

横断山区降水时空分布数据集 (1998-2012)

张 涛^{1,2}, 李宝林^{1,2*}, 袁焯城¹, 高锡章¹, 蒋育昊^{1,2}, 刘 岩^{1,2}

1. 中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室, 北京 100101;

2. 中国科学院大学, 北京 100049

摘 要: 1998-2012 年横断山区降水时空分布数据集是利用地基站降水数据对 1998-2012 年的 TRMM 3B43 V7 降水数据进行订正后生成的该区域降水数据及其相关对照或辅助数据。其中, 订正后的 TRMM 数据是基于发表在《自然资源学报》上的论文中提出的分区逐月回归法生成的。该方法包括三个步骤: (1) 采用基于 ISODATA 的动态聚类法和最大似然法相结合的方法对研究区进行分区, 并确保每个子区域内至少包含一个气象站点; (2) 以月为时间分析单元, 在每个子区域内以站点实测降水为因变量, TRMM 降水估值为自变量建立线性回归方程; (3) 对于任一栅格点, 用其所在子区域的回归方程修正 TRMM 降水估值。结果表明, 订正后的 TRMM 3B43 数据精度明显提高; 在 38 个气象站点中, 相对偏差值大于 10% 的站点数从 16 个减少到 8 个, 对应于从站点总数的 42.1% 减少到 21.1%; 相对偏差值大于 20% 的站点数从 12 个减少到 3 个, 对应于从站点总数的 31.6% 减少为 7.9%。数据集采用 WGS84 坐标系, 矢量数据采用 shapefile 格式存储, 栅格数据采用 Grid 格式存储。

关键词: 降水; TRMM 卫星; 数据订正; 横断山区

DOI: 10.3974/geodp.2019.02.07

1 前言

准确获取降水时空分布信息对水文学、气象学以及气候学等领域的研究具有重要意义。除了大气环流、海陆位置等因素, 山区降水还主要受到下垫面地形的影响, 具有极为强烈的时空异质性。基于卫星遥感反演的方法具有大范围空间连续观测的优势, 可以克服传统地面雨量计和地基测雨雷达受站点分布和观测范围限制的问题, 从而能够更好地反映山区降水的时空分布特征。但是, 卫星降水数据的精度具有空间异质性, 特别是在地形复杂的山区, 降水时空分布特征较为复杂, 直接使用卫星降水反演数据用于相关的应用分析具有很大的不确定性^[1-2]。因此, 需要对卫星降水反演产品进行订正后才能更为准确地反映区域的降水时空分布特征^[3-4]。

收稿日期: 2019-04-25; **修订日期:** 2019-06-17; **出版日期:** 2019-06-25

基金项目: 中华人民共和国科学技术部 (2016YFC0500205, 2015CB954103)

***通讯作者:** 李宝林 N-8884-2019, 中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室, libl@lreis.ac.cn

数据引用方式: [1] 张涛, 李宝林, 袁焯城等. 横断山区降水时空分布数据集(1998-2012) [J]. 全球变化数据学报, 2019, 3(2): 178-184. DOI: 10.3974/geodp.2019.02.07.

[2] 张涛, 李宝林, 何元庆等. 横断山区降水时空分布数据集(1998-2012) [DB/OL]. 全球变化科学研究数据出版系统, 2016. DOI: 0.3974/geodb.2016.06.06.V1.

横断山区降水时空分布数据集（1998–2012）^[5]由中华人民共和国科学技术部项目资助完成。本产品是基于目前可以方便获取的 0.25°×0.25°分辨率 TRMM 3B43 V7 数据产品^[6–7]研发的。针对该数据产品在横断山区精度较差的问题，假设其反映的降水分布信息在不同的空间子区域内与真实的地面降水之间具有系统性的偏差，从而采用分区逐月回归法对 TRMM 数据进行订正，即基于 ISODATA 动态聚类法和最大似然法^[8]对 TRMM 数据进行非监督分类，划分出不同的子区域，用地面站点数据在各个子区域内建立真实降水与 TRMM 降水估值之间的线性回归方程，并利用该方程对 TRMM 数据进行订正。订正后的产品与基于地面站点插值的方法生成的结果进行对比分析。

2 数据集元数据简介

横断山区降水时空分布数据集（1998–2012）^[5]的名称、作者、地理区域、数据年代、时间分辨率、空间分辨率、数据集组成、数据出版与共享服务平台、数据共享政策等信息见表 1。

表 1 横断山区降水时空分布数据集元数据简表

条 目	描 述
数据集名称	横断山区降水时空分布数据集（1998–2012）
数据集短名	SpaTempPrecipHengduanMt1998_2012
作者信息	张涛 N-8690-2019, 中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室, zhangtao@lreis.ac.cn 李宝林 N-8884-2019, 中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室, libl@lreis.ac.cn 何元庆, 中国科学院寒区旱区环境与工程研究所, yqhe@lzb.ac.cn 杜建括, 中国科学院寒区旱区环境与工程研究所, djklzu2008@163.com 牛贺文, 中国科学院寒区旱区环境与工程研究所, niuhw11@lzu.edu.cn 辛惠娟, 中国科学院寒区旱区环境与工程研究所, xinhj642@163.com
地理区域	24.5°N –34.5°N, 93.5°E–104.5°E。包括川西、滇西北以及藏东南等区域内的 38 个县或市，即八宿、巴塘、昌都、大理、道孚、稻城、德格、德钦、丁青、甘孜、贡山、红原、华坪、会理、九龙、康定、雷波、类乌齐、理塘、丽江、洛隆、马尔康、芒康、木里、攀枝花、若尔盖、色达、石渠、松潘、维西、西昌、香格里拉、小金、新龙、盐源、越西、昭觉、左贡
数据年代	1998–2012 年
时间分辨率	3 months
空间分辨率	0.25°×0.25°
数据格式	.shp、.Grid
数据量	773 KB（压缩后）

续表

条 目	描 述
数据集组成	<p>数据集由 6 个文件夹和 1 个 Excel 文件组成：</p> <p>(1) 文件夹“1_StudyArea_Stations”内包含“StudyArea”和“Stations_HDMt”两个子文件夹，分别存储了 StudyArea.shp 和 Stations_HDMt.shp 两个矢量文件，分别代表研究区范围和气象站点位置；</p> <p>(2) 文件夹“2_1_CorrectedTRMMdata_1998-2012”内包含 1998 至 2012 年各个年份 Grid 格式的文件，如 rhdm_1998、rhdm98_spr、rhdm98_sum、rhdm98_aut、rhdm98_win、rhdm98_mon 和 rhdm98_umon，分别代表：1998 年年度、春季、夏季、秋季、冬季、生长季以及非生长季等时段内订正后的 TRMM 3B43 累积降水量，其空间分辨率为 $0.25^{\circ}\times0.25^{\circ}$，单位为 mm</p> <p>(3) 文件夹“2_2_CorrectedAverageTRMMdata”内包含 7 个 Grid 格式文件：Ann_ave、Spr_ave、Sum_ave、Aut_ave、Win_ave、Mon_ave 和 Umon_ave 分别代表：1998 至 2012 年多年平均的年度、春季、夏季、秋季、冬季、生长季以及非生长季等时段内订正后的 TRMM 3B43 累积降水量，其空间分辨率为 $0.25^{\circ}\times0.25^{\circ}$，单位为 mm</p> <p>(4) 文件夹“2_3_Kriging_StationData”内包含 7 个 Grid 格式文件：Ann_kriging、Spr_kriging、Sum_kriging、Aut_kriging、Win_kriging、Mon_kriging 和 Umon_kriging，分别代表：1998 至 2012 年多年平均的年度、春季、夏季、秋季、冬季、生长季以及非生长季等时段内基于站点数据 kriging 插值生成的累积降水量，其空间分辨率为 $0.25^{\circ}\times0.25^{\circ}$，单位为 mm</p> <p>(5) 文件夹“2_4_RainfallTendencyRate”内包含 7 个 Grid 格式文件：Ann_RTR、Spr_RTR、Sum_RTR、Aut_RTR、Win_RTR、Mon_RTR 和 Umon_RTR，分别代表 1998 至 2012 年年度、春季、夏季、秋季、冬季、生长季以及非生长季等时段内由订正后的 TRMM 数据计算的降水倾向率，其空间分辨率为 $0.25^{\circ}\times0.25^{\circ}$，单位为 mm/年</p> <p>(6) 文件夹“2_4_SignificantLevel”内包含 7 个 Grid 格式文件：Ann_sig、Spr_sig、Sum_sig、Aut_sig、Win_sig、Mon_sig 和 Umon_sig，分别代表：1998 至 2012 年年度、春季、夏季、秋季、冬季、生长季以及非生长季等时段内由订正后的 TRMM 数据计算的降水倾向率对应的显著性水平，其空间分辨率为 $0.25^{\circ}\times0.25^{\circ}$</p> <p>(7) Excel 文件“Attachment.xlsx”内包含 2 个表格：Tab.1 和 Tab.2。其中，Tab.1 记录了横断山区 38 个气象站点逐月降水观测数据（1998-2012），Tab.2 记录了横断山区 38 个气象站所在 TRMM 格点逐月降水估值数据（1998-2012）</p>
基金项目	中华人民共和国科学技术部 (2016YFC0500205, 2015CB954103)
出版与共享服务平台	全球变化科学研究数据出版系统 http://www.geodoi.ac.cn
地址	北京市朝阳区大屯路甲 11 号 100101，中国科学院地理科学与资源研究所
数据共享政策	<p>全球变化科学研究数据出版系统的“数据”包括元数据（中英文）、实体数据（中英文）和通过《全球变化数据学报》（中英文）发表的数据论文。其共享政策如下：（1）“数据”以最便利的方式通过互联网系统免费向全社会开放，用户免费浏览、免费下载；（2）最终用户使用“数据”需要按照引用格式在参考文献或适当的位置标注数据来源；（3）增值服务用户或以任何形式散发和传播（包括通过计算机服务器）“数据”的用户需要与《全球变化数据学报》（中英文）编辑部签署书面协议，获得许可；（4）摘取“数据”中的部分记录创作新数据的作者需要遵循 10%引用原则，即从本数据集中摘取的数据记录少于新数据集总记录量的 10%，同时需要对摘取的数据记录标注数据来源^[9]</p>

3 数据研发方法

3.1 算法原理

横断山区受多种环流交替控制，因而不同月份或季节的降水时空分布特征差异非常大^[10-11]。受降水时空分布差异的影响，横断山区的 TRMM 降水估值精度也存在较大区域差异。算法原理是假设降水时空变化较为一致的空间单元内具有相同的降水形成机理，于

是在上述空间单元内分别建立订正模型，减少降水高时空变异性对订正模型的影响，从而提高降水估值的精度。首先，对横断山区进行分区，得到多个降水空间分布模式较为接近的子区域；然后，在每个子区域内建立 TRMM 降水与地面实测降水之间的回归方程；最后，在每个子区域内基于回归方程对 TRMM 降水格点值进行订正。

3.2 技术路线

1998-2012 年 TRMM 降水订正产品采用分区逐月回归法制作（图 1）。该方法的主要步骤包括：（1）确定空间和时间分析单元。采用非监督分类将研究区划分为不同的子区域作为 TRMM 数据订正的空间分析单元，并确保每个空间单元内至少包含一个气象站点。采用“月”作为时间分析单元，从而确保分析时段内降水具有相似的空间分布模式。（2）建立统计回归模型。对每个子区域内的任一月份，以站点降水观测值作为因变量，TRMM 降水估值作为自变量，建立回归方程。（3）利用回归模型进行降水订正。对于任一栅格点，用其所在子区域的回归方程修正 TRMM 降水估值。

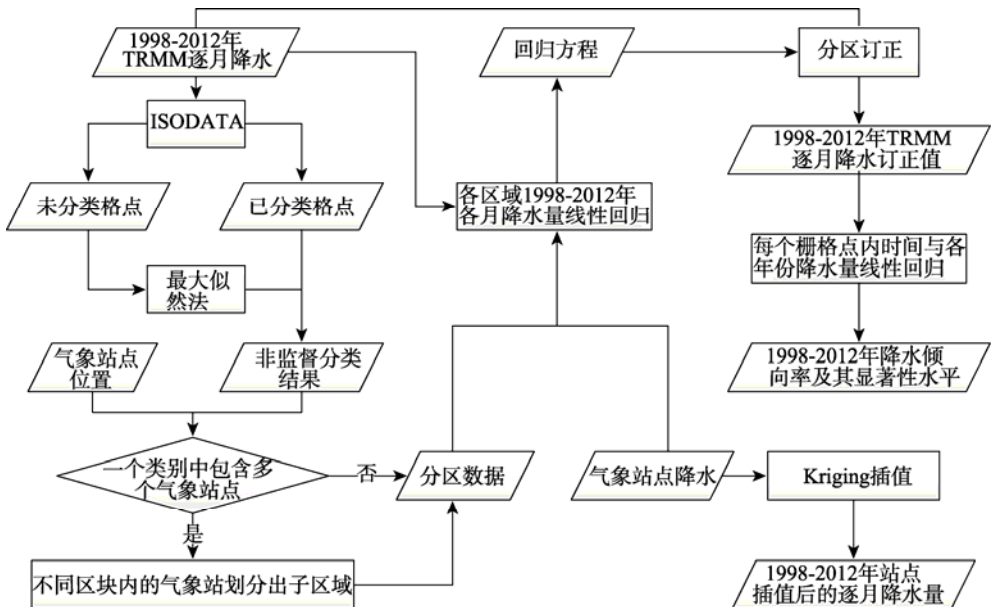


图 1 横断山区降水时空分布数据集生产技术路线图

划分空间分析单元采用 ISODATA 动态聚类法，并结合最大似然法实现。其具体步骤如下：（1）栅格点非监督分类。采用 ISODATA 聚类法对 TRMM 数据进行非监督分类。（2）判定未分类栅格点的类别。采用最大似然法将 ISODATA 算法迭代结束后仍然未确定类别的栅格点分配到与其降水变化特征最相似的类别中。（3）分类后处理。在同一类别中，根据不同区块内的气象站点划分出子区域，其余区块标记为距离其最近的气象站所在子区域。

1998-2012 年降水倾向率及其显著性水平数据采用相应时期订正后的 TRMM 数据与时间在每个栅格点作线性回归制作。1998-2012 年降水插值产品是根据研究区内气象站点在相应时期的累积降水量，利用 kriging 进行空间插值制作（图 1）。

4 数据结果与验证

数据集组成如表 1 所示。表 1 对数据集中包含的 6 个文件夹和 1 个 Excel 文件进行了详细的介绍。文件夹中主要包含了研究区边界以及气象站点位置数据、站点降水插值数据、订正后的 TRMM 降水数据、降水倾向率及其显著性数据，Excel 表中则包含了研究区内各气象站点处降水观测数据以及站点所在 TRMM 格点的降水估值数据。

4.1 数据空间分布

图 2 展示了 1998–2012 年横断山区多年平均降水量的空间分布情况。从图 2 可知，订正后的 TRMM 降水与基于站点实测数据插值生成的降水具有相似的空间分布特征，但订正后的 TRMM 降水具有更多空间细节信息。上述两种数据都显示出降水从研究区东南向西北逐渐递减的趋势，且降水高值和低值中心位置相近。但是，订正后的 TRMM 降水可较为准确地反映出无站点区域特别是研究区边缘附近的降水分布状况，比基于站点实测降水插值结果展示出更多的空间细节信息。

图 3 展示了 1998–2012 年横断山区年度降水倾向率的空间分布状况。从图 3 可知，该区域的年降水量主要呈现出减少趋势，降水减少的区域位于研究区南部以及中西部地区，且大部分呈现降水减少趋势的区域可以达到 0.05 的显著性水平；而降水呈现增加趋势的区域位于研究区东北部地区，但其趋势不具有显著性。

4.2 数据结果验证

图 4 展示了 1998–2012 年横断山区 TRMM 3B43 年平均降水估值相对误差空间分布状况。大部分站点订正后的 TRMM 多年平均降水偏差绝对值变小，精度明显改善。降水相对偏差大于 10% 的站点数从 16 个减少到 8 个，对应于从站点总数的 42.1% 减少为 21.1%；相对偏差大于 20% 的站点数从 12 个减少到 3 个，对应于从站点总数的 31.6% 减少为 7.9%；其中，严重高估的八宿、洛隆和左贡 3 个站点的降水相对偏差分别从 129.2%、76.2% 和 65.9% 减少为 37.7%、21.1% 和 14.6%。

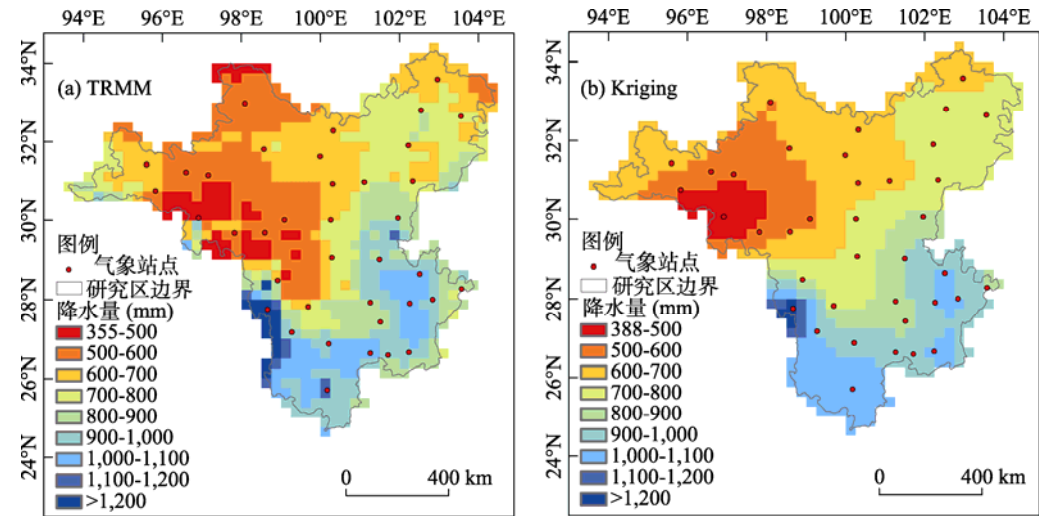


图 2 1998–2012 年横断山区多年平均降水空间分布^[4]

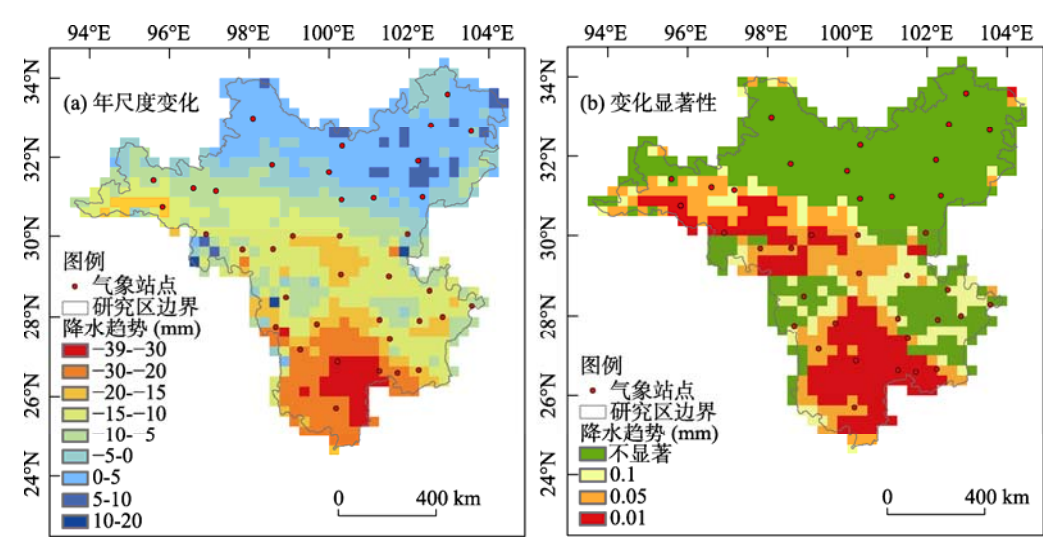


图 3 1998–2012 年横断山区年度降水倾向率空间分布^[4]

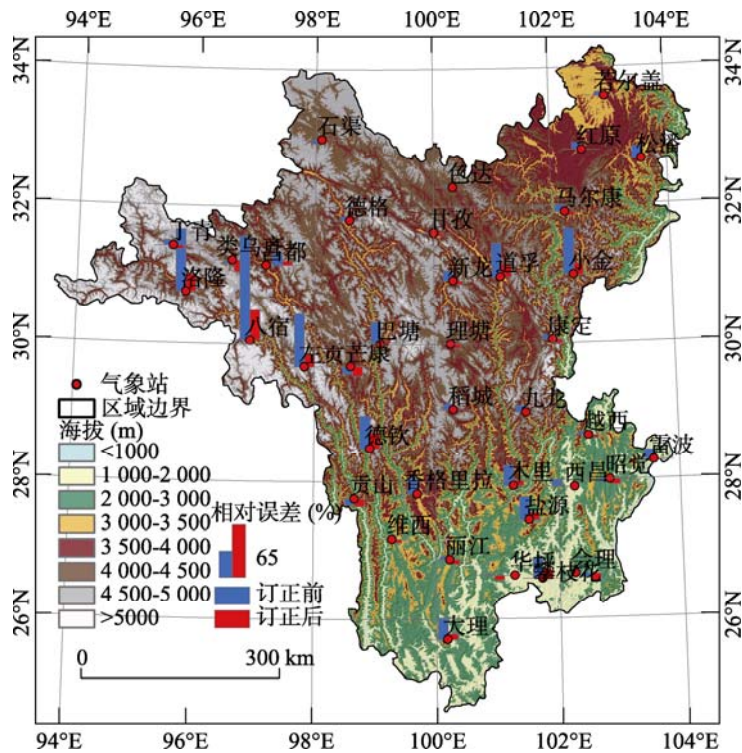


图 4 1998–2012 年横断山区 TRMM 3B43 年平均降水估值相对误差空间分布

同时，图 4 还表明相对误差较大的区域在订正后仍然具有相对较低的精度。如上述八宿、洛隆和左贡等站点的降水相对偏差仍然较高。这三个站点都位于研究区西北部的高山区，地形地貌尤为复杂。可见，通过本数据集中采用的订正方法，能够使下垫面条件较为复杂的高山区的降水反演精度得到提高，但仍然较差。

5 讨论和总结

本数据集的生成算法主要考虑了降水形成机理的时空异质性, 基于非监督分类得到降水相对一致的时空单元(子区域), 并在每个子区域内基于地面实测降水建立回归模型对 TRMM 数据进行订正, 从而提高了订正结果的精度, 为该区域的降水时空分布特征分析提供了更为准确的数据产品。

数据集生成算法的不足之处主要在于没有考虑该区域的地形地貌因素和降水观测的尺度差异。未考虑地形地貌因素可能带来的影响包括: (1) 基于非监督分类得到的子区域可能无法与特定的自然区域完全对应, 子区域边界附近的不确定性较大; (2) 用于降水订正的回归模型中没有考虑海拔对降水分布的影响, 模型的校正能力可能受到限制。降水观测的尺度差异主要是指 TRMM 网格点与地面气象站点之间的空间尺度差异。TRMM 网格点较粗, 网格内可能包含复杂的地形变化, 而气象站点一般位于沟谷平原上, 为单点观测, 两者在尺度上并不匹配, 存在一定的不确定性。

后续的研究中将针对数据集生成算法存在的上述两类问题开展进一步改进实验: (1) 考虑地形地貌差异, 根据地貌边界或流域边界划分子区域, 同时, 在模型中加入海拔作为解释变量, 以便得到更符合客观规律的订正方法; (2) 先对 TRMM 数据进行降尺度, 然后再订正, 从而减少 TRMM 网格与地面气象站点之间的尺度差异。

作者分工: 李宝林, 张涛对数据集的开发做了总体设计; 蒋育昊, 刘岩采集和处理了气象站点和 TRMM 降水估值数据; 张涛, 袁烨城设计了模型和算法; 高锡章做了数据验证; 张涛, 李宝林撰写了数据论文等。

参考文献

- [1] 曾红伟, 李丽娟. 澜沧江及周边流域 TRMM 3B43 数据精度检验[J]. 地理学报, 2011, 66(7): 994-1004.
- [2] 朱国锋, 蒲焘, 张涛等. TRMM 降水数据在横断山区的精度[J]. 地理科学, 2013, 33(9): 1125-1131.
- [3] Zhang, T., He, Y., Ma, J., *et al.* Spatial and temporal distribution of precipitation based on corrected TRMM data around the Hexi Corridor, China [J]. *Sciences in Cold and Arid Regions*, 2014, 6(2): 159-167.
- [4] 张涛, 李宝林, 何元庆等. 基于 TRMM 订正数据的横断山区降水时空分布特征[J]. 自然资源学报, 2015, 30(2): 260-270.
- [5] 张涛, 李宝林, 何元庆等. 横断山区降水时空分布数据集(1998-2012) [DB/OL]. 全球变化科学研究数据出版系统, 2016. DOI: 10.3974/geodb.2016.06.06.V1.
- [6] Kummerow, C., Barnes, W. The tropical rainfall measuring mission (TRMM) sensor package [J]. *Journal of Atmospheric and Oceanic Technology*, 1998, 15: 809-817.
- [7] 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.
- [8] 胡凯衡, 崔鹏, 韩用顺等. 基于聚类和最大似然法的汶川灾区泥石流滑坡易发性评价[J]. 中国水土保持科学, 2012, 10(1): 12-18.
- [9] 全球变化科学研究数据出版系统. 全球变化科学研究数据共享政策. DOI: 10.3974/dp.policy. 2014.05 (2017 年更新).
- [10] 李吉均, 苏珍. 横断山冰川[M]. 北京: 科学出版社, 1996.
- [11] 李宗省, 何元庆, 辛惠娟等. 我国横断山区 1960-2008 年气温和降水时空变化特征[J]. 地理学报, 2010, 65(5): 563-579.