

DOI:10.12171/j.1000-1522.20240180

基于 LSTM 的雅砻江流域多源降水产品融合研究

王世宇¹ 赵胤懋¹ 李佳倩²

(1. 北京林业大学水土保持学院, 北京 100083; 2. 北京林业大学林学院, 北京 100083)

摘要:【目的】在全球气候变暖背景下, 降水的时空格局发生变化致使极端降水事件发生频率增多且强度加剧。区域降水作为流域水文过程研究的重要基础, 因受到诸如高寒山区、冻土冰川等地理环境的限制而使得相关检测工作难以全面推进。因此本研究旨在提出一个优化多源降水精度的框架, 为少资料或无资料地区的降水及水文模拟提供参考。【方法】以青藏高原地区的雅砻江流域为研究对象, 基于 19 个气象站点的实测降水和 ERA5、CHIRPS 及 HAR 3 套降水数据集, 引入地形因子和降水的季节特征, 利用长短期记忆人工神经网络方法构建出一套多源降水融合的机器学习模型, 并输出一套 1981—2017 年 $0.25^\circ \times 0.25^\circ$ 的栅格降水数据。【结果】最终输出的多源融合降水产品在研究时段内与实测降水的皮尔逊相关系数由融合前的 0.43 提高至 0.82, 均方根误差从均值 6.12 mm/d 下降到 5.45 mm/d, 其平均临界成功指数、检测概率和误报率分别为 0.64、0.92 和 0.33。基于最终输出的降水产品运用 VIC 水文模型分析得到研究时段内的纳什效率系数、相对误差和确定系数分别为 0.80、1.39% 和 0.89。【结论】本研究构建的多源降水融合机器学习模型在一定程度上可以提高雅砻江流域的时空降水产品质量, 并为少资料或无资料区域获取高精度降水产品并进行水文模拟提供一定参考。

关键词: 雅砻江流域; 多源降水; LSTM; 融合; 精度优化

中图分类号: S715; TV121⁺4 文献标志码: A 文章编号: 1000-1522(2024)11-0092-11

引文格式: 王世宇, 赵胤懋, 李佳倩. 基于 LSTM 的雅砻江流域多源降水产品融合研究 [J]. 北京林业大学学报, 2024, 46(11): 92-102. Wang Shiyu, Zhao Yinmao, Li Jiaqian. Merge of multi-source precipitation products in Yalong River Basin of southwestern China based on LSTM [J]. Journal of Beijing Forestry University, 2024, 46(11): 92-102.

Merge of multi-source precipitation products in Yalong River Basin of southwestern China based on LSTM

Wang Shiyu¹ Zhao Yinmao¹ Li Jiaqian²

(1. School of Soil and Water Conservation, Beijing Forestry University, Beijing 100083, China;

2. School of Forestry, Beijing Forestry University, Beijing 100083, China)

Abstract: [Objective] With the intensification of global warming, the spatiotemporal pattern of precipitation has changed, leading to an increase in the frequency and intensity of extreme precipitation events. As an important basis for the study of watershed hydrological processes, regional precipitation is constrained by certain geographical environments such as alpine mountainous areas and permafrost glaciers, which hinders the comprehensive implementation of related detection efforts. Therefore, this paper aims to propose a framework for optimizing the accuracy of multi-source precipitation, which provides a reference for precipitation and hydrological simulation in areas with little or no data. [Method] This study took the Yalong River Basin in the Tibetan Plateau of southwestern China as the research object. Based on the measured precipitation of 19 meteorological stations and three sets of precipitation data sets such as ECMWF re-analysis 5 (ERA5), climate hazards center infraRed precipitation with station data (CHIRPS) and the high asia refined analysis (HAR), the topographic factors and seasonal characteristics of precipitation

收稿日期: 2024-06-04 修回日期: 2024-08-13

基金项目: 中央高校基本科研业务费(BLX202215), 国家自然科学基金项目(42401017), 北京林业大学大学生创新创业训练项目(X202410022066)。

第一作者: 王世宇. 主要研究方向: 自然地理学. Email: shiyuwang@bjfu.edu.cn 地址: 100083 北京市海淀区清华东路 35 号北京林业大学水土保持学院。

责任作者: 赵胤懋, 博士, 讲师. 主要研究方向: 水文水资源. Email: scirenc@bjfu.edu.cn 地址: 同上。

本刊网址: <http://j.bjfu.edu.cn>; <http://journal.bjfu.edu.cn>

were introduced. Using the long short-term memory (LSTM) method, a multi-source merged precipitation machine learning model was constructed, and a set of gridded precipitation data with a resolution of $0.25^\circ \times 0.25^\circ$ from 1981 to 2017 was outputted. [Result] The Pearson correlation coefficient between final output multi-source merged precipitation product and actual precipitation during the study period increased from 0.43 before merging to 0.82, and the root mean square error decreased from an average of 6.12 to 5.45 mm/d. The average critical success index (C_{SI}), probability of detection (P_{OD}), and false alarm ratio (F_{AR}) were 0.64, 0.92, and 0.33, respectively. At the same time, based on the final output multi-source merged precipitation product, the VIC hydrological model was used for hydrological effect analysis, and the N_{ash} , RE, and C during the study period were 0.80, 1.39%, and 0.89, respectively. [Conclusion] The above results indicate that the multi-source merged precipitation machine learning model constructed in this study can improve the spatiotemporal quality of precipitation products in the Yalong River Basin to a certain extent, and provide a certain reference for obtaining high-precision precipitation products and conducting hydrological simulation in areas with little or no data.

Key words: Yalong River Basin; multi-source precipitation; LSTM; merge; accuracy optimization

降水数据是地区水文循环系统的重要动态输入,是分析径流的基础资料之一,并且在气候变化、水资源开发管理及灾害预警与监测等方面都发挥着重要的作用^[1-2]。然而,在全球气候变暖大背景下,降水的时空格局发生变化且伴随着极端降水事件发生频率提高,最终使得区域降水因受到诸如高寒山区、冻土冰川等环境因素的限制而使得检测工作难以全面推进^[2-3]。

目前,降水资料通常来源于雨量站、气象雷达和卫星等多源途径^[4]。传统的地面雨量观测站能测得较高精度的站点降水数据但受限于地面站点的空间分布,尤其是在地势落差大、地形复杂的地区,测站分布不均匀、密度低等问题会直接影响降水资料的精度和完整性^[5]。相比地面固定站点,气象雷达的实时性更好,能够快速提供时空连续变化的实时降水资料,但同样受限于组网分布情况以及架设成本。此外,雷达的降水数据还容易受到诸如地形遮挡、超折射回波和零度层亮带等多种因素的影响^[6]。随着卫星遥感技术的不断发展,利用卫星遥感通过算法反演得到的降水产品也广泛应用于水文气象领域的研究中。卫星遥感数据因其时空分辨率高、覆盖范围广,成为弥补地面观测空间覆盖度不高的重要方式^[7]。然而,在实际应用过程中由于受到传感器性能、云层性质和反演算法等因素的影响,卫星降水信息仍然存在相对较大的误差^[8]。

依据以上分析可发现,由于各类降水产品的误差无法避免,特别是对于特定区域,不同降水产品的精度、覆盖度和完整性都经受着极大的考验,为此开展多源降水数据的融合,充分整合各类降水产品的优势,取长补短以达到优化降水产品精度的目的是十分迫切的^[9-11]。目前,机器学习方法因能较好地处理降水与其他影响因子的非线性关系而得到了广泛

的应用^[12]。现今,大部分研究主要致力于融合方法的改进,使得融合模型的内部越来越复杂,在应用计算增加难度的同时也提高了模型的不确定性,且结果精度的提升也相对较小^[13]。其次,现有研究通常采用基于实测站点数据引入多类降水产品进行融合以达到提高精度的目的,却往往忽略了外部环境因素对降水产生的影响^[14]。另外,大部分致力于精度提升的研究中所采用的多源降水产品时间序列较短,限制了长时间序列气候和水文的模拟研究^[3,6,14]。

雅砻江流域地处长江上游,位于青藏高原东南部,因受复杂地形和季风的综合影响,其内部呈现出明显的区域特性和季节特征^[15]。正是由于雅砻江流域地形条件复杂、实测站点分布不均且部分区域资料匮乏,在一定程度上制约了该区域降水研究的发展。因此,如何在资料缺乏的条件下获取较为精确的降水数据并完成水文效应研究是一项亟待解决的问题。为此,本研究以雅砻江流域为研究对象,采用实测站点数据以及再分析、区域气候模式和卫星降水等多源降水产品,在引入地形因子和降水季节特征的基础上采用长短期记忆人工神经网络(long short term memory, LSTM)构建一套多源降水融合机器学习模型(a multi-source merged precipitation machine learning model based on terrain factors and seasonal precipitation characteristics, TFSP),并对最终输出降水产品进行水文效应的精度评估,为该地区降水产品的精度优化提供参考,也为少资料甚至无资料区域降水数据的获取提供帮助。

1 研究区概况及数据

1.1 雅砻江流域的区域概况

雅砻江位于青藏高原东南部($96^\circ 52' \sim 102^\circ 48' E$, $26^\circ 32' \sim 33^\circ 58' N$),发源于巴颜喀拉山南

麓,是长江上游金沙江的第一大支流(图1)。雅砻江河源至河口天然落差约3 830 m,河口多年平均流量约为 $1\ 890\ \text{m}^3/\text{s}$,年径流量约占长江上游的13.3%,流域面积约13.6万 km^2 ,年降水量在河源区约为500~600 mm,中下游区逐渐增加至900~1 300 mm,其中中游地区最高可达1 500~1 800 mm。雅砻江流域受地形影响,气候条件复杂,降水空间分布大致为由南向北递减,由东向西递减,同时受到西南季风的影响,导致降水在5—10月较为集中,雨量约占全年的90%~95%^[16]。

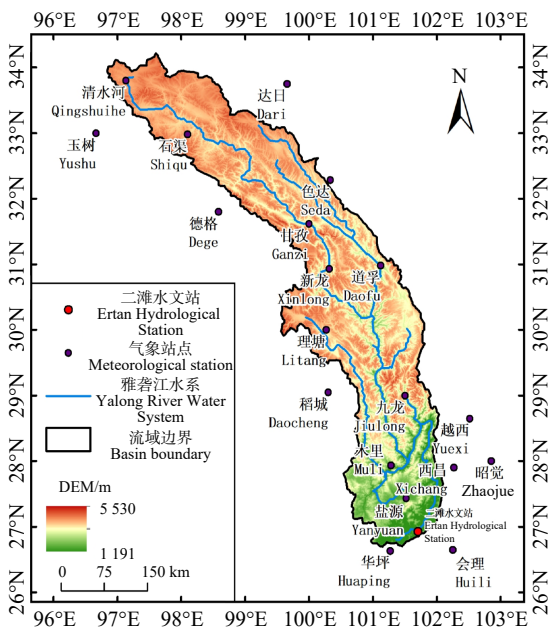


图1 研究区地理位置

Fig. 1 Geographical location of the study area

1.2 研究数据

在对比现有研究成果的基础上,为保证对降水资料稀少或缺失地区水文状况的有效探索,本研究在充分考虑降水时间序列完整性的基础上选取了站点实测降水和3套降水产品数据作为融合降水的基础数据源。

1.2.1 实测降水数据

实测降水数据来自于中国气象数据服务中心发布的《中国地面气候资料日值数据集(V3.0)》(<https://data.cma.cn/>)。该数据集的空间范围为 $75^{\circ}11' \sim 132^{\circ}58'E$, $16^{\circ}32' \sim 53^{\circ}29'N$ 。该数据集在数据质量方面较之前发布的同类产品更为准确和完整。各要素数据的实测值普遍在99%以上,数据正确率接近100%。在本研究中,基于对空间连续性和站点实测降水数据完整性的考虑,最终选择了研究区内及其附近的19个实测站点,时间范围为1981-01-01—2017-12-31。

1.2.2 多源降水产品

(1)全球气候的第五代大气再分析数据集

(ECMWF Reanalysis v5, ERA5)来自于欧洲中期天气预报中心(The European Centre for Medium-Range Weather Forecasts, ECMWF),覆盖了1940年以来的全球气候数据,以30 km的分辨率和137个垂直层次(地表到高空80 km)解析大气,并在降低空间和时间分辨率时提供所有变量的不确定性信息。ERA5结合了模型数据和全球观测数据,取代了ERA-Interim再分析^[17]。本研究采用ERA5-Land逐小时数据,时间范围为1981-01-01—2017-12-31,空间分辨率为 0.1° 。

(2)卫星降水数据(climate hazards center infraRed precipitation with station data, CHIRPS)是由美国地质调查局与加州大学气候灾害中心共同研发的全球降雨数据集(<https://data.chc.ucsb.edu/products/CHIRPSRPS-2.0/>)。时间起始于1981年,范围跨越 $50^{\circ}S \sim 50^{\circ}N$,以及所有经度,CHIRPS将 0.05° 分辨率的卫星图像与原位台站数据相结合,为趋势分析和季节性干旱监测创建网格化降雨时间序列^[18]。本研究采用CHIRPS-2.0逐日数据,时间范围为1981-01-01—2017-12-31,空间分辨率为 0.25° 。

(3)亚洲精细分析数据(the high asia refined analysis, HAR)是德国柏林工业大学在中亚气候和构造自然灾害项目(climatic and tectonic natural hazards in Central Asia)框架内生成的大气数据集(<https://www.klima.tu-berlin.de/>)。该套数据提供了青藏高原及其周边地区温度、降水和风速等要素的网格场,包含30 km和10 km两种精度^[19]。本研究采用HAR V2的逐日数据,时间范围为2000-01-01—2017-12-31,空间分辨率为 0.1° 。

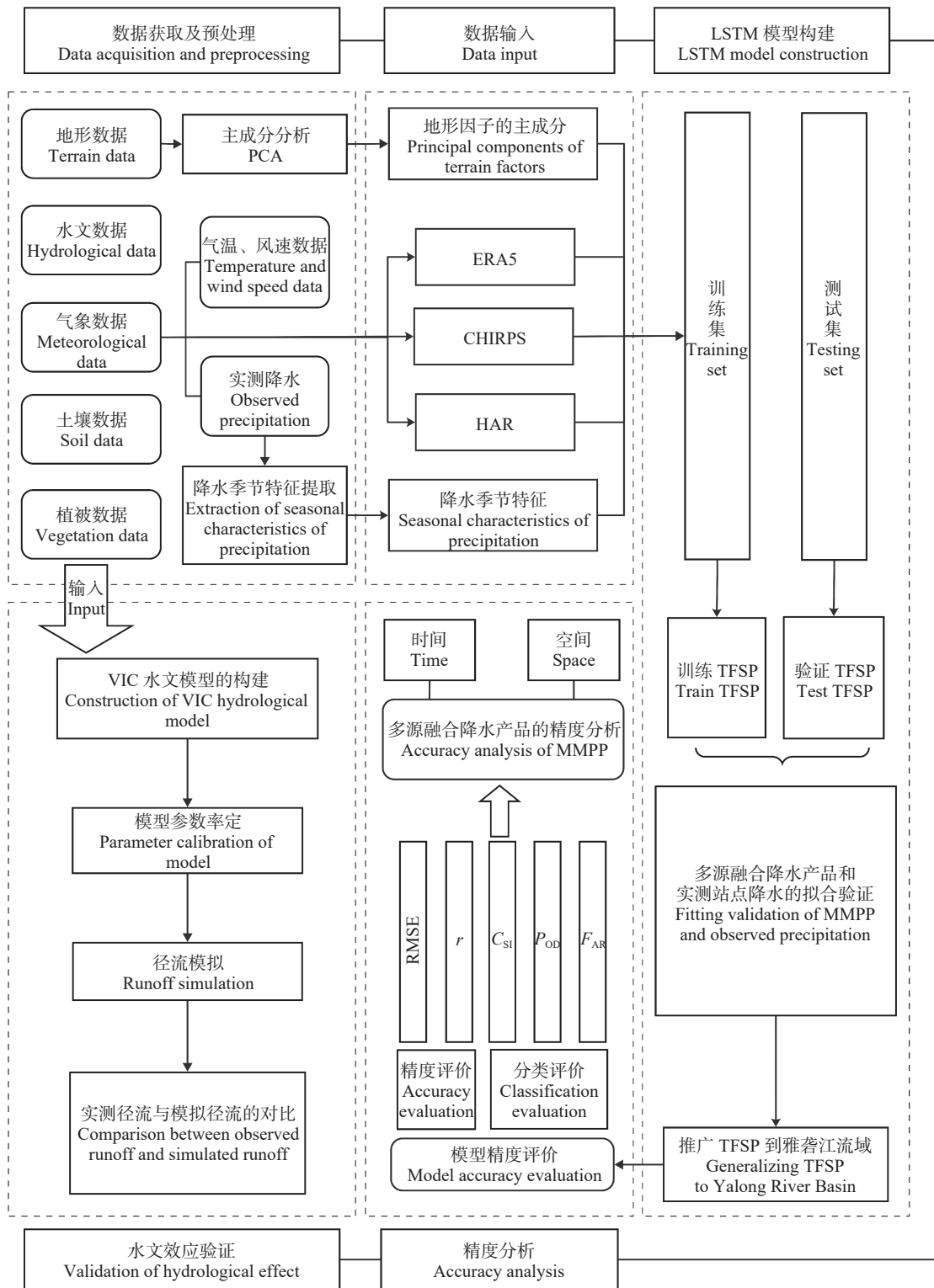
1.2.3 其他数据

逐日流量数据来源于1981—2017年长江流域水文年鉴中的二滩水文站(研究区出口)。植被数据基于美国马里兰大学研发的地表覆盖类型资料,土壤数据采用美国国家海洋和大气管理局(National Oceanic and Atmospheric Administration, NOAA)水文办公室提供的全球5'土壤质地分类资料。地形因子数据基于数字高程模型(digital elevation model, DEM)提取,DEM数据来自美国地质调查局(<https://earthexplorer.usgs.gov/>)。

2 研究方法与评估指标

2.1 研究方法

本研究拟基于地形因子和降水季节特征构建一套多源降水融合机器学习模型(TFSP),以期提高雅砻江流域降水产品的时空精度,具体构建方法如图2所示。



PCA.主成分分析; ERA5.全球气候第五代大气再分析数据集; CHIRPS.卫星降水数据; HAR.亚洲精细分析数据; LSTM.长短期记忆人工神经网络; TFSP.本研究构建的多源降水融合机器学习模型; MMPP.本研究生产的多源融合降水产品; VIC.VIC 模型; C_{SI} .临界成功指数; P_{OD} .检测概率; F_{AR} .误报率。PCA, principal component analysis; ERA5, ECMWF reanalysis v5; CHIRPS, climate hazards center infraRed precipitation with station data; HAR, high Asia refined analysis; LSTM, long short term memory; TFSP, a multi-source merged precipitation machine learning model based on terrain factors and seasonal precipitation characteristics; MMPP, a multi-source merged precipitation product; VIC, the variable infiltration capacity model; C_{SI} , critical success index; P_{OD} , detection probability; F_{AR} , false positive rate.

图 2 多源融合降水机器学习模型的构建流程

Fig. 2 Construction flowchart of multi-source precipitation merge machine learning model

2.1.1 机器学习模型

长短期记忆递归神经网络 (LSTM)是由

Hochreiter 和 Schmidhuber 于 1997 年提出的,增加了一个专门用于保存历史信息的记忆单元,解决了

循环神经网络(recurrent neural network, RNN)出现的梯度爆炸和梯度弥散的问题^[20-21]。历史信息通过遗忘门、输入门和输出门的控制进行更新,主要解决了数据分类问题,应用于自然语言翻译、图像字幕和语音识别等多个方面,也可用于时间序列的预测。

2.1.2 水文模型

VIC(variable infiltration capacity)模型是由华盛顿大学、加州大学伯克利分校和普林斯顿大学的研究学者基于 SVATS(soil vegetation atmospheric transfer schemes)的原理共同研发出的大尺度陆面分布式水文模型,目前应用较广泛的是具有3层土壤层的 VIC-3L 模型^[22-24]。本研究采用 Rosenbrock 算法^[25]对 VIC 水文模型产流模块的参数进行率定优化,具体参数如表1。该算法采用的是 Oth 阶段搜索,每一个搜索阶段均包含一组方向的探索性搜索。

2.2 评价指标

本研究从数值精度和分类两方面对 TFSP 及模型输出结果进行评价。降水产品的数值精度评价采用均方根误差(root mean squared error, RMSE)、Pearson 相关系数(Pearson correlation coefficient, r)、相对误差(relative error, RE)^[26-28]以及标准差(standard deviation, SD)进行评价;降水产品的分类评估能力采用临界成功指数(critical success index, C_{SI})、检测概率(probability of detection, P_{OD})和误报率(false alarm ratio, F_{AR})3个指标进行评价。 C_{SI} 表示成功预测一次降水事件的概率,取值范围为0~1,越接近于1表示成功概率越高; P_{OD} 表示对每次降水事件反演命中率,取值范围为0~1,取值为1表示命中率百分百,为0表示反演完全错误; F_{AR} 表示卫星降水产品对真实降水事件预报出现错误的概率,取值范围为0~1,取值越大,表示出现错误的概率越高。

$$C_{SI} = \frac{N_{11}}{N_{11} + N_{10} + N_{01}} \quad (1)$$

$$P_{OD} = \frac{N_{11}}{N_{11} + N_{01}} \quad (2)$$

$$F_{AR} = \frac{N_{10}}{N_{11} + N_{10}} \quad (3)$$

式中: N_{11} 表示降水产品与基准数据均探测有降水的频数; N_{01} 表示降水产品探测无降水而基准数据有降水的频数; N_{10} 表示降水产品探测有降水而基准数据无降水的频数。判断有无降水的阈值可自行设定,一般选定为0.1 mm。

采用纳什效率系数(Nash-Sutcliffe efficiency coefficient, N_{ash})、相对误差(RE)和确定系数(coefficient of determination, C)来评估径流模拟的准确性^[29], N_{ash} 的值介于 $(-\infty, 1]$,越接近1表示模型的拟合效果越好,即模型预测结果与实际观测数据的拟合程度较高。

$$N_{ash} = 1 - \frac{\sum_{i=1}^n (Q_{sim}^i - Q_{obs}^i)^2}{\sum_{i=1}^n (Q_{obs}^i - \bar{Q}_{obs})^2} \quad (4)$$

$$C = \left[\frac{\sum_{i=1}^n (Q_{sim}^i - \bar{Q}_{sim})(Q_{obs}^i - \bar{Q}_{obs})}{\sqrt{\sum_{i=1}^n (Q_{sim}^i - \bar{Q}_{sim})^2 \sum_{i=1}^n (Q_{obs}^i - \bar{Q}_{obs})^2}} \right]^2 \quad (5)$$

式中: n 代表时间序列的长度, Q_{obs}^i 是第*i*个时刻的观测流量, Q_{sim}^i 第*i*个时刻的模拟流量, \bar{Q}_{obs} 和 \bar{Q}_{sim} 分别是研究时段内观测流量和模拟流量的均值。

3 引入地形因子和降水季节特征的多源降水融合机器学习模型的构建

青藏高原地区下垫面状况复杂,加上气候多变,使得区域内降水量在时间和空间上呈现出非均匀分布,因此本研究在构建模型时引入地形因子和降水的季节特征以提高现有降水产品的精度。主要构建步骤:降水数据预处理、地形因子与降水季节特征的提取及预处理、无标签网格样本处理和 TFSP 模型构建。

3.1 降水数据预处理

基于实测站点的空间分布,选择了研究区及其周边的19个气象站点(图1)。对研究区实测站点资

表1 VIC 水文模型所需率定参数

Tab. 1 Calibration parameters for VIC hydrological model

参数符号 Parameter symbol	参数含义 Parameter meaning	上限 Upper limit	下限 Lower limit
b	可变渗透曲线参数 Variable permeability curve parameter	0	10
D_m	最大基流流速 Maximum baseflow velocity/($m^3 \cdot s^{-1}$)	0	30
D_s	基流值与最大基流值的比值 Ratio of baseflow to maximum baseflow	0	1
W_s	土壤湿度与最大湿度的比值 Ratio of soil moisture to maximum soil moisture	0	1
D_1	第1层土壤厚度 Soil thickness of layer 1/m	0	0.5
D_2	第2层土壤厚度 Soil thickness of layer 2/m	0	2
D_3	第3层土壤厚度 Soil thickness of layer 3/m	0	2

料进行严格的审查, 以确保其可靠性、一致性和代表性。利用这 19 个气象站点数据, 提取 CHIRPS、ERA5 和 HAR 降水产品在相应时段的逐日数据。为了规避数据过度波动, 增强模型构建的稳定性, 对 19 个气象站点的逐日降水数据以及提取的多源降水产品逐日降水数据进行标准化, 使其处于 [0, 1]。标准化处理有助于减少数据之间的量纲差异和异常值的影响, 从而使模型更加稳定和可靠。由于逐日降水数据中存在大量 0 mm 数据, 对训练集和验证集的尾部分布存在影响。因此, 为保证训练集和测试集的特征一致性, 对测试集进行归一化处理时采用相应训练集的特征值。

$$X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \quad (6)$$

式中: X_{norm} 是归一化后的降水值, X 是任一系列的降水值, X_{min} 和 X_{max} 分别是该系列中的最小值和最大值。

3.2 地形因子及降水季节特征的提取与预处理

本研究根据 Yin 等^[30]方法, 结合雅砻江流域的实际地形, 选取 22 个地形因子表征其对降水的影响。通过对雅砻江流域 19 个气象站点的统计, 得出 12.5 km 是提取该流域地形因子的适宜半径。因此, 以每个气象站点为圆心, 构建半径为 12.5 km 的缓冲区, 提取研究区范围内 DEM 数据中的各类地形因子, 如图 3 所示。由于各类地形因子并非相互独立且存在数据冗余, 采用主成分分析进行降维处理,

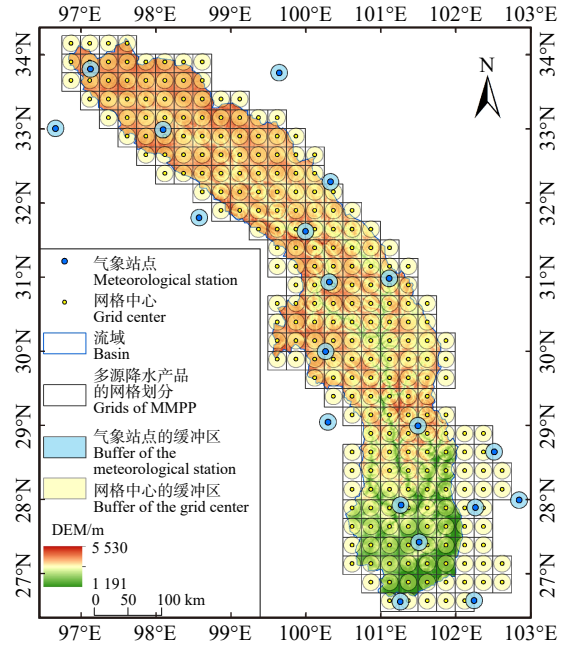


图 3 地形因子提取示意图

Fig. 3 Schematic diagram of terrain factor extraction

在保留主要信息的基础上提取有效的地形因子, 作为后续多源降水融合模型的输入特征。

雅砻江流域受西南季风的影响, 降水量呈现季节性变化, 夏秋季节降水相对较多, 冬春季节降水相对较少, 因此本研究拟引入季节特征进行多源降水融合。

3.3 无标签网格样本的处理

基于研究区集水面积和降水产品空间分辨率, 将雅砻江流域划分为 250 个网格(图 4)。若网格样

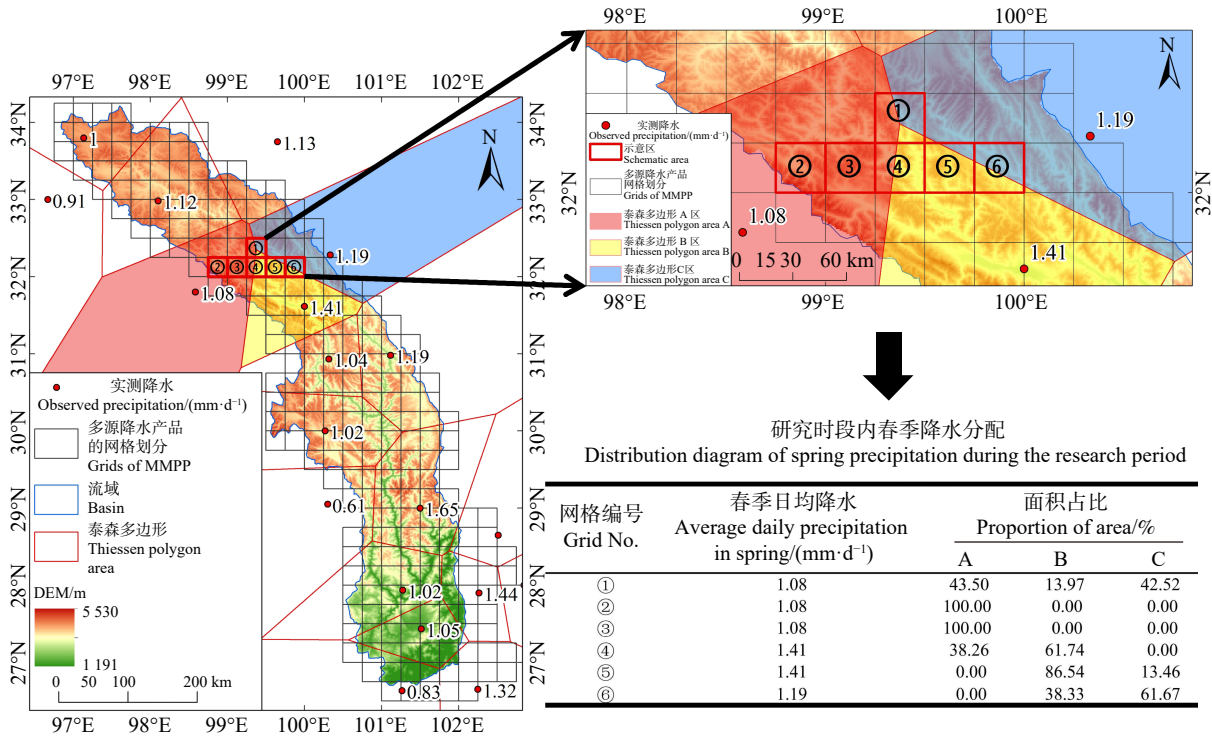


图 4 雅砻江流域的空间划分

Fig. 4 Spatial division of Yalong River Basin

本内有气象站点则认为该网格是具有标签的网格样本,若网格内没有气象站点则认为该网格不具有标签。有标签网格样本,按3.1和3.2小节完成预处理。对每个无标签网格内的3套降水产品数据按3.1小节进行降水数据的预处理,再按3.2小节提取无实测站点网格内的地形因子并进行主成分分析,最后对无标签网格样本赋予降水季节特征。主要操作如下:基于实测站点绘制泰森多边形,计算每个子泰森多边形区域内的多年日均降水量,再将各子泰森多边形内的降水季节特征分配给其包含的无标签网格。各网格的降水季节特征取决于由各气象站划分出的泰森多边形的子区域在该网格中的面积占比,进而赋值为子区域内气象站实测降水的季节特征。

3.4 构建 TFSP 模型

对实测站点和降水产品进行预处理后,设置 LSTM 中的相关参数。基于 Python 3.9 中的 Keras 库构建一个包含 20 个节点的单层结构,采用自适应学习率的 Adam 随机梯度下降优化算法,作为损失函数训练批量大小为 10 的模型。参考 Lu 等^[31]研究,设置输入序列长度为 60,初始学习率为 0.001。

基于初设的 LSTM、地形因子和降水季节特征构建 TFSP 模型。由于各降水产品时段不同,因此融合过程分为 1981—1999 年和 2000—2017 年两个时段。以 1981—1999 年为例,将 1981—1999 年时间序列分为 2 段,1981—1990 年数据为训练集,1991—1999 年为测试集。LSTM 模型输入包括 3 类降水产品(ERA5、CHIRPS 和 HAR),4 种地形因子主成分和处理后的降水季节特征。基于已构建并训练完成的 TFSP 模型,得到 1981—1999 年无标签网格样本的逐日数据(图 4)。类比上述操作,最终得到 1981—2017 年雅砻江流域 $0.25^{\circ} \times 0.25^{\circ}$ 的多源融合降水产品(multi-source merged precipitation product, MMPP)。

4 结果与分析

4.1 多源降水产品和多源融合降水产品的时序精度评估

表 2 为多源降水产品 ERA5(1981—2017 年)、CHIRPS(1981—2017 年)和 HAR(2000—2017 年)与实测日降水数据的精度评估结果。总体而言,3 类降水

产品的精度都呈现出较差的状态。相关系数在 0.36~0.53。CHIRPS 与实测降水的相对误差较小(5.63%),表现出轻微高估;HAR 表现较为明显的低估(-21.05%);ERA5 和实测降水的相对误差最大(54.36%)。在分类评估方面,HAR 最优,其 C_{SI} 、 P_{OD} 、 F_{AR} 分别为 0.52、0.88、0.44;ERA5 次之, C_{SI} 、 P_{OD} 、 F_{AR} 分别为 0.48、0.95、0.51;CHIRPS 表现最差, C_{SI} 、 P_{OD} 、 F_{AR} 分别为 0.35、0.45、0.38。

表 2 多源降水产品日尺度精度评估结果

Tab. 2 Evaluation results of daily scale accuracy of multi-source precipitation products

降水产品 Precipitation product	r	RE/%	RMSE/(mm·d ⁻¹)	SD/(mm·d ⁻¹)
ERA5	0.53	54.36	5.47	5.23
CHIRPS	0.39	5.63	7.37	6.53
HAR	0.36	-21.05	5.52	3.38

表 3 为基于实测降水序列对 MMPP 进行精度验证的结果。1981—2017 年,MMPP 与融合前各类降水产品相比,整体精度大幅提高。相关系数由融合前均值 0.43 提高到 0.82,均方根误差从融合前均值 6.12 mm/d 下降到 5.45 mm/d。此外,MMPP 与实测降水的相对误差为 7.75%,标准差 7.98 mm/d,表明融合后降水分布在实测降水附近且相较于实测降水呈现出正偏差,整体表现为略微高估。MMPP 平均 C_{SI} 、 P_{OD} 和 F_{AR} 分别为 0.64、0.92 和 0.33。上述结果表明,多源融合降水产品与实测降水的拟合效果较好。分时段研究,1981—1999 年,多源降水产品包括 ERA5 和 CHIRPS,2000—2017 年增加了 HAR。结果表明,2000—2017 年较 1981—1999 年的相应指标低,结合表 2 的 HAR 与实际降水相关性最低,推断 HAR 加入融合模型并不能提高模型的精度。

对 MMPP 分季节进行精度评价(表 4)。各季节降水产品与实测降水数据的一致性均较好,相关系数均值高达 0.84,均方根误差和标准差的均值分别为 2.33 mm/d 和 5.15 mm/d,以春秋两季最优,夏冬两季次之。其中冬季表现最差,相关系数为 0.81,相对误差为-4.89%。

4.2 多源融合降水产品的空间精度分析

图 5 可以看出多源融合降水较为准确地捕捉到

表 3 不同时段多源融合降水产品的日尺度精度评估结果

Tab. 3 Evaluation results of daily scale accuracy of multi-source merged precipitation products in different periods

时段 Period	输入源 Input source	r	RE/%	RMSE/(mm·d ⁻¹)	SD/(mm·d ⁻¹)
1981—1999	ERA5、CHIRPS	0.83	4.15	3.33	8.64
2000—2017	ERA5、CHIRPS、HAR	0.72	36.87	3.89	15.32
1981—2017	ERA5、CHIRPS、HAR	0.82	7.75	5.45	7.98

表 4 MMPP 的四季精度评估结果

Tab. 4 Four-season accuracy assessment results of MMPP

季节 Season	r	RE/%	RMSE/(mm·d ⁻¹)	SD/(mm·d ⁻¹)
春 Spring	0.85	1.25	1.97	4.11
夏 Summer	0.83	7.69	4.76	9.89
秋 Autumn	0.86	4.51	1.85	5.46
冬 Winter	0.81	-4.89	0.74	1.12

了区域实测降水的特征分布,反映了降水的空间纹理特性。融合后的降水呈现出由东南向西北内陆递减的趋势,最高值位于流域下游出口处,多年日均值为 3.24 mm/d; 最小值在高海拔地区,多年日均值为 1.10 mm/d。相较于 3 类降水产品,MMPP 使流域内绝大部分区域的降水数据更加平滑。

MMPP 季节降水数据与实测值在空间分布格局上较为相近(图 6)。春季,两类降水数据的最大值均出现在流域东南部,其中 MMPP 的最大降水为 2.19 mm/d; 夏、秋季,两类降水最大的区域均出现在流域下游并呈现由东南向西北递减的趋势,夏季 MMPP 降水量达到最大,为 8.22 mm/d; 冬季,两类降水均呈现流域东南部最大,西北部次之,中部最小的分布格局。冬季东南部降水量最大可能是受地形因素的影响,出现迎风坡的“地形雨”现象,西北部降水量次之可能是受冬季降雪的影响。多源融合降水产品的降水量在各季节的变幅与实测降水较为接近,在一定程度上反映了多源融合降水产品能够更好地拟合出实际降水的变化情况。

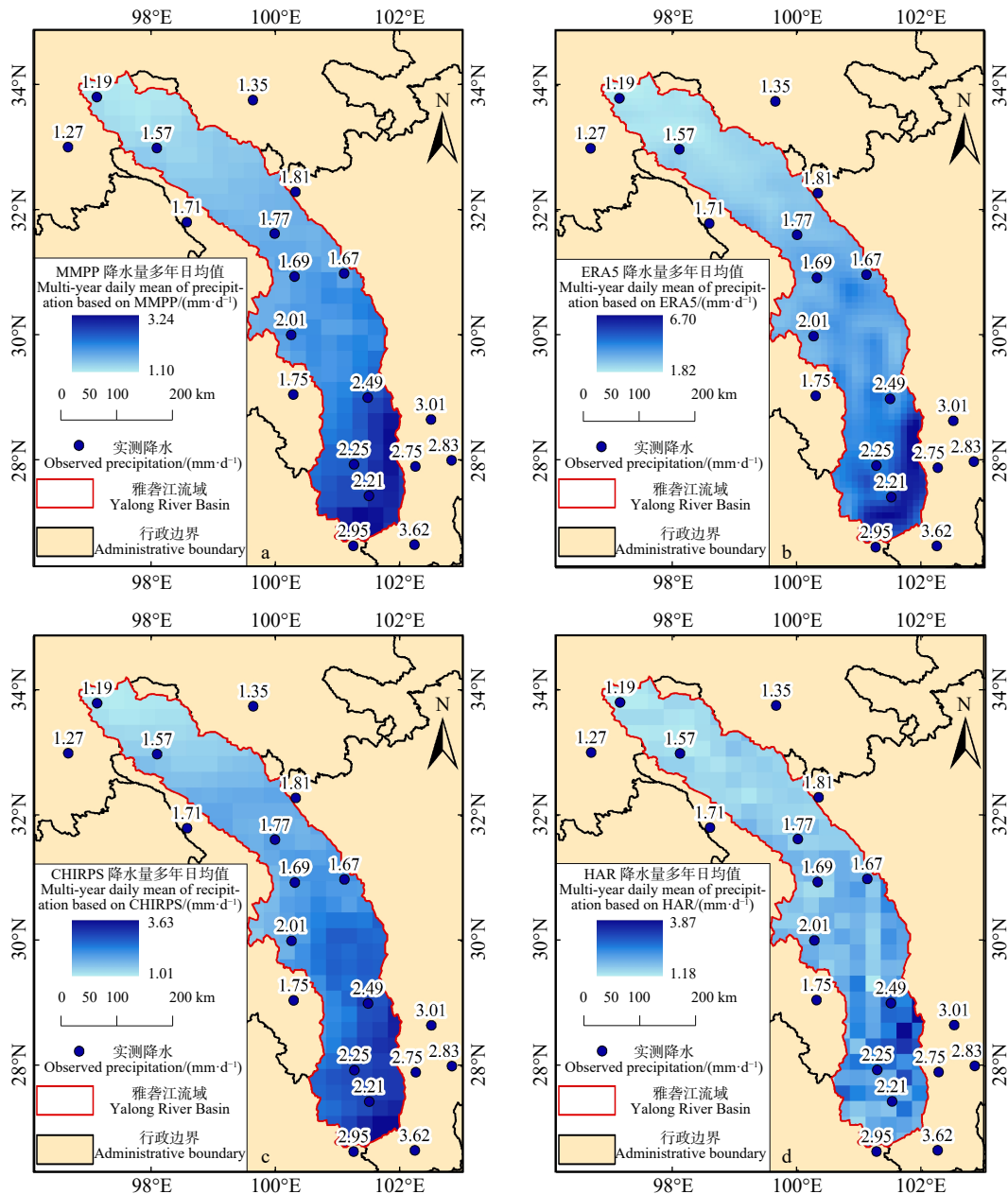


图 5 1981—2017 年各降水产品降水量多年日均值空间分布

Fig. 5 Spatial distribution of multi-year daily mean precipitation based on various precipitation products from 1981 to 2017

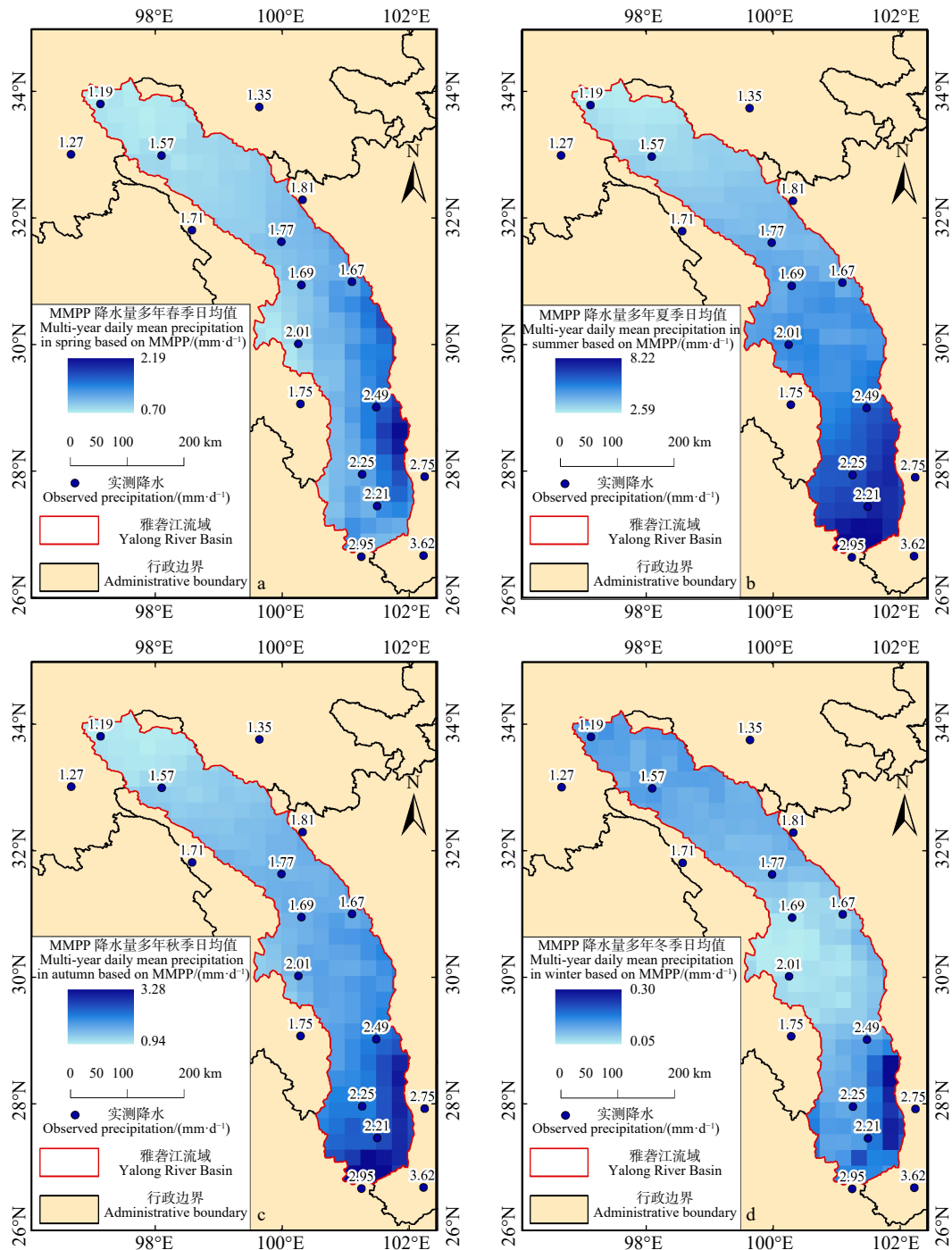


图6 1981—2017年MMPP和实测季节降水量日均值空间分布

Fig. 6 Spatial distribution of multi-year (1981—2017) daily mean precipitation in four seasons of measured precipitation and MMPP

4.3 水文效应评估

基于最终输出的MMPP在雅砻江流域二滩水文站进行水文效应验证,其中1981—1999年为率定期,2000—2017年为验证期(表5)。本研究利用融合后的降水进行径流模拟,并与二滩水文站的实测径流进行拟合(图7)。1981—2017年,实测径流的多年日均流量为 $1\ 603.51\ \text{m}^3/\text{s}$,模拟径流的多年日均流量为 $1\ 625.72\ \text{m}^3/\text{s}$,两者仅相差 $22.21\ \text{m}^3/\text{s}$,相对误差为1.39%。实测径流在春夏秋冬四季的多年日均流量分别为 640.97 、 $2\ 935.49$ 、 $2\ 246.82$ 和

表5 二滩水文站逐日径流拟合结果

Tab. 5 Daily runoff fitting results of Ertan Hydrological Station

研究时段 Research period	N_{ash}	RE/%	C
率定时段 Calibration period (1981—1999)	0.83	1.69	0.91
验证时段 Validation period (2000—2017)	0.76	0.99	0.88
整体时段 Whole period (1981—2017)	0.80	1.39	0.89

注: N_{ash} , 纳什效率系数; C, 确定系数。Notes: N_{ash} , Nash-Sutcliffe efficiency coefficient; C, coefficient of determination.

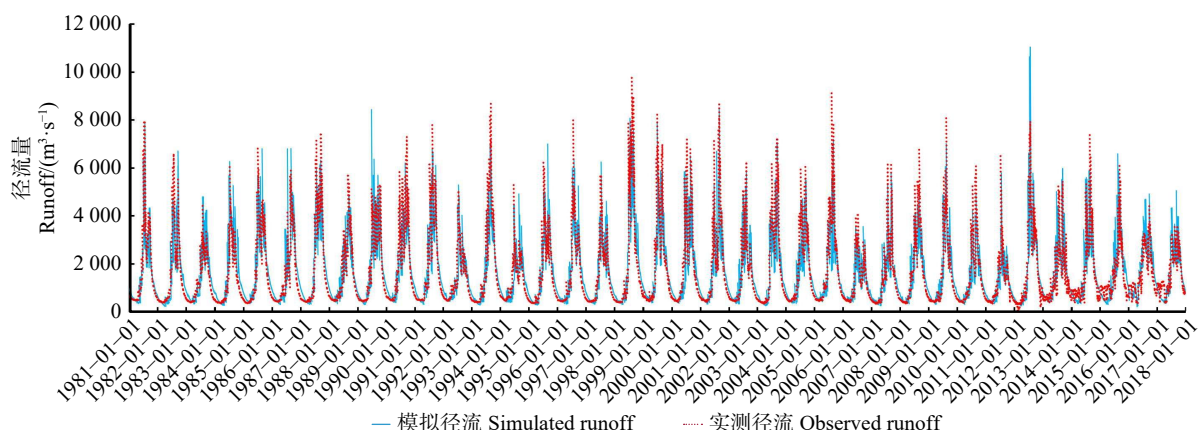


图 7 二滩水文站逐日径流拟合过程线

Fig. 7 Daily runoff fitting process line of Ertan Hydrological Station

576.15 m³/s, 模拟径流相应季节的多年日均流量分别为 595.19、2 793.44、2 362.29 和 740.83 m³/s。其中, 冬季的相对误差最大, 为 28.58%。

基于 MMPP 所得的逐日径流数据能够较为准确地模拟出研究区的实测流量过程线, 但在 1989 年和 2012 年出现了较为严重的高估状况。模拟径流对于枯水期整体呈现出轻微低估的现象, 尤其是 2015 年初。从评估指标来看, 验证期的精度相较于率定期有所下降, 但整体呈现出轻微高估但相似性较高的状况, 整个研究时段内的 N_{ash} 、RE 和 C 分别为 0.80、1.39% 和 0.89。

5 结论与讨论

本研究采用 19 个实测站点数据和 3 类多源降水产品, 考虑地形和季节因素, 利用 LSTM 方法构建了 TFSP 模型, 生成了 1981—2017 年雅砻江流域 $0.25^{\circ} \times 0.25^{\circ}$ 分辨率的逐日降水数据集。

雅砻江流域实测降水由东南向西北内陆递减, 降水产品空间分布与实测大体一致, 但 HAR 产品出现不连续性, ERA5 产品降水量偏高。四季尺度分析表明, 各产品对降水中心的捕捉能力以及在空间连续性特征上表现不一, 尤其在冬季表现较差。

融合模型有效整合了实测和降水产品数据, 显著提升了降水数据的相关系数 (r 从 0.43 提高到 0.82) 和降低了均方根误差 (RMSE 从 6.12 mm/d 降至 5.45 mm/d), 改善了空间分布误差。水文模型验证表明, MMPP 产品有效解决了水文资料缺乏问题。

研究还探讨了融合降水数据的多源性, 发现增加数据源并不能绝对提高模型精度, 输入数据的原始精度对融合结果影响较大, 单纯增加数据源并不能保证模型精度提升。

参 考 文 献

[1] Zhong R D, He Y H, Chen X H. Responses of the hydrological

regime to variations in meteorological factors under climate change of the Tibetan Plateau[J]. Atmospheric Research, 2018, 214: 296–310.

[2] Zhang X X, Jiang L G, Liu Z F, et al. Flow regime changes in the Lancang River, revealed by integrated modeling with multiple Earth observation datasets[J/OL]. Science of the Total Environment, 2023, 862: 170–186.

[3] 张婷, 李怡, 李建柱, 等. 多源降雨数据融合及其水文应用研究进展[J]. 自然灾害学报, 2022, 31(1): 15–28.

Zhang T, Li Y, Li J Z, et al. Research progress on multi-source rainfall data fusion and their hydrological application[J]. Journal of Natural Disasters, 2022, 31(1): 15–28.

[4] 南天一, 陈杰, 丁智威, 等. 基于深度学习的青藏高原多源降水融合[J]. 中国科学: 地球科学, 2023, 53(4): 836–855.

Nan T Y, Chen J, Ding Z W, et al. Deep learning-based multi-source precipitation merging for the Tibetan Plateau[J]. Science China Earth Sciences, 2023, 53(4): 836–855.

[5] Zhang L, Li X, Zheng D H, et al. Merging multiple satellite-based precipitation products and gauge observations using a novel double machine learning approach[J]. Journal of Hydrology, 2021, 594: 587–604.

[6] 潘咏, 谷军霞, 徐宾, 等. 多源降水数据融合研究及应用进展[J]. 气象科技进展, 2018, 8(1): 143–152.

Pan Y, Gu J X, Xu B, et al. Advances in multi-source precipitation merging research[J]. Advances in Meteorological Science and Technology, 2018, 8(1): 143–152.

[7] Anagnostou E N, Maggioni V, Nikolopoulos E I, et al. Benchmarking high-resolution global satellite rainfall products to radar and rain-gauge rainfall estimates[J]. IEEE Transactions on Geoscience and Remote Sensing, 2009, 48(4): 1667–1683.

[8] Andrew V, Balaji R, William K, et al. A Bayesian kriging approach for blending satellite and ground precipitation observations[J]. Water Resources Research, 2015, 51(2): 908–921.

[9] 张敬国. 基于深度学习的数值天气预报降水产品订正方法研究[D]. 南京: 南京信息工程大学, 2023.

Zhang J G. Research on precipitation product correction method of numerical weather forecast based on deep learning[D]. Nanjing: Nanjing University of Information Science &

- Technology, 2023.
- [10] 张永顺. 基于机器学习的数值天气预报降水产品偏差订正方法研究 [D]. 长沙: 国防科技大学, 2020.
- Zhang Y S. Research on the method of correcting precipitation product deviation in numerical weather prediction based on machine learning[D]. Changsha: National University of Defense Technology, 2020.
- [11] 杨轩, 曾燕, 邱新法, 等. 基于机器学习算法的多源月尺度融合降水产品在中国区域的检验评估 [J]. 暴雨灾害, 2023, 42(5): 595–605.
- Yang X, Zeng Y, Qiu X F, et al. Examination and evaluation of multi-source monthly scale fusion precipitation product in China based on machine learning algorithm[J]. *Torrential Rain and Disasters*, 2023, 42(5): 595–605.
- [12] 张钧民, 阮惠华, 许剑辉, 等. 基于 XGBoost 的多源降水数据融合方法研究 [J]. 热带地理, 2021, 41(4): 845–856.
- Zhang J M, Ruan H H, Xu J H, et al. An XGBoost-merging method for high-resolution daily precipitation estimation for a regional rainstorm event[J]. *Tropical Geography*, 2021, 41(4): 845–856.
- [13] Nguyen G V, Le X H, Van L N, et al. Application of random forest algorithm for merging multiple satellite precipitation products across South Korea[J]. *Remote Sensing*, 2021, 13(20): 3172–3189.
- [14] Zhang Z, Wang D G, Qiu J X, et al. Machine learning approaches for improving Near-Real-Time IMERG rainfall estimates by integrating cloud properties from NOAA CDR PATMOS-x[J]. *Journal of Hydrometeorology*, 2021, 22(10): 2767–2781.
- [15] 董立俊. 雅砻江流域径流对未来气候变化的时空响应研究 [D]. 宜昌: 三峡大学, 2020.
- Dong L J. Spatiotemporal response of runoff to climate change in Yalongjiang River Basin[D]. Yichang: China Three Gorges University, 2020.
- [16] 于家瑞, 艾萍, 袁定波, 等. 雅砻江流域径流演变特征及成因 [J]. 水电能源科学, 2019, 37(10): 13–16.
- Yu J R, Ai P, Yuan D B, et al. Characteristics of evolution of runoff and its causes in Yalongjiang River Basin[J]. *Water Resources and Power*, 2019, 37(10): 13–16.
- [17] Zhang X, Lu N, Jiang H, et al. Evaluation of reanalysis surface incident solar radiation data in China[J/OL]. *Scientific Reports*, 2020, 10(1): 3494[2024-04-30]. <https://www.nature.com/articles/s41598-020-60460-1>.
- [18] Funk C, Peterson P, Landsfeld M, et al. The climate hazards infrared precipitation with stations: a new environmental record for monitoring extremes[J]. *Scientific data*, 2015, 2(1): 1–21.
- [19] Orsolini Y, Wegmann M, Dutra E, et al. Evaluation of snow depth and snow cover over the Tibetan Plateau in global reanalyses using in situ and satellite remote sensing observations[J]. *The Cryosphere*, 2019, 13(8): 2221–2239.
- [20] 李嘉源. 基于层叠长期记忆模型的日径流预报研究 [D]. 南京: 南京信息工程大学, 2023.
- Li J Y. Research on daily runoff forecasting based on cascade long short-term memory model[D]. Nanjing: Nanjing University of Information Science & Technology, 2023.
- [21] 殷仕明. 基于 LSTM 网络的流域水文模型建模与迁移重构研究 [D]. 重庆: 重庆交通大学, 2023.
- Yin S M. Research on watershed hydrological model modeling and migration reconstruction based on LSTM network[D]. Chongqing: Chongqing Jiaotong University, 2023.
- [22] Liang X, Lettenmaier D P, Wood E F, et al. A simple hydrologically based model of land surface water and energy fluxes for general circulation models[J]. *Journal of Geophysical Research: Atmospheres*, 1994, 99(D7): 14415–14428.
- [23] Xie Z H, Su F G, Liang X, et al. Applications of a surface runoff model with Horton and Dunne runoff for VIC[J]. *Advances in Atmospheric Sciences*, 2003, 20(2): 165–172.
- [24] 刘谦. VIC 大尺度陆面水文模型在中国区域的应用 [D]. 长沙: 湖南大学, 2004.
- Liu Q. An application of variable infiltration capacity (VIC) macro scale land surface hydrological model in China[D]. Changsha: Hunan University, 2004.
- [25] Rosenbrock H. An automatic method for finding the greatest or least value of a function[J]. *The Computer Journal*, 1960, 3(3): 175–184.
- [26] Brier G W. Verification of forecasts expressed in terms of probability[J]. *Monthly Weather Review*, 1950, 78(1): 1–3.
- [27] Zhu S Y, Ma Z Q, Xu J T, et al. A morphology-based adaptively spatio-temporal merging algorithm for optimally combining multisource gridded precipitation products with various resolutions[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1–21.
- [28] Tang G Q, Clark M P, Papalexiou S M, et al. Have satellite precipitation products improved over last two decades? A comprehensive comparison of GPM IMERG with nine satellite and reanalysis datasets[J]. *Remote Sensing of Environment*, 2020, 240: 473–490.
- [29] Nilawar A P, Waikar M L. Use of SWAT to determine the effects of climate and land use changes on streamflow and sediment concentration in the Purna River Basin, India[J]. *Environmental Earth Sciences*, 2018, 77(23): 89–101.
- [30] Yin Z Y, Zhang X Q, Liu X D, et al. An assessment of the biases of satellite rainfall estimates over the Tibetan Plateau and correction methods based on topographic analysis[J]. *Journal of Hydrometeorology*, 2008, 9(3): 301–326.
- [31] Lu D, Konapala G, Painter S L, et al. Streamflow simulation in data-scarce basins using Bayesian and physics-informed machine learning models[J]. *Journal of Hydrometeorology*, 2021, 22(6): 1421–1438.

(责任编辑 孟瑶
责任编辑 孟平)