



基于XGBoost模型的广西区域PWV估算及极端降雨分析

闫立伟, 林买金, 谢劭峰, 黄良珂, 李向红, 方琼玉

引用本文:

闫立伟, 林买金, 谢劭峰, 等. 基于XGBoost模型的广西区域PWV估算及极端降雨分析[J]. 大地测量与地球动力学, 2026, 46(6): 702–709.

Yan Liwei, Lin Maijin, Xie Shaofeng, et al. Estimation of PWV and Analysis of Extreme Rainfall in Guangxi Region Based on XGBoost Model[J]. Journal of Geodesy and Geodynamics, 2026, 46(6): 702–709.

在线阅读 View online: <https://doi.org/10.14075/j.jgg.2025.08.295>

您可能感兴趣的其他文章

Articles you may be interested in

基于PSO-BP神经网络的中国区域大气加权平均温度模型

An Atmospheric Weighted Mean Temperature Model of China Region Based on PSO-BP Neural Network

大地测量与地球动力学. 2023, 43(12): 1300–1306 <https://doi.org/10.14075/j.jgg.2023.12.016>

基于地基GNSS观测数据的贵州高原地区水汽层析精度分析

Accuracy Analysis of Water Vapor Tomography Based on Ground-Based GNSS Observation Data in Guizhou Plateau

大地测量与地球动力学. 2023, 43(11): 1162–1166 <https://doi.org/10.14075/j.jgg.2023.11.011>

海潮模型对陆态网络GNSS观测的影响

Influence of Ocean Tide Model on GNSS Observations of CMONOC

大地测量与地球动力学. 2023, 43(11): 1193–1198 <https://doi.org/10.14075/j.jgg.2023.11.017>

基于FFT和小波变换的台风天气GNSS ZTD周期特征分析

Typhoon Weather GNSS ZTD Cycle Characteristics Analysis Based on FFT and Wavelet Transform

大地测量与地球动力学. 2023, 43(10): 1032–1038 <https://doi.org/10.14075/j.jgg.2023.10.008>

一种基于Emardson-H的GPT3改进模型构建方法

An Improved GPT3 Model Established by Emardson-H Method

大地测量与地球动力学. 2023, 43(10): 1039–1044 <https://doi.org/10.14075/j.jgg.2023.10.009>

基于 XGBoost 模型的 广西区域 PWV 估算及极端降雨分析

闫立伟¹ 林买金² 谢劭峰¹ 黄良珂¹ 李向红³ 方琼玉³

¹ 桂林理工大学测绘地理信息学院, 桂林, 541006

² 长安大学地质工程与测绘学院, 西安, 710054

³ 桂林市气象局, 桂林, 541001

摘要: 针对气象参数缺失情况下实时/近实时 GNSS PWV 反演受限的问题, 基于 XGBoost 模型分别建立广西区域 3 种无需实测气象参数的 PWV 估算模型。首先建立输入为站点时间(DOY 和 HOD)、位置(Longitude, Latitude 和 Height)和 GNSS ZTD, 输出特征为 GNSS PWV 的 XGBZ-PWV 模型, 然后在 XGBZ-PWV 模型的基础上分别加入 2 种经验 PWV, 建立 XGBZG-PWV 模型和 XGBZE-PWV 模型。同时, 基于 GNSS ZTD, 利用 GPT3 模型提供的气压与温度反演 PWV(GPT3-PWV 模型), 对所建 PWV 模型进行对比, 并以 2022 年广西区域基于 GNSS ZTD、ERA5 地表气压和温度反演的 GNSS PWV 作为参考值对所建立的模型进行精度验证。结果表明, 相比于 GPT3-PWV 模型, XGBZ-PWV、XGBZG-PWV 和 XGBZE-PWV 模型的估算精度分别提升 22.98%、29.03% 和 31.45%, 其中, XGBZE-PWV 模型最优。在 2022 年 2 次极端降雨过程中, 对 PWV 与降雨的时空演变特征进行分析, 结果表明, XGBZE-PWV 模型在极端天气条件下依然具有较好的适用性。

关键词: ZTD; PWV; GNSS; XGBoost; 降雨

中图分类号: P228

文献标识码: A

大气可降水量(PWV)是指垂直大气柱中从地球表面到指定对流层高度的总水蒸气含量, 在气候研究及极端天气监测中具有重要意义^[1-3]。地基 GNSS PWV 反演是通过 GNSS 天顶对流层延迟(ZTD)结合地表气压和温度来实现的, GNSS ZTD 可通过精密单点定位(PPP)技术获取。由于大部分 GNSS 站并未配备气象传感器, 获取实时或近实时高精度气象参数困难, 限制了 GNSS PWV 的实时/近实时反演。

因此, 有学者研究并建立无需实测气象参数的 GNSS PWV 反演模型。李黎等^[4]、韦云等^[5]基于线性模型建立无需气象参数的区域 ZTD 直接获取 PWV, 但此类模型不能精确描述 ZTD 和 PWV 之间的非线性关系, 模型精度较低。随着人工智能算法的发展, 深度学习和机器学习被广泛应用于各种回归和分类任务。此类算法具有较强学习数据特征的能力, 能够较好地学习数据之间的非线性特征。因此, Zheng 等^[6]、Zhang

等^[7]和 Huang 等^[8]先后使用深度学习算法建立中国、全球区域的 ZTD, 直接转换 PWV 模型。由于数据驱动模型的估算性能依赖于输入特征对输出特征变化规律的刻画能力和学习性, 这些模型仅使用 ZTD 建模难以精确估计 PWV。另有研究表明, PWV 不仅是云和降水形成的必要条件, 还在强降水、寒潮、台风及严重洪涝灾害等极端天气过程中起主要作用^[9-11]。Xiong 等^[12]通过分析广西“龙舟水”期间的 GNSS PWV 及其驱动因素, 发现 GNSS PWV 受风速、温度及平均海平面气压的共同影响。江婷等^[13]研究表明, 在充足的水汽条件下降雨的形成依赖于强烈的水汽辐合和上升作用, 结合热动力条件, 可以有效预测降水时间和强度。PWV 值的变化与降水事件的发生存在显著的相关性, 尤其是当 PWV 值达到一定阈值时, 降雨发生的概率明显增加^[14-16]。

针对以上模型的局限性, 首先建立广西区域仅需输入站点时间(DOY 和 HOD)、位置(Longi-

收稿日期: 2025-08-24

项目来源: 广西自然科学基金(2023GXNSFAA026434)。

第一作者简介: 闫立伟, 主要研究方向为 GNSS 大气水汽反演, E-mail: 13347633621@163.com。

通讯作者简介: 谢劭峰, 教授, 主要研究方向为 GNSS 大气水汽监测, E-mail: xieshaofeng@glut.edu.cn。

tude、Latitude 和 Height) 和 GNSS ZTD 的 XGBZ-PWV 模型。然后引入 2 种经验 PWV (GAskne_PWV 和 GXE_PWV) 作为先验信息, 其中 GAskne_PWV 基于 GPT3 和 Askne-Nordius 模型获取, GXE_PWV 基于经验 ERA5 格网数据获取; 将经验 PWV 值融入模型, 分别建立 XGBZG-PWV 和 XGBZE-PWV 模型, 并以 2022 年广西区域基于 GNSS ZTD、ERA5 地表气压和温度反演的 GNSS PWV 作为参考值, 系统评估模型的估算精度。最后, 基于广西 2022 年 2 次极端降雨过程, 利用 GNSS PWV、模型估算的 PWV 和气象站的降雨数据, 分析 PWV 与极端降雨之间的时空变化特征, 并探讨所建立模型在极端气象条件下的适用性。

1 实验数据与方法

1.1 实验数据

所使用的数据主要包括 2020-01-01—2022-12-31 的广西地区地基 GNSS 站数据、气象站数据(图 1)和 $0.25^\circ \times 0.25^\circ$ 的逐小时 ERA5 地表气压和温度数据。基于武汉大学 PRIDE 团队开发的 PRIDE PPP-AR 软件对 GNSS 观测数据进行处理得到小时 ZTD 数据。基于 ERA5 提供的地表气压和温度, 使用反距离加权和垂直插值模型获取 GNSS 站点高度的小时气象参数进行 GNSS PWV 反演^[17]。气象站根据“最近原则”向 GNSS 站提供的逐小时降雨数据^[18]。

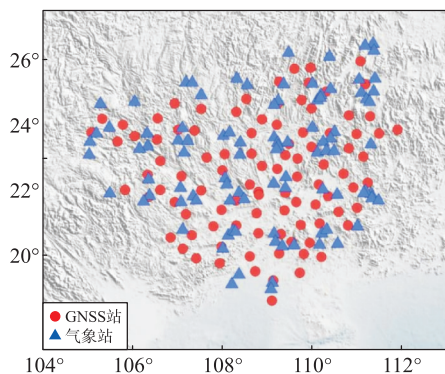


图 1 广西区域 GNSS 站和气象站分布
Fig.1 Distribution of GNSS stations and meteorological stations in Guangxi

1.2 GNSS PWV 反演原理

GNSS ZTD 由天顶静力学延迟(ZHD)和天顶湿延迟(ZWD)组成, 见式(1)。其中 ZHD 可由 Saastamoinen 模型结合地面压力数据计算获取, 见式(2)。

$$\text{ZTD} = \text{ZHD} + \text{ZWD} \quad (1)$$

$$\text{ZHD} = (2.2779 \pm 0.0024) P_s /$$

$$(1 - 0.00266 \cdot \cos 2\varphi - 0.00028 \cdot h) \quad (2)$$

式中, h 为站点椭球高; φ 为站点纬度; P_s 为气压; ZTD、ZHD 和 ZWD 单位均为 mm。最后, ZWD 结合转换系数 Π 即可得到 PWV, 具体公式见式(3)和(4)。

$$\text{PWV} = \Pi \cdot \text{ZWD} \quad (3)$$

$$\Pi = \frac{10^6}{\rho_w R_v [k_3/T_m + k'_2]} \quad (4)$$

式中, Π 为水汽转换系数; ρ_w 为液态水密度; R_v 为水汽比气体常数, $R_v = 461.5 \text{ J}/(\text{kg} \cdot \text{K})$; k_3 和 k'_2 均为大气折射率常数, 其中, $k_3 = 375463 \text{ K}^2/\text{hPa}$, $k'_2 = 22.97 \text{ K/hPa}$; T_m 为加权平均温度, 可由 Bevis 模型计算获得, 此处的 ZWD 单位为 mm。由于广西地区气象站分布较为稀疏, 因而在 GNSS PWV 反演中所需的气压与温度数据均采用 ERA5 再分析资料获取, 以保证 GNSS PWV 的精度^[17]。

1.3 XGBoost 算法原理

XGBoost 是一种基于梯度提升的算法, 通过迭代构建决策树使损失函数最小化。它主要通过集成多个弱学习器形成一个强学习器, 提高模型的性能。XGBoost 不仅具有较高预测精度, 计算效率也较为出色, 因而被广泛应用于各种回归任务中。XGBoost 模型的目标函数为:

$$\begin{cases} L = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \\ \Omega(f_k) = rT + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \end{cases} \quad (5)$$

其中, L 为目标函数; l 为损失函数; y_i 、 $\hat{y}_i$ 分别为样本 i 的实际值和模型估算值; n 为样本数量; Ω 为正则化项; K 为决策树数量; (f_k) 为第 K 棵树的复杂度; r 为正则化参数; λ 为惩罚系数; T 为树的叶节点数; ω_j^2 为叶节点 j 的权重。设置的 XGBoost 模型树的数量为 20, 最大深度为 7, 学习率为 0.1, 其余参数均使用默认值。

2 模型建立与精度验证

2.1 模型建立

1) 首先基于 XGBoost 模型建立一个无需气象参数的广西区域 PWV 估算模型 (XGBZ-PWV), 模型的输入变量为站点时间信息 (DOY 和 HOD)、位置信息 (Longitude、Latitude 和 Height) 和 GNSS ZTD, 输出变量为 GNSS PWV。

2) 基于 Askne-Nordius 模型, 联合式(3)和

式(4)得到式(6):

$$\text{PWV}=\frac{R_d e}{\rho_w R_v(\lambda+1) g_m}$$

(6)

式中, e 为地表水汽压; λ 为水汽压递减因子; g_m 为重力系数; $R_d=287.058\text{ J}/(\text{kg}\cdot\text{K})$ 为干空气气体常数;其余参数与 § 1.2 相同。GPT3 模型可提供 $1^\circ\times 1^\circ$ 的 λ 和 e 。

将式(6)得到的 PWV 命名为 GAskne_PWV,将其作为 PWV 先验信息并作为 XGBZ-PWV 模型的输入,建立 XGBZG-PWV 模型。

3)通过对 ERA5 分层数据进行数值积分可获得地表 PWV 值,根据式(7)建立一个 $0.25^\circ\times 0.25^\circ$ 的广西区域地表经验 PWV 模型(GXE_PWV):

$$\begin{aligned}\text{PWV}^i &= A_0^i + A_1^i \cos\left(2\pi \frac{\text{DOY}}{365.25}\right) + \\ &A_2^i \sin\left(2\pi \frac{\text{DOY}}{365.25}\right) + A_3^i \cos\left(4\pi \frac{\text{DOY}}{365.25}\right) + \\ &A_4^i \sin\left(4\pi \frac{\text{DOY}}{365.25}\right)\end{aligned}$$

(7)

式中, i 为格网点数量; PWV^i 为第 i 个格网点的 PWV 值; A_0^i 为第 i 个格网的年均值; A_1^i 、 A_2^i 为第 i 个格网点的年周期振幅系数; A_3^i 、 A_4^i 为第 i 个格网点的半年周期振幅系数;DOY 为年积日。

利用林买金等^[19]建立的 C-GPWV 模型可将地表经验 PWV 值改正到 GNSS 站点高度。基于 XGBZ-PWV 模型,将 GXE_PWV 模型得到的经验 PWV 值作为模型先验值参与建模,建立 XGBZE-PWV 模型。3 个模型的详细信息见表 1。

表 1 广西区域 3 个 PWV 估算模型详细信息

Tab.1 Detailed information on the three PWV estimation models for Guangxi

模型	模型输入	模型输出
XGBZ-PWV	时间信息(DOY 和 HOD)、位置信息(Longitude、Latitude 和 Height)和 GNSS ZTD	
XGBZG-PWV	时间信息(DOY 和 HOD)、位置信息(Longitude、Latitude 和 Height)和 GNSS ZTD,GAskne_PWV	GNSS PWV
XGBZE-PWV	时间信息(DOY 和 HOD)、位置信息(Longitude、Latitude 和 Height)和 GNSS ZTD,GXE_PWV	

以上 3 个模型均使用 2020-01-01—2021-12-31 共 2 a 的数据进行建模,2022-01-01—12-31 数据用于模型验证,数据的时间分辨率均为 1 h。同时,基于 GNSS 获得的 ZTD,利用 GPT3 模型提供的气压和温度由式(1)、式(2)、式(3)和式

(4)反演 PWV(称为 GPT3-PWV 模型),与所建 PWV 模型进行对比。

2.2 精度评价指标

使用 bias 与均方根误差(RMSE)评价模型精度:

$$\text{bias}=\frac{1}{n}\sum_{i=1}^n(y_i-\hat{y}_i)$$

(8)

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^n(y_i-\hat{y}_i)^2}$$

(9)

式中, y_i 为 GNSS PWV 实际值; $\hat{y}_i$ 为模型 PWV 值; n 为样本数量。

2.3 精度评估

以 2022 年广西区域基于 GNSS ZTD、ERA5 地表气压和温度反演的 GNSS PWV 作为参考值,验证所建立的 XGBZ-PWV、XGBZG-PWV 和 XGBZE-PWV 3 个模型的估算精度,并与 GPT3-PWV 模型进行对比。分别计算 4 个模型的平均 bias 和 RMSE(表 2),并绘制站点误差空间分布(图 2)。

表 2 4 个模型估计 PWV 的 bias 和 RMSE 统计

Tab.2 The bias and RMSE statistics of the four models for estimated PWV

模型	bias/mm			RMSE/mm		
	最小值	最大值	平均值	最小值	最大值	平均值
GPT3-PWV	-1.57	1.33	0.05	1.74	3.38	2.48
XGBZ-PWV	-0.28	0.78	0.10	1.42	2.56	1.91
XGBZG-PWV	-0.33	0.36	0.03	1.17	2.49	1.76
XGBZE-PWV	-0.37	0.70	0.13	1.10	2.47	1.70

由表 2 可知,相比于 GPT3-PWV 模型,新建立的 3 个模型的估算精度均有所提升。从 bias 来看,GPT3-PWV 模型的 bias 范围变化较大,而 XGBZ-PWV、XGBZG-PWV 和 XGBZE-PWV 三个模型的 bias 变化范围较小。从 RMSE 来看,所建立 3 个模型的平均 RMSE 远小于 GPT3-PWV 模型。相比于 GPT3-PWV 模型,XGBZ-PWV、XGBZG-PWV 和 XGBZE-PWV 估算精度分别提升 22.98%、29.03% 和 31.45%,其中 XGBZE-PWV 提升幅度最大。相比于 XGBZ-PWV 模型,XGBZG-PWV 和 XGBZE-PWV 由于增加了先验条件,从而提高了模型的 PWV 估算能力。

由图 2 可知,GPT3-PWV 模型存在较大的偏差,且地域性分布明显,主要表现为东部和南部

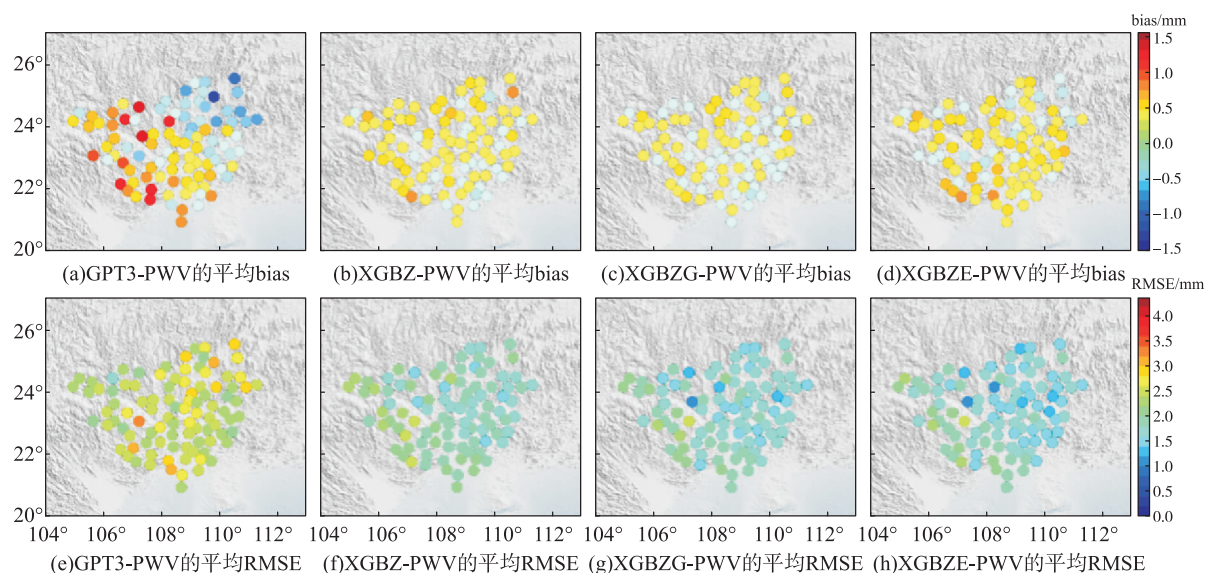


图2 4个模型的平均 bias 和 RMSE 空间分布

Fig.2 Spatial distribution of mean bias and RMSE for the four models

地区为正 bias,而北部和西部地区表现为负 bias。XGBZ-PWV、XGBZG-PWV 和 XGBZE-PWV 整体表现为较小的正 bias,分布较稳定。从 RMSE 分布可知,GPT3-PWV 模型的整体 RMSE 均较大,所建立 3 个模型的整体 RMSE 均较小。整体上,低海拔区域的 PWV 估算精度要高于高海拔区域。相比于 XGBZ-PWV 模型,XGBZG-PWV 和 XGBZE-PWV 在大多数站点具有更小的

RMSE,主要表现在部分中部地区以及东南部区域,说明在建模过程中加入经验 PWV 是有必要的。XGBZG-PWV 和 XGBZE-PWV 在空间分布上相差较小,均较稳定。

为分析 4 个模型在不同季节的估算性能,分别计算各模型在春季、夏季、秋季和冬季的平均 bias 和 RMSE,结果见图 3。

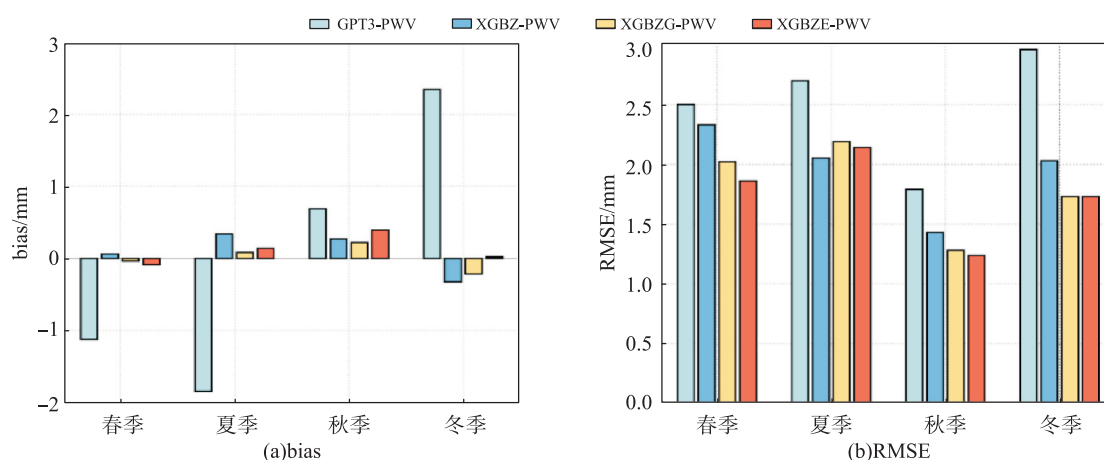


图3 4个模型季节性误差统计

Fig.3 Statistics on seasonal errors in four models

由图 3 可知,GPT3-PWV 模型在 4 个季节的绝对 bias 均最大,相比之下,XGBZ-PWV、XGBZG-PWV 和 XGBZE-PWV 整体 bias 较小,其中 XGBZG-PWV 和 XGBZE-PWV 较稳定。由 RMSE 的季节性变化可知,GPT3-PWV 模型均表现出较大的 RMSE。在夏季,XGBZE-PWV 和 XGBZG-PWV 模型引入了先验 PWV 信息,经验 PWV 的时间和空间分辨率有限,并且只能反映

长时间 PWV 的变化趋势,不能刻画详细的 PWV 变化,对于强对流天气反而会引入数据误差,从而在复杂气象条件下增加模型的拟合难度。但整体上 XGBZG-PWV 和 XGBZE-PWV 模型表现较好。基于 XGBoost 模型建立的 3 个估算模型均表现出更好的性能,其中 XGBZE-PWV 模型在每个季节的估算性能都更优,说明此模型能够较好地学习 PWV 的季节性变化。

以上结果表明,XGBoost 模型能够较好地学习 ZTD 和 PWV 之间的非线性特征,并且在建模过程中通过增加经验 PWV 先验值,可提高无气象参数情况下 PWV 的估算精度。

3 GNSS PWV 与降雨的时空分析

由 § 2.3 可知,XGBZE-PWV 模型表现更优异,因此后续使用该模型估算的 PWV 参与降雨分析。基于 GNSS PWV、XGBZE-PWV 估算的 PWV 和最近的气象站提供的降雨数据分析 2 个时间段(广西“龙舟水”期间(2022-06-15—20)和台风“暹芭”期间(2022-07-01—05))PWV 与降雨之间的关系(图 4)。

由图 4 可知,JZ95 与 JZ18 在 06-16、06-17 和

06-20 期间降雨密集,PWV 表现为高频上升。JZ33 与 JZ05 在 7 月初的极端降雨事件中,PWV 分别上升至 75~80 mm,显示长时间水汽积累与极端降水的直接关联。在降雨峰值区,PWV 主要呈现上升趋势,并且 PWV 峰值与降雨峰值之间存在时间差,这也验证了大气水汽积累是降雨形成的主要前提。降雨过程中或降雨后,PWV 出现短暂下降,主要是因为降雨过程消耗大气中的大部分水汽。从 4 个子图可以看出,XGBZE-PWV 模型的 PWV 估算值与 GNSS PWV 数值上有差异,但其时序变化基本一致,峰值区间也基本相同,这也表明 XGBZE-PWV 模型估算的 PWV 在极端环境下仍然具有较好的性能。

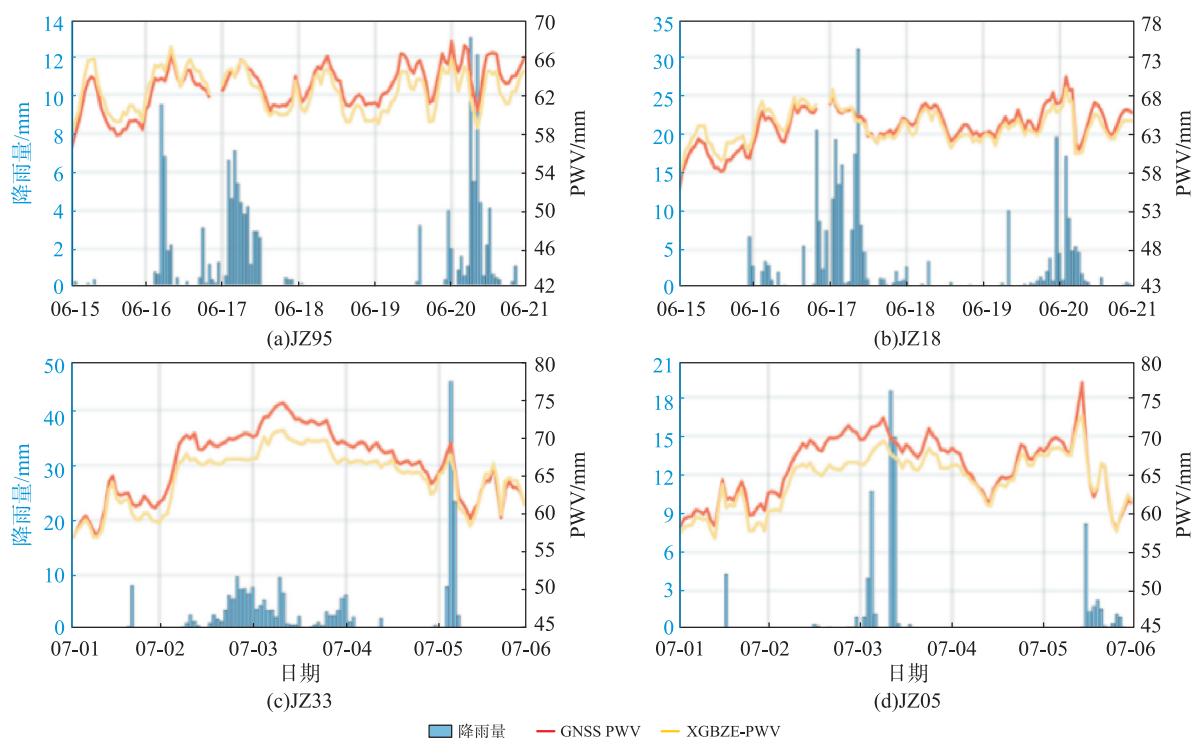


图 4 4 个站点逐小时 GNSS PWV、模型 PWV 与降雨时序

Fig.4 Hourly GNSS PWV, model PWV, and precipitation time series at four stations

为分析 PWV 和降雨之间的短期空间变化,分别选择 2022-06-15、2022-06-17 和 2022-06-20, 2022-07-01、2022-07-03 和 2022-07-05 作为分析时间,并使用自然邻域插值获取该时间段的 GNSS PWV、XGBZE-PWV 模型估算的 PWV 及气象站提供的降水数据的连续 PWV 和降雨量空间图。结果见图 5(“龙舟水”期间)和图 6(台风“暹芭”期间)。

由图 5 可知,XGBZE-PWV 模型估算的 PWV 与 GNSS PWV 具有较强的相似性,主要表现在 PWV 高值区域及水汽梯度变化。由 3 d 的

PWV 变化情况可知,在广西中部,PWV 值达 60~72 mm。并且随着时间的推移,水汽从东南逐渐向中部和北部移动。广西地势呈现西北高、东南低的走势,而水汽也主要集中在东南地区,西北区域值较低,这说明地形对水汽具有阻滞与聚集作用。降雨量的空间分布主要表现为“点-片”特征,并且主要降雨区域从东南区域逐渐向中部和北部移动。PWV 高值区主要对应降雨的核心区,其降雨量主要在 5~6 mm。降雨主要持续在东南地区 and 中部,这些区域 3 d 的 PWV 值均较高,表明充足的水汽条件是维持降水过程的

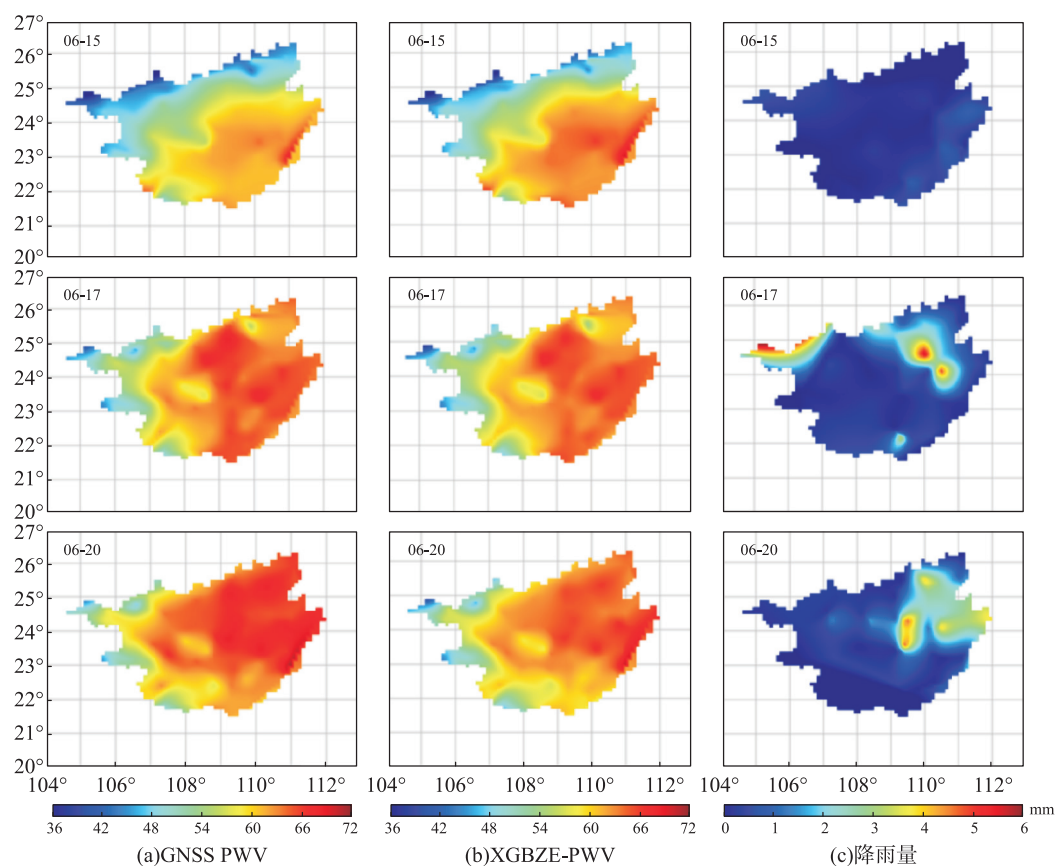


图 5 GNSS PWV、XGBZE-PWV 和降雨量的空间分布

Fig.5 Spatial distribution of GNSS PWV, XGBZE-PWV, and rainfall

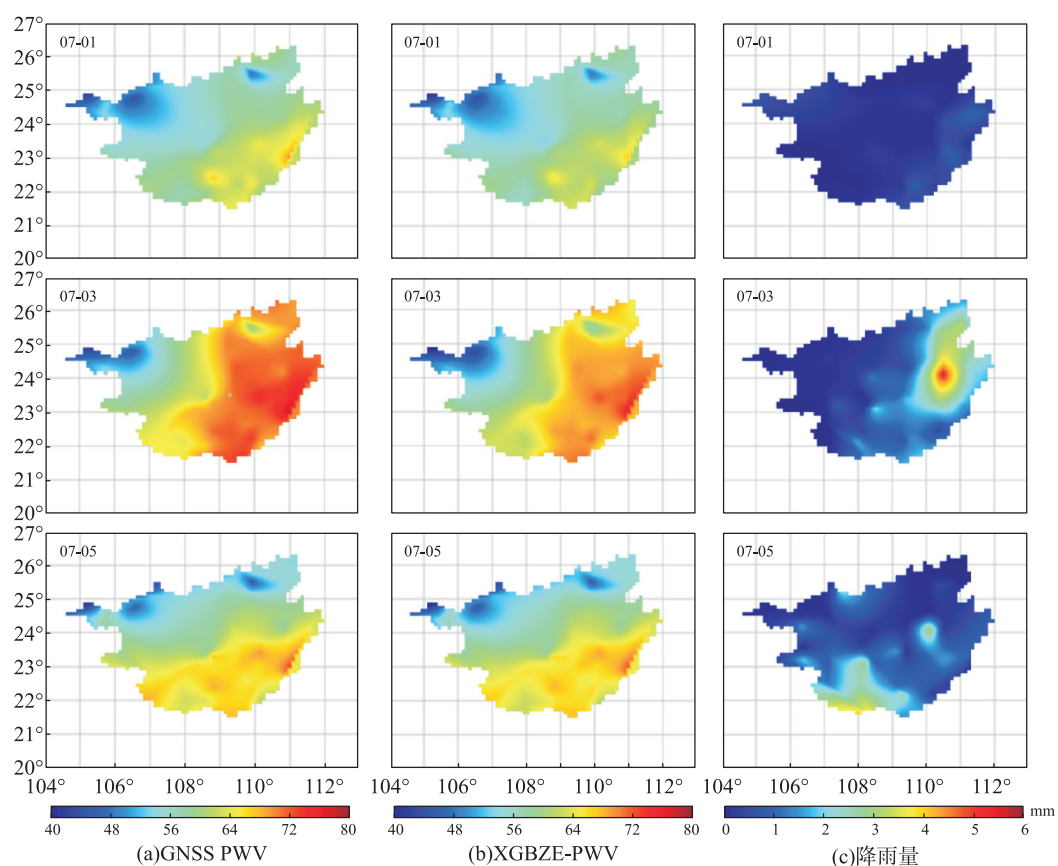


图 6 GNSS PWV、XGBZE-PWV 和降雨量的空间分布

Fig.6 Spatial distribution of GNSS PWV, XGBZE-PWV, and precipitation

重要前提。

由图 6 可知,虽然 3 d XGBZE-PWV 估算的 PWV 值略低于 GNSS PWV,但模型 PWV 值与 GNSS PWV 空间分布较为一致。07-01,除了南部与东南地区 PWV 值略高,其余地方 PWV 均较小,并无明显的降雨区。07-03,除西北部,其余区域 PWV 值均较高,在东南和中部地区值最大,且此时降雨的核心区域与 PWV 高值区相同,表明水汽充足的区域对应降雨集中区。07-05, PWV 高值区主要集中在东南区域,降雨区与 PWV 高值区有差异,这说明天气系统的演变仍然是影响降水空间分布和强度变化的重要原因。

整体来看,XGBZE-PWV 模型估算的 PWV 和 GNSS PWV 空间分布上较为一致,说明模型在极端降雨环境下精度相对较好,可用于极端环境下的水汽分析。

4 结 语

针对气象参数缺失情况下实时/近实时 PWV 反演受限的问题,基于 XGBoost 算法分别构建 XGBZ-PWV 模型、XGBZG-PWV 模型和 XGBZE-PWV 模型,并评价模型的精度;后续分析极端降雨环境下模型估算的 PWV、GNSS PWV 和降雨的时空变化。主要结论如下:

1)3 个模型相比于 GPT3-PWV 模型均较为优异,XGBZ-PWV、XGBZG-PWV 和 XGBZE-PWV 估算精度分别提升 22.98%、29.03% 和 31.45%。XGBZG-PWV 和 XGBZE-PWV 由于增加了经验 PWV 建模,提升了模型的精度,其中,XGBZE-PWV 精度最高。

2)在 2022 年 2 次极端降雨过程中,对 PWV 与降雨的时空演变特征进行分析。结果表明,水汽积累是降雨形成的主要前提,而地形对水汽具有阻滞与聚集作用。XGBZE-PWV 模型估算的 PWV 与 GNSS PWV 具有较好的时空一致性,在极端水汽条件下依然具有较高的适用性。

参考文献

- [1] King M D, Kaufman Y J, Menzel W P, et al. Remote Sensing of Cloud, Aerosol, and Water Vapor Properties from the Moderate Resolution Imaging Spectrometer (MODIS)[J]. IEEE Transactions on Geoscience and Remote Sensing, 1992, 30(1): 2-27
- [2] 何琦敏,张克非,李黎,等. 基于 PWV 到达时间差估计台风运动状态的四参数模型[J]. 测绘学报, 2024, 53(11): 2125-2137(He Qimin, Zhang Kefei, Li Li, et al. A Four-Parameter Model for Estimating Typhoon Motion States Based on Time Difference of PWV Arrival[J]. Acta Geodaetica et Cartographica Sinica, 2024, 53(11): 2125-2137)
- [3] 邓南山,刘洋,李小伟,等. 基于北斗/GNSS PWV 数据构建的短临强降雨模型及其应用[J]. 测绘通报, 2024(9): 96-100(Deng Nanshan, Liu Yang, Li Xiaowei, et al. Constructing a Short Term Heavy Rainfall Model Based on Beidou/GNSS PWV Data and Its Application[J]. Bulletin of Surveying and Mapping, 2024 (9): 96-100)
- [4] 李黎,宋越,易金花,等. 对流层延迟与可降水量直接转换模型研究[J]. 大地测量与地球动力学, 2019, 39(5): 492-495(Li Li, Song Yue, Yi Jinhua, et al. Study on the Direct Conversion Model between Tropospheric Delay and Precipitable Water Vapor[J]. Journal of Geodesy and Geodynamics, 2019, 39(5): 492-495)
- [5] 韦云,王迅,王浩,等. 中国东南沿海地区 PWV 直接转换模型研究[J]. 大地测量与地球动力学, 2022, 42(7): 750-754 (Wei Yun, Wang Xun, Wang Hao, et al. Research on the Direct Conversion Model of PWV in Coastal Areas of Southeastern China[J]. Journal of Geodesy and Geodynamics, 2022, 42(7): 750-754)
- [6] Zheng Y X, Lu C X, Wu Z L, et al. Machine Learning-Based Model for Real-Time GNSS Precipitable Water Vapor Sensing[J]. Geophysical Research Letters, 2022, 49(3)
- [7] Zhang H, Yao Y B, Xu C Q, et al. A Novel Multilayer Perceptron-Based Nonmeteorological Parameters PWV Retrieval Model[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62
- [8] Huang L K, Lu D H, Chen F D, et al. A Deep Learning-Based Approach for Directly Retrieving GNSS Precipitable Water Vapor and Its Application in Typhoon Monitoring [J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62
- [9] Liu Y, Zhang B, Yao Y B, et al. Revealing the Spatiotemporal Patterns of Water Vapor and Its Link to North Atlantic Oscillation over Greenland Using GPS and ERA5 Data [J]. Science of the Total Environment, 2024, 918
- [10] Zhao Q Z, Liu K, Sun T T, et al. A Novel Regional Drought Monitoring Method Using GNSS-Derived ZTD and Precipitation[J]. Remote Sensing of Environment, 2023, 297
- [11] Manandhar S, Lee Y H, Meng Y S, et al. GPS-Derived PWV for Rainfall Nowcasting in Tropical Region[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(8): 4835-4844
- [12] Xiong L, Huang L K, Mo Z X, et al. Evolution of GNSS-Derived Precipitable Water Vapor and Its Driving Factors during the “Dragon Boat Water” Rainfall Event in Guangxi, China[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2025, 18: 14308-14323
- [13] 江婷,李黎,田莹,等. 基于湖南 CORS 网的 PWV 时空变化分析及其在暴雨落区预报中的应用[J]. 大地测量与地球动

- 力学, 2018, 38(7): 707-713 (Jiang Ting, Li Li, Tian Ying, et al. Analysis of PWV Spatiotemporal Variation and Its Application to Forecast the Falling Area of Rainstorm Based on Hunan CORS[J]. Journal of Geodesy and Geodynamics, 2018, 38(7): 707-713)
- [14] Huang L K, Mo Z X, Liu L L, et al. Evaluation of Hourly PWV Products Derived from ERA5 and MERRA-2 over the Tibetan Plateau Using Ground-Based GNSS Observations by Two Enhanced Models[J]. Earth and Space Science, 2021, 8(5)
- [15] Ke F Y, Zhao P, Yu W, et al. Response of Meiyu Process Considering the Temporal and Spatial Characteristics of GNSS PWV [J]. Theoretical and Applied Climatology, 2024, 155: 1301-1319
- [16] Wang Q, Yao Y B, Liu C, et al. Coordinated Variation of Precipitable Water Vapor and Rainfall during Typhoon[C]. 2022 3rd International Conference on Geology, Mapping and Remote Sensing (ICGMS), Zhoushan, 2022
- [17] 赵庆志, 刘康, 李祖锋, 等. GNSS 和非实测气象参数的 PWV 反演方法及其精度评估[J]. 武汉大学学报: 信息科学版, 2024, 49(3): 453-464 (Zhao Qingzhi, Liu Kang, Li Zufeng, et al. PWV Inversion Method Based on GNSS and Non-Measured Meteorological Parameters and Accuracy Evaluation[J]. Geomatics and Information Science of Wuhan University, 2024, 49(3): 453-464)
- [18] Huang L K, Wang X, Xiong S, et al. High-Precision GNSS PWV Retrieval Using Dense GNSS Sites and In-Situ Meteorological Observations for the Evaluation of MERRA-2 and ERA5 Reanalysis Products over China[J]. Atmospheric Research, 2022, 276
- [19] 林买金, 张露露, 谢劲峰, 等. 基于高斯函数的中国区域时变性水汽垂直剖面模型[J]. 大地测量与地球动力学, 2024, 44(10): 1016-1021 (Lin Maijin, Zhang Lulu, Xie Shaofeng, et al. Vertical Profile Model of Regional Time-Varying Water Vapor in China Based on Gaussian Function[J]. Journal of Geodesy and Geodynamics, 2024, 44(10): 1016-1021)

Estimation of PWV and Analysis of Extreme Rainfall in Guangxi Region Based on XGBoost Model

YAN Liwei¹ LIN Maijin² XIE Shaofeng¹ HUANG Liangke¹
LI Xianghong³ FANG Qiongyu³

¹ College of Geomatics and Geoinformation, Guilin University of Technology, Guilin 541006, China

² College of Geological Engineering and Geomatics, Chang'an University, Xi'an 710054, China

³ Guilin Meteorological Bureau, Guilin 541001, China

Abstract: Aiming at the limitation of real/near-real-time GNSS PWV retrieval under missing meteorological parameters, three PWV estimation models without the need for measured meteorological parameters were established in the Guangxi region based on the XGBoost model. First, the XGBZ-PWV model was developed with inputs including station time (DOY and HOD), location (Longitude, Latitude, and Height), and GNSS ZTD, and the output feature being GNSS PWV. Then, based on the XGBZ-PWV model, two empirical PWV values were incorporated to establish the XGBZG-PWV and XGBZE-PWV models, respectively. For comparison, the GPT3 model was used to provide pressure and temperature for PWV retrieval based on GNSS ZTD (GPT3-PWV model). The accuracy of the established models was validated using GNSS PWV retrieved from GNSS ZTD, ERA5 surface pressure, and temperature in the Guangxi region in 2022 as the reference value. The results show that, compared to the GPT3-PWV model, the estimation accuracy of the XGBZ-PWV, XGBZG-PWV, and XGBZE-PWV models improved by 22.98%, 29.03%, and 31.45%, respectively, with the XGBZE-PWV model performing the best. During two extreme rainfall events in 2022, the spatiotemporal evolution characteristics of PWV and rainfall were analyzed. The results demonstrate that the XGBZE-PWV model maintains good applicability even under extreme weather conditions.

Key words: ZTD; PWV; GNSS; XGBoost; rainfall