DOI:10.20040/j.cnki.1000－7709.2026.20250617

基于GWR与CatBoost算法的沁河流域汛期GPM产品降尺度校正

吴晨晨[1],任 黎[1],金 哲[2],赵 杰[1],李小迈[1],熊良冯[1]

(1. 河海大学水文水资源学院,江苏 南京 210098;2. 吉林省水文水资源局延边分局,吉林 延吉 133001)

摘要:卫星降水产品因空间范围大、连续分布等优势而广泛应用于水文研究,但空间分辨率低、复杂地形精度低等不足阻碍了该类产品在具有显著时空异质性研究区的直接应用。对此,采用多种插值方法,基于GWR和CatBoost算法建立了沁河流域GPM汛期降水产品降尺度校正模型,并分析了校正前后的空间分布细节与数据精度。结果表明,GPM月尺度降水数据在沁河流域具有较好的适用性,但空间分布细节和数据精度上仍有提高的空间;在多种组合降尺度校正模型中,CatBoost_linear模型效果最好,不仅有效改善了降水数据的空间分布细节,而且与站点实测数据的相关性达到最优。该研究可得到一套沁河流域高质量网格降水数据,为进一步研究该地区的降水特征、构建水文模型等工作提供了数据支持。

关键词:GPM;降尺度;GWR;CatBoost;RBF

中图分类号:TV125;P407;P412 **文献标志码**:A **文章编号**:1000-7709(2026)02-0042-05

Downscaling and Correction of GPM Products During Flood Season in the Qinhe River Basin Using GWR and CatBoost Algorithms

WU Chen-chen[1], REN Li[1], JIN Zhe[2], ZHAO Jie[1], LI Xiao-mai[1], XIONG Liang-feng[1]

(1. College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China;
2. Yanbian Branch, Jilin Provincial Hydrology and Water Resources Bureau, Yanji 133001, China)

Abstract: Satellite precipitation products have been widely used in hydrological research due to their large spatial range and continuous distribution, but the lack of low spatial resolution and low accuracy of complex terrain hinders the direct application of such products in research areas with significant spatiotemporal heterogeneity. In view of these shortcomings, this study uses a variety of interpolation methods to establish a downscaling correction model of GPM precipitation products in the Qinhe River Basin based on GWR and CatBoost algorithms, and analyzes the precipitation distribution details and data accuracy before and after downscaling. The results show that the GPM monthly precipitation data has good applicability in the Qinhe River Basin, but there is still room for improvement in data quality. The results of each downscaling improved the spatial details of precipitation, and the CatBoost_linear model had the best effect, and the correlation with the measured data at the stations were the best. This study can obtain a set of high-quality grid precipitation data in the Qinhe River Basin, which provides data support for further research on the precipitation characteristics and hydrological modeling in this region.

Key words: GPM; downscaling; GWR; CatBoost; RBF

1 概况

沁河位于北纬34°54′～36°59′、东经111°56′～113°49′之间,地势自北向南逐渐降低,年降水量在600～720 mm之间,流域面积13 532 km^2,全长485.1 km。流域概况见图1。作为黄河一级支流,沁河是黄河中游关键治理河段,其水文过程及防洪安全对黄河中游综合治理有重要影响。目前,该流域降水监测主要依赖雨量站实测,数据精

收稿日期:2025-04-07,**修回日期**:2025-05-18,**接受日期**:2025-05-22

基金项目:国家重点研发计划(2023YFC3209302)

作者简介:吴晨晨(2000-),女,硕士研究生,研究方向为水资源规划与管理,E-mail:wuchenchen2000@163.com

通讯作者:任黎(1978-),女,教授,研究方向为水资源系统优化规划与管理,E-mail:renli@hhu.edu.cn

中国知网 https://www.cnki.net

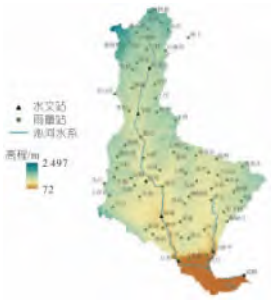


图 1　沁河流域地理概况图

Fig. 1　Geographical overview of the Qinhe River Basin

度高,但存在分辨率不足和数据不连续等局限。相比之下卫星反演降水数据凭借其大范围覆盖和良好的时空连续性,为降水监测提供了有效补充。近年来,众多卫星降水产品已普遍应用于相关研究。其中全球卫星降水计划(GPM)作为新一代卫星降水观测系统,空间覆盖度和时空分辨率进一步优化且对微量降水和固态降水的观测精度显著提升[1]。受传感器精度、反演算法及采样技术制约,GPM 产品在实际应用中仍存在显著误差,需通过降尺度校正优化其适用性。当前降尺度方法已由传统统计模型(如线性回归、地理加权回归 GWR)[2]逐步演进至机器学习方法(如随机森林、梯度提升)[3],但 CatBoost 等新型算法应用仍显不足;数据校正所用到的插值方法局限于 Kriging、IDW 等经典插值法,而对径向基函数(RBF)等更具灵活性的方法探索较少[4]。鉴此,本文基于 GWR 和 CatBoost 算法,结合多源环境变量,构建沁河流域 2001～2021 年汛期(6～9 月)GPM 数据的降尺度模型,并采用 IDW 和 RBF 插值法融合站点实测数据进行偏差校正,以期获得更高时空分辨率的流域降水数据,为沁河流域水资源管理、洪水预报及生态保护提供更可靠的数据支持,同时也为卫星降水产品的区域化应用提供方法借鉴。

2　数据来源与研究方法

2.1　数据来源

高程数据 DEM 来自地理空间数据云(https://www.gscloud.cn),归一化植被指数 NDVI 和地表温度 LST 数据来自美国太空总署(NASA)(https://modis.gsfc.nasa.gov)发布的 MODIS13A3 和 MOD11C3 数据集。

卫星降水数据选用 GPM IMERGE 终期多卫星产品 Final Run V07B 月尺度数据,时间范围为 2001～2021 年汛期(6～9 月)。

雨量站数据摘录自《中华人民共和国水文年鉴》第 4 卷第 6 册黄河下游区(伊洛河、沁河水系)年鉴,本文选择沁河流域内 67 个雨量站的月降水量。

2.2　研究方法

2.2.1　GWR

GWR 作为一种局部加权回归空间分析技术突破了全局模型的限制[5]。模型公式为:

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

式中,y_i 为点 i 处降尺度后的降水量;$\beta_0(u_i,v_i)$ 为 i 点处的常数项回归参数;(u_i,v_i) 为 i 点的坐标;$\beta_j(u_i,v_i)$ 为第 j 个影响因子在第 i 个点的回归参数;n 为影响因子个数;x_{ij} 为 i 点处第 j 个影响因子的变量值;$\varepsilon(u_i,v_i)$ 为第 i 个点的残差。

回归参数方程:

$$\hat{\beta}(u_i,v_i)=(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{W}(u_i,v_i)\boldsymbol{x})^{-1}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{W}(u_i,v_i)\boldsymbol{y} \tag{2}$$

式中,$\hat{\beta}(u_i,v_i)$ 为点 (u_i,v_i) 处的系数;$\boldsymbol{x}$、$\boldsymbol{y}$ 分别为自变量、因变量的向量;$\boldsymbol{W}(u_i,v_i)$ 为权重矩阵,距离 i 越近的点权重越大。

2.2.2　CatBoost

CatBoost 是 GBDT 框架下的一种改进算法,简化了数据预处理流程,能更好地处理数据特征中 categorical features。计算公式为:

$$\hat{x}_k^i=\frac{\sum_{j=1}^{p-1}[x_{\sigma_{j,k}}=x_{\sigma_{p,k}}]Y_{\sigma_j}+ap}{\sum_{j=1}^{p-1}[x_{\sigma_{j,k}}=x_{\sigma_{p,k}}]+a} \tag{3}$$

式中,$\hat{x}_k^i$ 为第 i 个样本的第 k 个类别特征转换后的数值;$x_{\sigma_{j,k}}$ 为排列中第 j 个样本的第 k 个类别特征的值;$x_{\sigma_{p,k}}$ 为当前要编码的样本的第 k 类别特征的值;Y_{σ_j} 为排列中第 j 个样本的标签/目标值(0/1)或连续值;p 为添加的先验项;a 为大于 0 的权重系数。

2.2.3　IDW

IDW 作为插值方法多应用于将离散的已知位置样点数据转化为连续分布数据。计算公式为:

$$w_i=1/d_i^p \tag{4}$$

式中,w_i 为第 i 个样点的权重;d_i 为样点与待插值点的距离;p 为距离指数,常取 2 或 3。

$$z(x_i,y_i)=\sum_{i=1}^{N}w_iz_i\Big/\sum_{i=1}^{N}w_i \tag{5}$$

式中,$z(x_i,y_i)$ 为 (x_i,y_i) 点处的估算数值;N 为已知样点的个数;(x_i,y_i) 为第 i 个样点的坐标;z_i 为函数值。

中国知网　https://www.cnki.net

2.2.4 RBF

RBF 插值法不要求采样点在空间中具有结构化的网格形式，通过对采样点构建径向基函数表达式得到插值点数值。计算公式为：

$$\overline{P}(x)=\sum_{i=1}^{n}W_i\varphi(\|x_i-x_j\|) \tag{6}$$

式中，$\overline{P}(x)$为在 x 点处的估计值；W_i 为第 i 个基函数的权重；φ 为径向基函数，常见的有 Gaussian、Multiquadrics、Linear 等。

2.2.5 精度评价指标

选取决定系数 R^2、均方根误差 R_{RMSE} 和平均绝对误差 M_{MAE}3 个指标进行评价。

2.2.6 研究流程

具体流程为：①数据处理。基于 DEM 数据提取地形因子，在 Arcgis 10.8 软件将地形因子、NDVI、LST 数据裁剪并分别重采样至 0.1°、0.01°。②构建降尺度模型。将 GPM 数据作为目标变量，0.1°地形因子、NDVI、LST 数据为环境变量，分别利用 GWR、CatBoost 算法建立回归关系，即降尺度模型。③运行模型。将 0.01°环境变量输入降尺度模型运算生成 0.01°降水数据。④校正。计算 GPM 与站点值残差，分别用 IDW、RBF 两种方法插值到 0.01°，与运行模型所得的 0.01°数据叠加，得到最终的 0.01°高精度网格数据。⑤精度评估。选取 R^2、R_{RMSE}、M_{MAE} 对各降尺度校正方法所得结果进行精度评估。

3 研究结果与分析

3.1 GPM 降水产品适用性分析

目前 GPM 产品在沁河流域缺乏针对性应用，因此进行降尺度校正前需评估适用性。以 2001～2021 年汛期(6～9 月)雨量站观测月降水数据为 X 轴，以相应的 GPM 数据为 Y 轴绘制精度评估散点图(图 2)，并计算精度评估指标 R^2=0.698、R_{RMSE}=43.389 mm、M_{MAE}=28.885 mm，3 个评价指标均表现良好，表明 GPM 产品与站点实测数据之间有一定的相关性和精度，因此可以认为 GPM 产品在沁河流域的研究与应用具有一定的可行性和适用性，但数据质量仍存在提升空间，有待开展更为深入和系统的研究，以期为该流域提供更加精确可靠的降水数据支持。

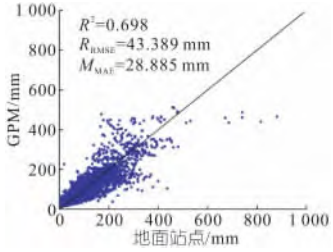


图 2 汛期 GPM 数据精度评估

Fig. 2 Accuracy assessment of GPM data during flood season

同样地，分别绘制各月精度评估散点图，见图 3。由图 3 可看出，GPM 在 9 月的拟合效果最好，R^2=0.894，在汛期 4 个月份中达到最高，其次为 7、8 月，R^2 分别为 0.661、0.519，而在 6 月拟合效果最差，R^2 仅为 0.353。在数据精度方面，9 月 R_{RMSE}、M_{MAE} 值为 4 个月份中最小，分别为 25.489、18.725 mm，数据精度最高，而 7 月数据精度最差，R_{RMSE}、M_{MAE} 分别为 58.982、41.192 mm，在 4 个月份中最高。

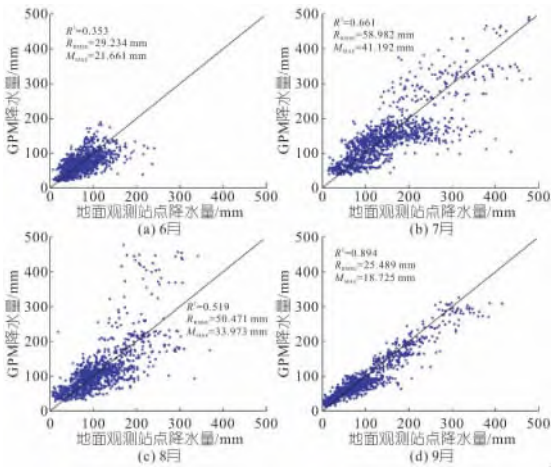


图 3 汛期 GPM 各月份降水数据精度评估

Fig. 3 Accuracy assessment of GPM precipitation data during flood season by month

3.2 沁河流域多年平均降水量降尺度结果

分别采用 GWR 和 CatBoost 方法将 GPM 数据空间分辨率降尺度到 0.01°，在初步完成降尺度处理之后，利用 IDW 和 RBF(4 种插值函数)将 GPM 降水数据和站点观测数据的残差插值到整个流域，并与初步降尺度结果叠加得到 10 组经过降尺度及校正的降水数据集。原始 GPM 数据、站点实测值以及各降尺度结果的多年平均降水分布见图 4。

由图 4(a)可看出，沁河流域汛期多年平均降水量整体上呈现西北部向东南部递增的趋势，低值区出现在中上游，降水量最低为 398 mm，高值区出现在中下游，最高达 474 mm。与图 4(b)雨量站数据相比，二者对流域降雨量空间分布的呈现存在一定差异且原始网格数据有明显锯齿状，分辨率和精度有改善空间。由图 4(c)～(l)可看出，校正结果对降水量空间分布的描述更加接近实测数据。沁河流域汛期多年平均降水量高值区

中国知网 https://www.cnki.net

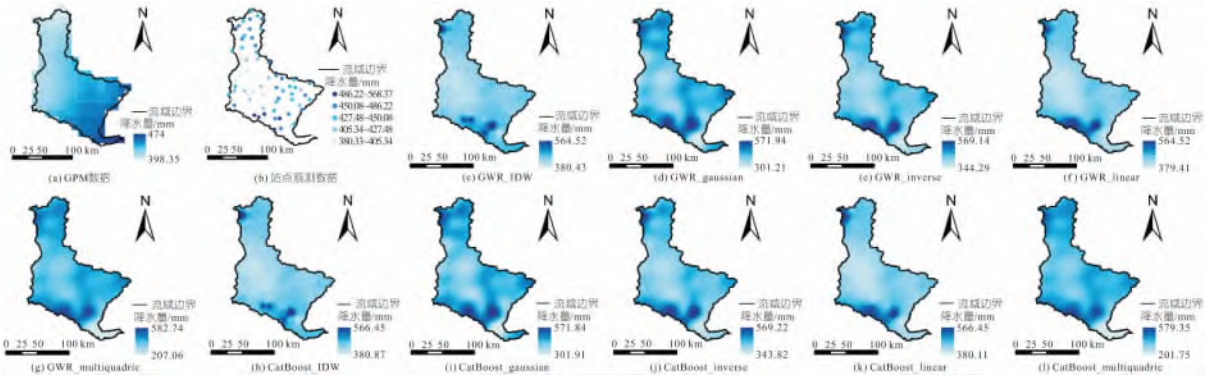


图 4　降水数据降尺度结果

Fig. 4　Downscaling results of precipitation data

位于流域上游和中下游西南地区,降水量可达 500～700 mm,低值区位于中游和下游东南地区,降水量在 201～380 mm 之间。降水量时空分布整体上与站点数据更为相似,对降水量分布的细节刻画能力更强。各降尺度校正模型所得数据与原始 GPM 数据相比均能更好地刻画降水量空间分布的细节,与站点数据空间分布的一致性也有不同程度的提升。

3.3　降尺度模型精度对比分析

为定量评价各降尺度校正模型效果,计算 10 组结果数据的 R^2、R_{RMSE}、M_{MAE} 值并绘制雷达图,见图 5。由图 5 可知,相较于原始 GPM 数据,各降尺度校正结果数据的 R^2 均有不同程度提高,而 R_{RMSE} 和 M_{MAE} 明显降低,表明降尺度校正后的数据与实测数据有更好的一致性和精度。为进一步对比分析出最适用于沁河流域 GPM 降水数据降尺度及校正的方法,绘制每组降尺度校正结果数据的精度评估散点图进行精度评估对比分析。

基于 GWR 的结果数据精度评估散点图见图 6。由图 6 可看出,在相关性方面,各模型的 R^2 在 0.930～0.962 之间,与原始 GPM 数据 R^2 = 0.698 相比均有一定程度的提升,其中 GWR_linear 表现最优,R^2 达到 0.962。在观测误差方面,各模型 R_{RMSE} 在 15.223～20.947 mm 之间,与原始 GPM 数据适用性评估中 R_{RMSE} = 43.389 mm 相比均有明显降低,其中 GWR_linear 最优,R_{RMSE} 仅为 15.223 mm,GWR_multiquadric 表现最差,R_{RMSE} 高达 20.947 mm;各模型 M_{MAE} 在 8.997～12.527 mm 之间,与原始 GPM 的 M_{MAE} = 28.885 mm 相比也明显降低,其中 GWR_linear 表现最优,M_{MAE} = 8.997 mm,GWR_gaussian 表现最差,R_{RMSE} = 12.527 mm。综合 3 个指标,在 GWR 的各结果中 GWR_linear 最优,分辨率提升至 0.01°的同时与站点值呈现最优相关性及最小误差。

基于 CatBoost 的降尺度校正结果数据精度评估散点图见图 7。由图 7 可看出,在相关性方

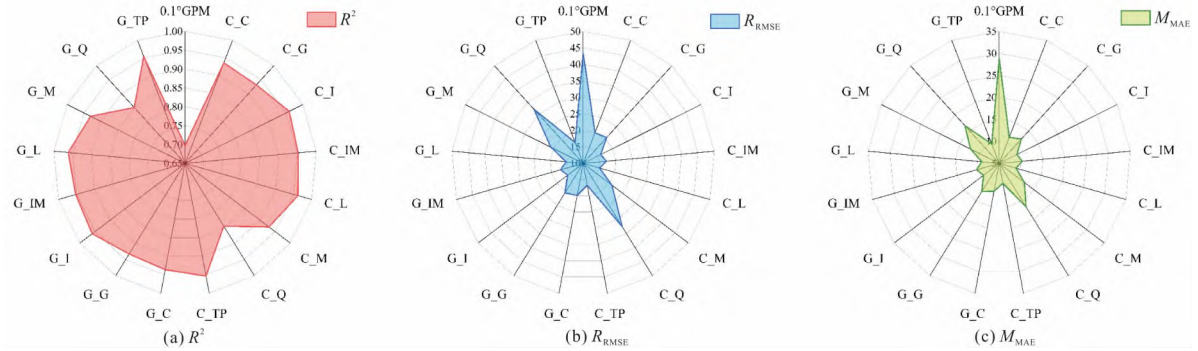


图 5　精度指标对比雷达图

Fig. 5　Radar chart of accuracy metrics comparison

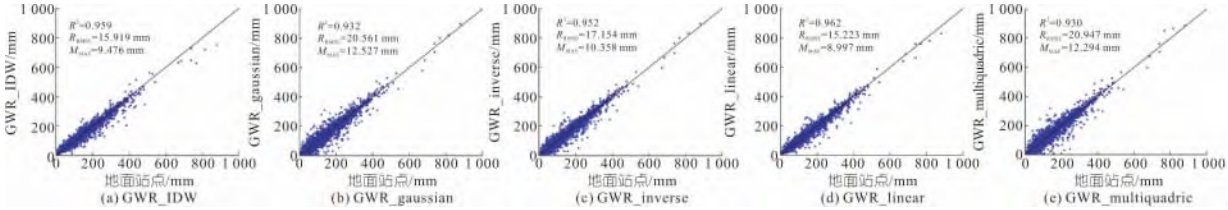


图 6　基于 GWR 的降尺度校正精度评估散点图

Fig. 6　Scatter plot of downscaling correction accuracy evaluation based on GWR

中国知网　https://www.cnki.net

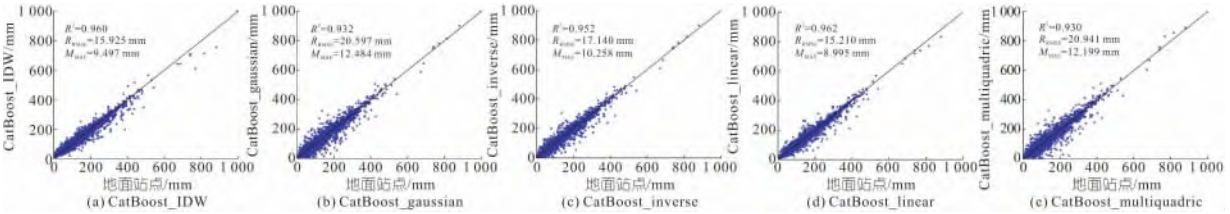


图 7 基于 CatBoost 的降尺度校正精度评估散点图

Fig. 7 Scatter plot of downscaling correction accuracy assessment based on CatBoost

面，各模型的 R^2 在 0.930～0.962 之间，与原始 GPM 数据的 $R^2=0.698$ 相比，均有一定程度的提升，其中 CatBoost_linear 表现最优，R^2 达到 0.962。在观测误差方面，各模型的 R_{RMSE} 在 15.210～20.941 mm 之间，与原始 GPM 数据适用性评估中 $R_{RMSE}=43.389$ mm 相比均有明显的降低，其中 CatBoost_linear 最优，R_{RMSE} 仅为 15.210 mm，CatBoost_multiquadric 表现最差，R_{RMSE} 高达 20.941 mm；各模型的 M_{MAE} 在 8.995～12.484 mm 之间，与原始 GPM 的 $M_{MAE}=28.885$ mm 相比也有明显降低，其中 CatBoost_linear 表现最优，$M_{MAE}=8.995$ mm，CatBoost_gaussian 表现最差，$R_{RMSE}=12.484$ mm。综合 3 个评价指标，在 CatBoost 模型的各个降尺度及校正结果中，与 RBF 插值中的 Linear 函数相结合的 CatBoost_linear 结果最优。

综上所述，基于 GWR 的不同校正方法中 GWR_linear 模型效果最优，基于 CatBoost 算法的不同校正方法中 CatBoost_linear 模型效果最优，对比可知，在拟合程度方面，2 个模型 R^2 均为 0.962，在数据精度方面二者有细微差别，CatBoost_linear 模型的 R_{RMSE}、M_{MAE} 分别比 GWR_linear 模型低 0.013、0.002 mm。因此，CatBoost_linear 模型在沁河流域 2001～2021 年汛期的 GPM 降水产品的降尺度校正中适用性最好。

4 结论

a. GPM 月尺度降水数据在沁河流域汛期具有一定的适用性，$R^2=0.698$、$R_{RMSE}=43.389$ mm、$M_{MAE}=28.885$ mm，数据质量有提升空间。

b. 降尺度校正后数据降水量空间分布刻画能力和精度显著提升。沁河流域汛期多年平均降水量高值区位于上游和中下游西南地区，降水量可达 500～700 mm，低值区位于中游和下游东南地区。

c. CatBoost_linear 适用性最好，R^2 达 0.962，R_{RMSE}、M_{MAE} 分别为 15.210、8.995 mm。

参考文献：

[1] 徐秀丽，张艺铭，张俊瑞，等. 基于星地多源数据与 GA-DNN 模型的江苏省 GPM 产品降尺度研究[J]. 水电能源科学，2021，39(10)：18-21，62.(XU X L，ZHANG Y M，ZHANG J R，et al. Downscaling of GPM products in Jiangsu Province based on satellite ground multi-source data and GA-DNN model[J]. Water resources and power，2021，39(10)：18-21，62.(in Chinese))

[2] 王浩钰，付国永，李黎，等. 一种基于 Xgboost+GWR 算法的中国地区降水量空间插值模型[J]. 水电能源科学，2021，39(12)：7-11.(WANG H Y，FU G Y，LI L，et al. A spatial precipitation distribution prediction model for Chinese regions based on Xgboost+GWR algorithm[J]. Water resources and power，2021，39(12)：7-11.(in Chinese))

[3] 苗聪聪，闫金凤. 永兴岛 GPM-IMERG 降水数据空间降尺度研究[J]. 水电能源科学，2024，42(12)：24-27，45.(MIAO C C，YAN J F. Spatial downscaling study of Yongxing Island GPM-IMERG data[J]. Water resources and power，2024，42(12)：24-27，45.(in Chinese))

[4] 齐化龙. 基于 ATPRK 算法的 2019 年华中地区 GPM 降水数据降尺度研究[J]. 水电能源科学，2021，39(12)：12-16.(QI H L. Downscaling study of GPM precipitation data in Central China in 2019 by applying ATPRK algorithm[J]. Water resources and power，2021，39(12)：12-16.(in Chinese))

[5] BRUNSDON C，FOTHERINGHAM S，CHARLTON M. Geographically weighted regression—Modelling spatial non-stationarity[J]. Journal of the royal statistical society series D：the statistician，1998，47(3)：431-443.

中国知网 https://www.cnki.net