份青藏高原月降水图
Fig.5 Monthly precipitation of the Tibetan Plateau in April, 1982-1997

被指数与降水相关性的影响。具体而言,图 6b 中,均方根误差 $RMSE$ 的年内周期性变化较明显,其中 6 月、7 月、8 月 $RMSE$ 值较高,其他月份较低。未校正月降水的 $RMSE$ 明显大于校正后月降水 $RMSE$ 值,且校正后模型模拟月降水的 $RMSE$ 通常在 50 mm 以下。图 6c 中 MAE 的总体趋势与 $RMSE$ 较相似,周期性明显,且校正后有明显改善。图 6d 可以看出模型模拟月降水偏差较小,精度较高,只有极少的月份偏差较大。例如 1984 年 12 月偏差为

1.393,1986 年 12 月偏差为 1.597,主要是由于冬季 NDVI 与降水的相关性较低,因此在模拟冬季降水时精度会受到影响,偏差较大。

5 讨论

(1) 青藏高原地区年降水自西北向东南逐渐增加。西北部地区降水量较少,降水量在 200 mm 以下,主要是由于柴达木盆地海拔相对较低,距离水

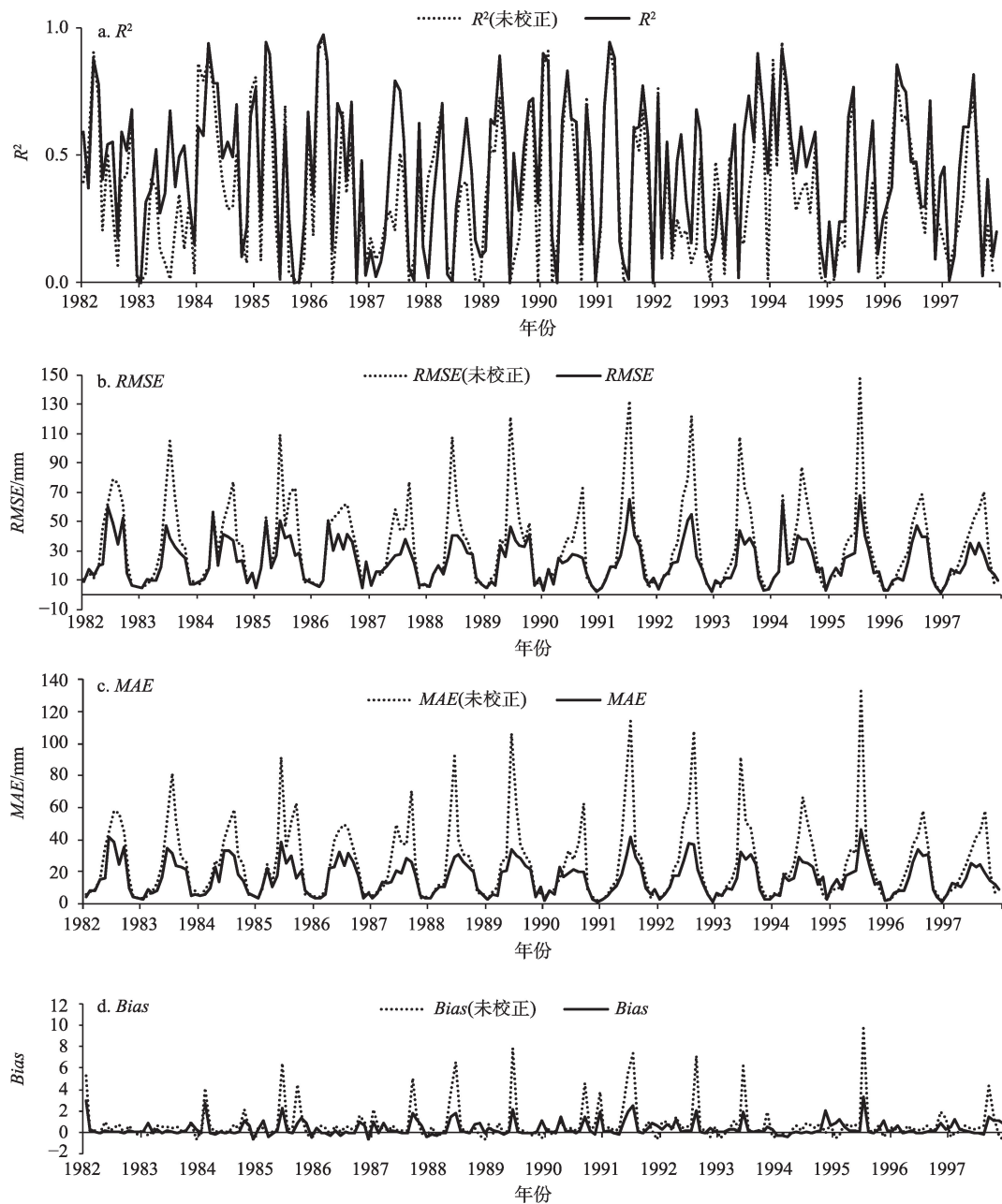


图6 1982-1997年青藏高原地区未校准和校准后月降水数据

Fig.6 Non-calibrated and calibrated monthly precipitation data of the Tibetan Plateau, 1982-1997

源较远,以及喜马拉雅山、昆仑山等山脉对于水汽的阻截作用;中部地区降水量在200~600 mm之间;东南部地区降水较多,降水量在600 mm以上,在东南部边缘地区,有一个向北伸展的降水充沛的“舌状地带”,降水量高达1300 mm以上,是由于受到青藏高原南部孟加拉湾水汽的影响。

(2) 根据重建的青藏高原0.1°分辨率的数据可知,青藏高原月降水在空间上,具有季节性变化。相较于站点获取的离散数据,重建数据更有利于降

水空间分析。同时,对比低分辨率的降水遥感数据,重建数据可获取更多局地降水细节信息。春季大部分地区降水量在100~200 mm之间,且高原雨季开始从东南部逐渐向西北移动;夏季降水明显增多,降水量主要在300~600 mm之间;秋季降水量明显减少,高原雨季开始向东南地区缩小;冬季高原大部分地区几乎没有降水。

(3) 将年降水降尺度到月降水时,采用了比例系数法,将年降水分配到各个月份,而并非直接在

月尺度上构建随机森林模型模拟降水,主要是由于NDVI与降水的关系具有滞后性,不适合进行月降水的模拟。

6 结论

(1) 本文重建了青藏高原地区1982–1997年间0.0833°分辨率的月降水数据,从而解决了青藏高原实测降水数据缺失,观测站点较少且分布不均匀的问题,并且弥补了TRMM 3B43数据分辨率较低的不足,为该地区提供了较高分辨率月尺度的降水数据。

(2) 月降水模型模拟精度有季节性差异。在春季模拟效果最好,冬季模拟效果最差,夏季和秋季模拟效果介于两者之间。主要是由于植被生长与降水的相关性具有季节性变化,NDVI作为随机森林模型的输入因子之一,因此降水模拟的精度会受到两者相关性的影响。

(3) 本文重建的月降水总体精度较高,但在冬季表现较差。在日后青藏高原降水研究中,可尝试添加其他相关因子,如地表温度等,进一步提高冬季月降水模拟精度。

参考文献(References)

陈少勇, 董安祥, 韩通. 2007. 祁连山东、西部夏季降水量时空分布的差异及其成因研究[J]. 南京气象学院学报, 30(5): 715–719. [Chen S Y, Dong A X, Han T. 2007. Differences in summer precipitation between the east and west of the Qilian Mountains and its contributing factors[J]. Journal of Nanjing Institute of Meteorology, 30(5): 715–719.]

杜加强, 舒俭民, 赵晨曦, 等. 2016. 两代 AVHRR GIMMS NDVI 数据集的对比分析: 以新疆地区为例[J]. 生态学报, 36(21): 6738–6749. [Du J Q, Shu J M, Zhao C X, et al. 2016. Comparison of GIMMS NDVI3G and GIMMS ND-VIG for monitoring vegetation activity and its responses to climate changes in Xinjiang during 1982–2006[J]. Acta Ecologica Sinica, 36(21): 6738–6749.]

傅抱璞. 1992. 地形和海拔高度对降水的影响[J]. 地理学报, 47(4): 302–314. [Fu B P. 1992. The effects of topography and elevation on precipitation[J]. Acta Geographica Sinica, 47(4): 302–314.]

何红艳, 郭志华, 肖文发, 等. 2005. 利用GIS和多变量分析估算青藏高原月降水[J]. 生态学报, 25(11): 2933–2938. [He H Y, Guo Z H, Xiao W F, et al. 2005. Mapping monthly precipitation for Tibetan Plateau with GIS and multivariate analysis based on dem data[J]. Acta Ecologica Sinica, 25(11): 2933–2938.]

侯光良, 鄂崇毅, 肖景义. 2012. 青藏高原全新世降水序列的集成重建[J]. 地理科学进展, 31(9): 1117–1123. [Hou G L, E C Y, Xiao J Y. 2012. Synthetical reconstruction of the precipitation series of the Qinghai-Tibet Plateau during the holocene[J]. Progress in Geography, 31(9): 1117–1123.]

黄聪敏, 陆德辉, 许艾米. 2014. 华南地区几种降水产品的对比分析[J]. 广东气象, 36(4): 19–24. [Huang C M, Lu D H, Xu A M. 2014. Comparisons and analysis of a number of precipitation products for the region of South China[J]. Guangdong Meteorology, 36(4): 19–24.]

李欣海. 2013. 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, 50(4): 1190–1197. [Li X H. 2013. Using "random forest" for classification and regression[J]. Chinese Journal of Applied Entomology, 50(4): 1190–1197.]

林振耀, 吴祥定. 1990. 青藏高原水汽输送路径的探讨[J]. 地理研究, 9(3): 33–40. [Lin Z Y, Wu X D. 1990. A preliminary analysis about the tracks of moisture transportation on the Qinghai-Xizang Plateau[J]. Geographical Research, 9(3): 33–40.]

刘永和, 郭维栋, 冯锦明, 等. 2011. 气象资料的统计降尺度方法综述[J]. 地球科学进展, 26(8): 837–847. [Liu Y H, Guo W D, Feng J M, et al. 2011. A summary of methods for statistical downscaling of meteorological data[J]. Advances in Earth Science, 26(8): 837–847.]

明均仁, 肖凯. 2012. 基于R语言的面向需水预测的随机森林方法[J]. 统计与决策, (9): 81–83. [Ming J R, Xiao K. 2012. Jiyou R yuyan de mianxiang xushui yuce de suijisenlin fangfa[J]. Statistics & Decision, (9): 81–83.]

齐文文, 张百平, 庞宇, 等. 2013. 基于TRMM数据的青藏高原降水的空间和季节分布特征[J]. 地理科学, 33(8): 999–1005. [Qi W W, Zhang B P, Pang Y, et al. 2013. TRMM-data-based spatial and seasonal patterns of precipitation in the Qinghai-Tibet Plateau[J]. Scientia Geographica Sinica, 33(8): 999–1005.]

吴国雄. 2004. 我国青藏高原气候动力学研究的近期进展[J]. 第四纪研究, 24(1): 1–9. [Wu G X. 2004. Recent progress in the study of the Qinghai-Xizang Plateau climate dynamics in China[J]. Quaternary Sciences, 24(1): 1–9.]

徐祥德, 赵天良, 施晓晖, 等. 2015. 青藏高原热力强迫对中国东部降水和水汽输送的调制作用[J]. 气象学报, 73(1): 20–35. [Xu X D, Zhao T L, Shi X H, et al. 2015. A study of the role of the Tibetan Plateau's thermal forcing in modulating rainband and moisture transport in eastern China[J]. Acta Meteorologica Sinica, 73(1): 20–35.]

Beck H E, McVicar T R, van Dijk A I J M, et al. 2011. Global evaluation of four AVHRR-NDVI data sets: Intercomparison and assessment against Landsat imagery[J]. Remote Sensing of Environment, 115(10): 2547–2563.

Chen C, Zhao S H, Duan Z, et al. 2015. 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, 8(9): 4592-4604.
- Duan Z, Bastiaanssen W G M. 2013. First results from version 7 TRMM 3B43 precipitation product in combination with a new downscaling-calibration procedure[J]. Remote Sensing of Environment, 131: 1-13.
- Immerzeel W W, Rutten M M, Droogers P. 2009. Spatial downscaling of TRMM precipitation using vegetative response on the Iberian Peninsula[J]. Remote Sensing of Environment, 113(2): 362-370.
- Jia S F, Zhu W B, Lü A F, et al. 2011. A statistical spatial downscaling algorithm of TRMM precipitation based on NDVI and DEM in the Qaidam Basin of China[J]. Remote Sensing of Environment, 115(12): 3069-3079.
- Maronna R. 2011. Richard Berk: Statistical learning from a regression perspective[J]. Statistical Papers, 52(4): 981-982.
- Pang G J, Wang X J, Yang M X. 2017. Using the NDVI to identify variations in, and responses of, vegetation to climate change on the Tibetan Plateau from 1982 to 2012[J]. Quaternary International, 444: 87-96.
- Prakash S, Mitra A K, Momin I M, et al. 2015. A review of recent evaluations of TRMM multisatellite precipitation analysis (TMPA) research products against ground-based observations over Indian land and oceanic regions[J]. Mausam, 66(3): 355-366.
- Shi Y L, Song L. 2015. Spatial downscaling of monthly TRMM precipitation based on EVI and other geospatial variables over the Tibetan Plateau from 2001 to 2012[J]. Mountain Research and Development, 35(2): 180-194.
- Xu S G, Wu C Y, Wang L, et al. 2015. A new satellite-based monthly precipitation downscaling algorithm with non-stationary relationship between precipitation and land surface characteristics[J]. Remote Sensing of Environment, 162: 119-140.

Reconstruction of high resolution monthly precipitation data of the Tibetan Plateau

XU Ming, SHI Yuli, WANG Bin

(School of Geography and Remote Sensing, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: The Tibetan Plateau is of great significance to the study of global climate. Precipitation data are also important to hydrological, meteorological, and ecological research. With the change of research content and scale, the demand for historical precipitation data with high spatial and temporal resolutions is increasingly more urgent. This study selected six factors including the normalized vegetation index(NDVI), elevation, slope, longitude, latitude, and precipitation to obtain the historical precipitation data with high spatial and temporal resolutions of the Tibetan Plateau from 1982 to 1997. This research is based on the precipitation data of TRMM 3B43, AVHRR NDVI, and SRTM DEM. The factors were entered into the random forest model. A historical precipitation model was constructed to obtain the spatial resolution of 0.0833° for annual precipitation. Monthly precipitation was estimated according to the proportional coefficient. In order to improve estimation accuracy, the simulated monthly precipitation was corrected using station data. The results show that the method can simulate historical precipitation with high spatial and temporal resolutions. The coefficient of determination R^2 is between 0.4 and 0.9, with an average value of 0.6767. The model performed better in summer and worse in winter. The root mean square error and average absolute error is below 50 mm. *RMSE* mean is 22.66 mm. *MAE* mean is 15.97 mm. Deviation bias is between 0.0 and 0.1.

Key words: random forest; historical precipitation; reconstruction; TRMM 3B43; Tibetan Plateau