

大规模 GNSS 观测数据一体化预处理 系统设计与实现

任营营^{1,2}

- 中移(上海)信息通信科技有限公司, 上海 201210;
- 中移智行网络科技有限公司, 上海 201210)

摘要: 针对全球 GNSS 观测数据规模持续增长及多源数据管理、质量控制和自动化处理的迫切需求, 设计并实现了一种面向大规模 GNSS 观测数据的一体化预处理系统; 该系统采用模块化分层架构, 集成格式转换、数据编辑、质量检查、单点定位、结果可视化及报告生成等功能, 具有功能完备性、结果可靠性、输出多样性、高度自动化和高效并行化等显著技术特征; 基于 135 个基站的观测数据进行性能测试, 结果表明: 在六核十二线程的硬件环境下, 该系统可将数据处理耗时由 98 分钟缩短至约 17 分钟, 效率提升约 576.5%, 当并行度设置接近 CPU 物理核心数时可获得最佳性能; 该系统有效降低了大规模 GNSS 数据预处理的技术门槛, 显著提升了处理效率和可靠性, 为 GNSS 精密定位、误差建模及气象反演等应用提供了高效的数据支撑。

关键词: 关键词: GNSS; RINEX; 预处理; 并行化; 一体化

Design and Implementation of Integrated Preprocessing System for Large-Scale GNSS Observation Data

REN Yingying^{1,2}

- China Mobile (Shanghai) Information and Communication Technology Co., Ltd., Shanghai 201210, China;
- China Mobile Intelligent Mobility Network Technology Co., Ltd., Shanghai 201210, China)

Abstract: In response to the rapid growth of global navigation satellite system (GNSS) observation data and urgent demands for multi-source data management, quality control, and automated processing, an integrated preprocessing system for large-scale GNSS observation data is designed and implemented, which adopts a modular layered architecture and integrates functionalities including format conversion, data editing, quality checking, single-point positioning, result visualization, and report generation. It is characterized by comprehensive functionality, reliable results, diverse outputs, high automation, and efficient parallelization. A performance test based on the observation data from 135 reference stations shows that, in a six-core twelve-thread computing environment, the system reduces the data processing time from 98 minutes to approximately 17 minutes, improving a 576.5% efficiency, and the optimal performance is obtained with the parallelism level close to the number of physical CPU cores. This system effectively reduces the complexity for preprocessing large-scale GNSS data, significantly enhances its efficiency and reliability, and provides a strong data support for applications such as GNSS precise positioning, error modeling, and atmospheric inversion.

Keywords: GNSS; receiver independent exchange format (RINEX); preprocessing; parallel; integration

0 引言

近年来, 在全球定位技术进步和多元化应用需求推动下, 全球导航卫星系统 (GNSS, global navigation

satellite system) 连续运行参考站 (CORS, continuous-operating reference stations) 网络呈现指数级增长态势^[1]。截至当前, 全球已建成覆盖各大洲及主要区域的

收稿日期: 2025-08-11; 修回日期: 2025-10-14。

基金项目: 导航与时空技术国家级重点实验室自主基金课题 (WXDHZZJJ20250104); 北斗和 5G 融合的多领域规模化应用示范 (2024YFB3910105)。

作者简介: 任营营 (1994-), 男, 博士, 工程师。

引用格式: 任营营. 大规模 GNSS 观测数据一体化预处理系统设计与实现[J]. 计算机测量与控制, 2026, 34(3): 208-215.

数千个 CORS 站点，形成多层次观测体系，涵盖国际 GNSS 服务 (IGS, international GNSS service) 全球跟踪网、日本 GNSS 地球观测网络系统、美国国家 CORS 网络、欧洲永久 GNSS 观测网络以及中国大陆构造环境监测网络等。这些基础设施不仅为地壳形变监测、高精度定位测量、气象学水汽反演及导航增强服务等应用领域提供了重要的数据支撑，更促进了多系统兼容、多频段融合的 GNSS 观测数据的快速积累^[2-4]。当前，规模化发展的 GNSS 观测网络显著提升了全球卫星导航数据的时空分辨率与覆盖能力，为高精度时空基准构建、地球动力学机理研究及多学科交叉应用奠定了坚实的数据基础。然而，伴随观测规模的扩大，多源异构数据的管理整合、标准化处理流程的优化以及数据质量控制体系的完善，已成为 GNSS 数据资源高效利用的挑战之一。

为实现跨网络、跨设备的 GNSS 数据共享与联合处理，接收机无关交换格式 (RINEX, receiver independent exchange format) 被提出。自 1989 年首次提出以来，RINEX 格式经历了多次迭代，从最初仅支持全球定位系统 (GPS, global positioning system) 单系统观测数据，逐步发展为支持多星座、多频段及多类观测信息的通用文件格式标准^[5]。随着格洛纳斯系统 (GLO-NASS, global navigation satellite system)、伽利略系统 (Galileo, galileo navigation satellite system)、北斗卫星导航系统 (BDS, BeiDou navigation satellite system)、准天顶卫星系统 (QZSS, quasi-zenith satellite system) 等卫星系统的建成与运行，RINEX 格式不断扩展数据字段以满足多系统集成需求，特别是 RINEX3 及后续版本在数据结构、信号标识及导航电文解析等方面进行了系统性优化，实现了对现代化 GNSS 全面兼容^[6]。最新发布 RINEX4 更进一步支持新型导航信息的记录与交换，使 GNSS 数据能够在国际 GNSS 服务等大规模观测网络中高效共享与处理^[7]。然而，随着数据量的急剧增长与数据源的多样化，仅依赖 RINEX 格式标准，尚不足以解决数据质量控制、错误检测与自动化处理等问题。这也促使 GNSS 数据预处理技术成为保障多源、多星座数据可靠性的关键环节。在此背景下，GNSS 数据预处理的重要性愈发凸显。

GNSS 数据预处理通常包括观测文件的编辑与合并、质量检查、接收机时钟跳变修正、周跳探测以及电离层延迟改正等关键步骤。高质量的数据预处理不仅是保障多源、多星座观测数据可靠性的基础环节，也是精密解算软件 (如 GAMIT、Bernese、GIPSY) 获得稳定、可信结果的前提条件^[8-10]。目前较为知名的相关软件工具包括 TEQC、GFZRNX 与 G-Nut/Anubis 等^[11-13]，并已有学者 (如游学真等) 基于上述工具开展了相关研究^[14-19]。然而，这些工具在功能完善性与适应

性方面仍存在一定局限：TEQC 作为经典命令行工具，虽集成了数据编辑与质量检查功能，但自 2019 年停止更新后已无法支持 RINEX3 及更高版本数据；GFZRNX 虽可兼容所有 RINEX 版本，但主要面向文件编辑，缺乏对质量检查与自动化处理的全面支持；G-Nut/Anubis 虽能实现多系统质量检查，但免费版功能受限，且在批量化与并行化处理能力方面存在不足。用户往往需要组合多款软件，甚至自行开发脚本，以完成完整的多星座数据预处理流程。这不仅增加了使用门槛，也显著降低了处理效率，制约了大规模 GNSS 数据在精密定位、误差建模与气象反演等领域的进一步应用。因此，开发一种集成多功能、支持最新 RINEX 版本、具备高效自动化能力的新型数据预处理工具，已成为 GNSS 数据处理领域的迫切需求。近年来，一些新型的 GNSS 数据预处理方法应运而生。随着云计算和大数据技术的发展，基于云平台的 GNSS 数据处理框架逐渐受到关注。例如，利用分布式计算资源进行并行处理，不仅能够提升数据处理的效率，还能通过动态扩展计算资源解决大规模数据集处理中的瓶颈^[20]。此外，人工智能和机器学习技术的引入，为 GNSS 数据质量检查提供了新的思路。基于深度学习和自适应算法的异常检测方法，能够自动识别并修复数据中的错误或异常，从而提升数据预处理的精度和效率^[21]。

针对上述问题，本文设计并实现了一款面向大规模 GNSS 观测数据的一体化预处理软件——TECPVR (Translation, Editing, Checking, Positioning, Visualization and Reporting)。该工具包基于 Python 语言开发，融合现有工具的功能优势，构建了高度集成、流程自动化、并行加速的处理框架。通过模块化设计，TECPVR 可自动完成数据解压、格式转换、观测文件编辑、质量检查、单点定位及可视化输出等全流程操作；多核并行计算的引入显著提升了处理效率，而可视化交互界面的设计则有效降低了使用门槛，适用于科研机构、高精度定位服务提供商及测绘单位等多类用户。本文将围绕 TECPVR 的系统架构、功能设计、性能评估与实验分析展开论述，旨在提供一种通用性强、灵活性高且具备可扩展性的大规模 GNSS 数据高效预处理解决方案，为推动 GNSS 在更广泛时空信息应用领域的高质量发展奠定坚实基础。

1 系统架构设计

TECPVR 软件的系统架构设计遵循模块化、自动化与可扩展性原则，主要由数据层、配置层、处理层和输出层四个核心模块构成，各模块之间通过基于文件的数据交互机制实现高效传输与无缝集成 (如图 1 所示)。该架构不仅支持多星座、多频段 GNSS 数据的批量处

理，还能在保证结果可追溯性的前提下，实现灵活的处理策略与高效的并行计算。

1.1 数据层

数据层是整个系统的基础，主要负责原始输入数据的管理及预处理过程中产生的中间文件存储。该层支持多种数据源，包括：

- 1) GNSS 观测数据：RINEX 格式（压缩或非压缩）、接收机原始二进制文件、RTCM（Radio Technical Commission for Maritime Services）数据。
- 2) 辅助数据：导航电文、精密星历等。
- 3) 为便于批量任务的组织与溯源，TECPVR 在运行时会自动创建项目文件夹，并采用规范化目录结构（见表 1），实现数据的有序化存储与管理。

表 1 TECPVR 项目文件夹目录结构

一级目录	二级目录	三级目录
项目名称	0_config	TECPVR.ini
	1_input	0_obs
		1_nav
		2_sp3
	2_output	0_log
		1_translation
		2_editing
		3_checking
		4_positioning
		5_visualization
6_report		

1.2 配置层

配置层用于定义数据处理的策略与流程，确保整个软件可在无需人工干预的情况下自动化运行。该层采用可读性强的配置文件，支持以下功能。

- 1) 处理参数的全局定义（如处理时间跨度、并行线程数）；
- 2) 功能模块开关控制（如是否启用数据质量检查或单点定位计算）；
- 3) 多任务批量处理的配置模板，以便快速适应不同应用场景。

1.3 处理层

处理层是 TECPVR 的核心，负责 GNSS 数据的预处理与计算任务。该层通过多线程架构与任务队列调度机制，能够高效完成以下功能：

- 1) 数据转换：将原始二进制文件、RTCM 及压缩格式文件转换为标准化 RINEX 格式，支持 RINEX2 至 RINEX4 的多版本兼容。
- 2) 数据编辑：实现 RINEX 文件的拼接、拆分、格式修正、文件重命名及观测值提取；同时对异常观测（如周跳、钟跳、数据缺口）进行修复。
- 3) 质量检查：统计观测量、历元数、可见卫星数、

信噪比等指标，并生成多系统、多频点的数据质量报告。

- 4) 单点定位：针对 GNSS 多系统执行单点定位计算，验证数据可用性并为后续精密处理提供参考结果。

1.4 输出层

- 1) 输出层负责对处理结果进行存储、组织与可视化呈现。TECPVR 支持多种输出形式；
- 2) 日志文件：记录处理流程及错误信息，便于溯源与调试；
- 3) 质量统计：以图表形式展示各类质量指标（如周跳分布、可见卫星数、数据缺口）；
- 4) 结果文件：生成标准化的文本、表格和文档报告，便于统计使用。

如图 1 所示，TECPVR 提供了统一且完整的数据处理流程。用户仅需通过软件界面导入待处理的 GNSS 观测数据，并设定具体处理策略，即可一键启动数据预处理流程。软件随后将自动按照用户的设置执行数据转换、编辑、检查、定位、可视化与报告生成等多项任务。

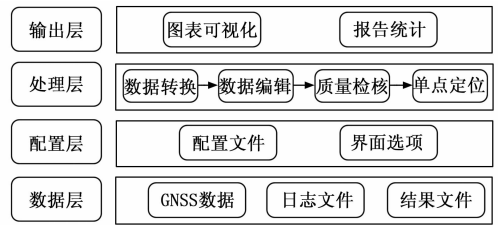


图 1 软件架构设计

一般情况下，用户可按照如下步骤进行 GNSS 数据预处理：

- 1) 导入多种格式的原始 GNSS 观测数据，软件支持二进制文件及压缩或未压缩的 RINEX 文件格式。
- 2) 通过用户界面详细设置数据处理策略，包括各功能模块的具体处理选项。此外，用户还可自定义并行处理的线程数，默认采用计算机 CPU 的物理核心数量进行并行处理。
- 3) 读取原始 GNSS 观测数据文件并转换为标准的 RINEX 格式，支持压缩文件、接收机日志和 RTCM 数据等。
- 4) 进行 RINEX 文件的拼接、拆分、数据提取和修改，同时实现 RINEX 格式转换、检查修复与文件重命名。
- 5) 统计 GNSS 观测数据的观测量及历元数，对观测数据质量进行检查。
- 6) 针对 GPS、GLONASS、Galileo 及北斗等卫星导航系统，执行单点定位计算。
- 7) 对 GNSS 数据的质量检查与统计结果进行可视化展示。
- 8) 以文本、表格、图像和 Word 文档等多种格式

输出最终处理结果。

2 系统功能设计

2.1 主要功能设计

TECPVR 工具包主要包含数据转换、数据编辑、质量检查、定位计算、结果可视化与报告生成六大功能模块。

2.1.1 数据转换

- 1) 文件压缩与解压；
- 2) RINEX 观测文件格式压缩与解压；
- 3) 接收机原始数据向 RINEX 格式转换；
- 4) RINEX 文件的批量自动重命名。

2.1.2 数据编辑

- 1) RINEX 文件格式转换；
- 2) RINEX 文件格式检查；
- 3) RINEX 文件格式修复；
- 4) RINEX 文件拼接与拆分；
- 5) RINEX 头文件信息编辑与修正；
- 6) RINEX 文件的批量重命名；
- 7) RINEX 文件的数据采样、观测类型与卫星系统选择；
- 8) RINEX 空缺或稀疏的观测数据类型剔除；
- 9) 卫星编号的重命名管理。

2.1.3 质量检查

- 1) GNSS 信号、频段与卫星观测数据的统计分析；
- 2) 数据缺口及碎片化情况的检测；
- 3) 载波相位数据预处理（如钟跳检测、周跳识别与修复）；
- 4) 卫星的方位角与高度角信息分析；
- 5) 伪距多路径效应与载波相位信噪比分析；
- 6) 导航数据及卫星健康状态信息核查；
- 7) 文件头信息及数据格式的有效性检查。

2.1.4 定位处理

- 1) 支持 GPS、GLONASS、Galileo 和 BDS 四大卫星系统的单点定位；
- 2) 接收机钟差改正；
- 3) 定位精度因子的计算与分析。

2.1.5 可视化展示

- 1) 观测数据质量统计图形化展示；
- 2) GNSS 观测历元数量统计与分析；
- 3) 单点定位结果的图形化展示；
- 4) 数据质量检查结果的直观可视化呈现。

2.1.6 报告输出

支持文本、表格、图像及文档等多种格式的处理结果报告生成。

2.2 软件特色与优势

TECPVR 工具包在架构设计与实现过程中，充分

结合 GNSS 数据预处理的工程化需求与学术研究场景，形成了功能高度集成、结果可靠性强、输出形式多样、自动化程度高以及数据处理高效五大技术特色。

2.2.1 功能集成

TECPVR 工具包实现了从原始数据导入到结果输出的全流程集成化设计，涵盖 GNSS 数据预处理的各个环节，包括：

- 1) 数据格式转换：支持多种输入格式（RINEX、二进制、RTCM 等）的标准化转换；
- 2) 数据编辑：提供文件拼接、拆分、观测值修复、文件重命名等功能；
- 3) 数据质量检查：可自动生成多维度的质量评估指标；
- 4) 单点定位：快速计算多系统单点定位解，验证数据可用性；
- 5) 结果可视化与报告生成：自动输出图表、统计报告及处理日志。

借助上述一体化流程，用户无需依赖多种外部软件工具，即可在同一平台上完成 GNSS 数据预处理的全流程操作，从而显著提高了工作效率与可重复性。

2.2.2 结果可靠

TECPVR 内置集成了多种国际学术界广泛应用并经验证的工具组件，该工具均经过长期迭代优化与大量工程实践验证，确保数据处理结果的科学性与稳定性。此外，TECPVR 在调用这些工具的同时，通过自动化的日志记录与异常检测机制，保障了处理流程的可追溯性与输出结果的高可信度。

2.2.3 输出多样化

针对不同科研与工程应用场景，TECPVR 提供了多元化的结果输出形式，包括如下。

文本与表格文件：便于后续数据分析与再处理；

文档报告：自动生成可直接用于科研报告或项目文档的标准化结果说明；

图像文件：内置 30 余种可视化图表，覆盖 GNSS 数据质量检查的核心指标（如观测量分布、周跳统计、信噪比特性等），帮助用户直观评估数据质量并快速定位潜在异常。

这一特性显著降低了用户进行二次处理或图表生成的工作量，有助于实现数据处理结果的直接工程化或科研应用。

2.2.4 自动化

TECPVR 采用高度自动化的处理框架，用户仅需通过图形化界面导入原始观测数据并设定处理策略，即可一键启动批处理流程。软件将自动执行数据格式转换、质量检查、单点定位、可视化与报告生成等任务，最大限度减少了人工干预的可能性，显著降低了数据处

理的操作复杂度与时间成本。

此外, TECPVR 支持基于配置文件的批量处理模式, 用户可通过模板化配置快速完成大规模 GNSS 数据集的自动化处理, 尤其适用于连续运行的监测项目与大数据分析场景。

2.2.5 处理高效

TECPVR 采用高性能并行计算架构, 充分利用现代多核处理器的计算资源, 显著提升了处理效率, 并保证了结果的可靠性。实验表明, 在 Intel i5-10400 (六核心十二线程) 处理器环境下, TECPVR 的整体处理效率相比传统单线程方式可提升 5~6 倍, 能够在短时间内完成大规模 GNSS 数据集的预处理任务。这一特性使其特别适用于全球 GNSS 观测网、区域连续运行站网及地学科研中海量数据的高效处理需求。

2.3 数据预处理关键算法

2.3.1 周跳探测与修复

对于高精度的应用而言, 由于需要为每颗卫星及每个频率建立唯一的整周模糊度, 只有在卫星信号连续跟踪的情况下, 观测数据才能被有效利用。如果某一卫星信号的连续性中断, 并且无法估算相关的周跳, 那么必须在解算模型中引入额外的待估参数来加以处理。

由于原始观测噪声较大, 仅能用于检测较大幅度的周跳, 因此需引入不同波长组合: 通过长波长的宽巷组合及短波长的窄巷组合, 可有效提升周跳探测的灵敏度和精度。借助这些组合方法, 可检测幅度高达一周的周跳^[22]。

以 GPS 三频载波相位观测 (Φ_1 、 Φ_2 、 Φ_3) 为例, 可以构建以下 3 种组合形式:

$$\begin{cases} \varphi_{\text{ewl}} = \varphi_2 - \varphi_3 \\ \varphi_{\text{wl}} = \varphi_1 - \varphi_3 \\ \varphi_{\text{nl}} = 2\varphi_1 - \varphi_2 \end{cases} \quad (1)$$

其中: φ_{ewl} 、 φ_{wl} 、 φ_{nl} 分别对应超宽巷、宽巷和窄巷组合。

该方法主要包括以下 3 个步骤。

1) 超宽巷周跳初探测:

基于 Melbourne-Wübbena (MW) 组合形成超宽巷组合, 并计算相邻历元之间的差分。如果差值超过 0.5 周, 则判定发生周跳。MW 组合的最大优势在于其可直接用于超宽巷周跳的快速探测。

2) 宽巷周跳精确判定:

对宽巷组合与超宽巷组合进行历元间差分, 计算相邻历元的差值。当该差值大于 0.5 周时, 即可确认宽巷周跳。此步骤中, 由于波长较长, 电离层延迟对结果影响可忽略。

3) 窄巷周跳精细修复:

在最后一步, 将窄巷组合从宽巷组合中扣除, 并与

前述步骤相同, 计算相邻历元间的差分。周跳判定的阈值依然为 0.5 周。由于窄巷波长显著更短 (约 0.156 m), 因此必须考虑电离层延迟的影响。

经过上述步骤后, 可得到 3 类不同的周跳量 (ΔN_{ewl} 、 ΔN_{wl} 、 ΔN_{nl})。该周跳量可通过下式转化为原始观测量上的周跳量:

$$\begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \\ 2 & -1 & 0 \end{bmatrix} \begin{bmatrix} \Delta N_1 \\ \Delta N_2 \\ \Delta N_3 \end{bmatrix} = \begin{bmatrix} \Delta N_{\text{ewl}} \\ \Delta N_{\text{wl}} \\ \Delta N_{\text{nl}} \end{bmatrix} \quad (2)$$

其中: ΔN_1 、 ΔN_2 、 ΔN_3 分别为原始观测频率的周跳量。

2.3.2 多路径分析

多路径效应同时作用于全球导航卫星系统的两类基本观测, 即伪距与载波相位。然而, 相较于载波相位, 伪距观测对多路径误差更为敏感, 且不同类型接收机在该误差幅值上存在显著差异。多路径误差是影响伪距观测精度的主要因素之一, 尤其在单点定位和精密单点定位 (PPP, precise point positioning) 等方法中, 对整体定位精度具有直接影响。

在具备双频观测数据的条件下, 可通过特定的线性组合消除卫星与接收机间的几何关系及全部大气延迟效应, 从而对伪距多路径误差进行估计。然而, 该线性组合无法消除整周模糊度及任何差分偏差。尽管差分偏差在时间尺度上通常保持稳定, 但由于周跳的潜在存在, 模糊度并非始终恒定。因此, 在多路径误差估计中, 开展可靠的观测数据预处理 (周跳检测与修复) 是确保精度的关键步骤。

适用于所有频率伪距多路径 (MP, Multipath) 估计的通用线性组合公式如下^[23]:

$$MP_k = P_k - L_i - \beta(L_i - L_j) = P_k + \alpha L_i + \beta L_j \quad (3)$$

式中, $\alpha = -\frac{f_2^2 + f_k^2}{f_i^2 - f_j^2} \cdot \frac{f_j^2}{f_k^2}$, $\beta = -\frac{f_2^2 + f_k^2}{f_i^2 - f_j^2} \cdot \frac{f_j^2}{f_k^2}$, k, i, j

为频率 (或频段) 索引; P 为伪距观测值; L 为载波相位观测值; f 为对应频率的载波频率。

当 $k=i=1$ 且 $j=2$ 时, 可得第一频点伪距多路径误差组合公式:

$$MP_1 = P_1 - L_1 - \frac{2f_2^2}{f_1^2 - f_2^2}(L_1 - L_2) \quad (4)$$

同理, 当 $k=i=2$ 且 $j=2$ 时, 第二频点伪距多路径误差可表示为:

$$MP_2 = P_2 - L_2 - \frac{2f_1^2}{f_1^2 - f_2^2}(L_2 - L_1) \quad (5)$$

对于 $k=5, i=1$ 且 $j=2$ 或其他频率, 其多路径组合可推广为:

$$MP_5 = P_5 - L_5 - \frac{(f_1^2 + f_5^2)f_2^2}{(f_1^2 - f_2^2)f_5^2}(L_1 - L_2) \quad (6)$$

多路径误差的统计量通过在连续历元序列内计算观

测值的标准差来获得，其中均值部分用于表征残余系统偏差并加以扣除。该方法能够有效降低观测噪声的瞬时波动影响，从而提高多路径误差估计的稳定性和精度。

与传统依赖单一频点或需要额外数据预处理的多路径估计方法不同，本方法无需针对不同 GNSS 星座或信号类型进行专门化适配。基于线性组合的简易周跳检测与修复机制，可在不影响正常观测序列的情况下剔除受周跳干扰的历元，确保多路径误差估计结果的可靠性与一致性。

此外，该方法支持多频观测条件下的扩展应用。通过引入频率相关权重和广义线性组合模型，能够在 GNSS 多系统、多频段观测环境下保持一致的统计性能。这种通用性不仅适用于常规的单点定位和精密单点定位数据质量评估，还可作为多星座多频 PPP 及网络解算中多路径建模与噪声约束的前置环节，为后续精密定位提供更高质量的观测数据支撑。

3 实验结果与分析

3.1 预处理结果分析

为验证 TECPVR 软件 GNSS 数据预处理结果的准确性与可靠性，选取不同测站的 GNSS 数据预处理结果并与 TEQC 进行对比分析。对比的指标包括多路径误差 (MP1、MP2) 和周跳检测数量。具体数据如表 2 所示。

表 2 预处理结果对比

TECPVR			TEQC		
MP1	MP2	周跳	MP1	MP2	周跳
0.39	0.35	853	0.42	0.33	783
0.52	0.53	2980	0.55	0.55	2910
0.36	0.52	526	0.37	0.49	500
0.51	0.46	188	0.47	0.45	137
0.61	0.67	468	0.64	0.68	451
0.3	0.4	2794	0.32	0.44	2754
0.41	0.38	1264	0.4	0.39	1265

由表可知，TECPVR 与 TEQC 的预处理分析结果呈现出一定的一致性。在多路径误差 (MP1、MP2) 方面，两者数值基本一致，不同测站结果差异小于 0.1 m。在周跳探测数量上，TECPVR 整体上比 TEQC 检测到的数量更多。这种差异源于两种软件在周跳探测算法上有所不同。TECPVR 使用了更加严格的检测标准，因此对信号的异常值处理更加敏感。

3.2 软件性能分析

为验证 TECPVR 软件在大规模 GNSS 数据预处理中的性能与适用性，本研究基于全球 135 个 IGS 参考站的单日观测数据开展了系统实验。如图 2 所示，实验站点的分布覆盖了不同纬度带，包括中高纬度和赤道区域，确保了结果的广泛适用性。实验数据来源于 IGS

发布的 Binex (Binary Exchange Format) 格式压缩文件，单个文件大小约 5 MB，总计约 600 MB。解压缩并转换为标准 Binex 格式后，数据总量达到约 5 GB。实验环境采用搭载 Intel (R) Core (TM) i5-10400 处理器 (主频 2.9 GHz，六核十二线程) 和 1 TB 固态硬盘的工作站平台，以保证数据读写与计算的稳定性。

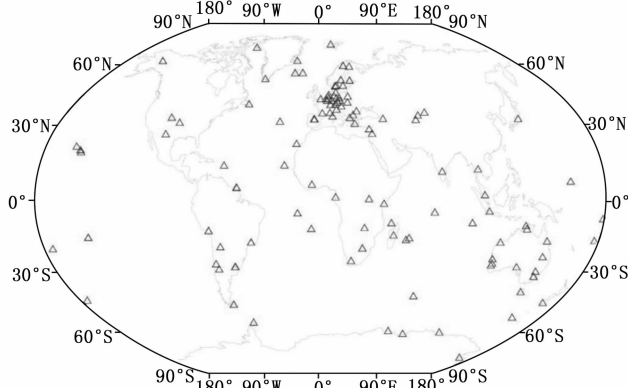


图 2 实验 GNSS 站点分布图

实验分别采用串行模式 (单线程) 和并行模式 (并行度设定为 12) 进行测试，测试功能包括文件解压、格式转换、数据编辑、信息统计及数据质量检查。各功能模块所用测试时间对比如图 3 及表 3 所示。结果表明，并行模式下整体处理效率较串行模式显著提升：串行模式完成整个数据集的处理耗时约 98 分钟，而并行模式仅需约 17 分钟，总体效率提升约 576.5%。其中，数据质量检查模块耗时最长，占总耗时约 43%；文件解压模块耗时最短，仅占 1%。进一步分析发现，除文件解压模块受限于 I/O 特性导致效率提升幅度相对较低 (约 125%) 外，其余功能模块在并行模式下的效率均实现约 6 倍的提升。

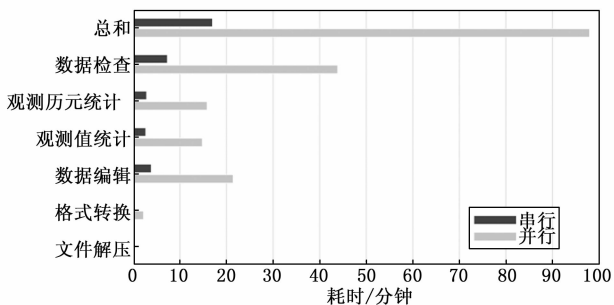


图 3 各功能模块串行与并行模式的处理耗时对比

表 3 各功能模块串行与并行模式处理耗时统计

(分钟)							
模式	解压缩	格式转换	数据编辑	观测值统计	观测历元统计	数据检查	总和
串行	0.25	2.18	21.42	14.82	15.83	43.9	98
并行	0.20	0.33	3.85	2.67	2.85	7.30	17
提升率	125.0%	660.6%	556.4%	555.1%	555.4%	601.4%	576.5%

在实验中,数据检查占总处理时间约 40%。通过进一步分析,发现周跳探测和信噪比分析功能是该模块耗时的主要部分。周跳检测采用多频段组合方法,对每个历元进行差分处理,计算相邻历元之间的差值。该步骤尤其在高频次和长时间段的观测数据中,处理时间显著增加。信噪比的统计分析需要对每个观测值进行计算,并对不同频段的信号进行分组。这一过程需要遍历整个数据集,且随着数据量的增加,计算量呈线性增长,因此是该模块的一个重要耗时环节。

为进一步探究并行度设置对数据处理效率的影响,本研究将并行度分别设为 1 至 15,并对比不同并行度下的整体处理耗时(见图 4)。结果表明:

1) 当并行度小于处理器物理核心数(6)时,处理效率随并行度呈近似线性提升;

2) 当并行度超过物理核心数但不超过逻辑核心数(12)时,效率仍有一定提升,但增益逐渐减弱,且 CPU 负载逐步接近其最大利用率,内存占用也开始增加;

3) 当并行度超过逻辑核心数(12)后,受限于线程调度与切换开销,整体效率出现下降。同时,CPU 利用率波动较大,I/O 等待时间增加,导致资源分配不均,影响整体系统性能。

综上,建议并行度的设置不应超过处理器的逻辑核心数。较优配置应接近或等于物理核心数,以避免因过度调度导致的性能瓶颈,同时保证计算性能与系统稳定性的平衡。

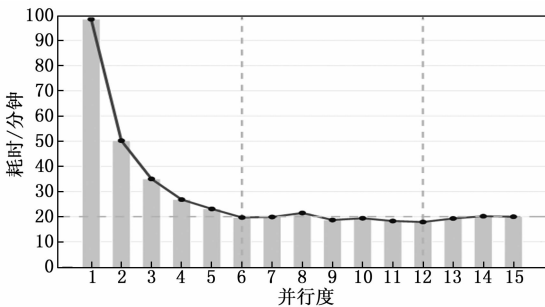


图 4 不同线程数下的 GNSS 数据预处理耗时对比

4 结束语

针对现有 GNSS 数据预处理工具在功能、效率及易用性方面的不足,设计并实现了一款面向大规模 GNSS 观测数据的一体化预处理系统 TECPVR (Translation, Editing, Checking, Positioning, Visualization and Reporting)。该系统采用模块化分层架构,集成数据转换、编辑、质量检查、单点定位、结果可视化及报告生成等核心功能,测试结果表明 TECPVR 与 TEQC 结果一致性较高,并引入多核并行计算与自动化任务调度机制,

显著提升了处理效率与结果可靠性。

基于全球 135 个 IGS 参考站单日观测数据的性能测试表明:在六核十二线程的硬件环境下,TECPVR 可将整体处理耗时由约 98 分钟缩短至 17 分钟,效率提升约 5~6 倍;其中,除文件解压模块外,各主要功能模块在并行模式下均获得了超过 5 倍的加速比。实验结果还表明,将并行度设置为接近或等于 CPU 物理核心数可在性能与稳定性之间取得较优平衡。

TECPVR 的开发有效降低了大规模 GNSS 数据预处理的技术门槛,提升了多源、多星座观测数据的处理能力与可视化水平,为精密定位、误差建模、气象反演等多领域应用提供了高质量的数据支撑。下一步将重点优化数据检查模块,尤其是信噪比分析、周跳检测和多路径效应修复等关键操作。计划通过引入更高效的自适应算法、分块计算和可选化功能,使系统能够根据数据质量和应用需求灵活调整计算资源,从而显著提升处理速度和灵活性。

未来工作将面向实时数据流和云端处理架构,进一步拓展系统对实时高精度定位、连续运行监测及大数据分析等新兴应用场景的支持能力。结合云计算和分布式技术,系统将能够处理更加庞大的实时数据集,提升系统的可扩展性和应对突发数据量的能力。此外,随着 AI 技术的进步,未来将进一步集成智能算法,自动优化处理流程 and 数据分析,增强系统的智能化水平,满足不断增长的多样化需求。

参考文献:

- [1] 李林阳,张学东,黄 娟,等. 大规模 GNSS 网发展及数据处理现状 [J]. 测绘通报, 2018 (10): 1-5.
- [2] SNAY R A, SOLER T. Continuously operating reference station (CORS): history, applications, and future enhancements [J]. Journal of Surveying Engineering, 2008, 134 (4): 95-104.
- [3] BRUYNINX C, LEGRAND J, FABIAN A, et al. GNSS metadata and data validation in the EUREF permanent network [J]. GPS Solutions, 2019, 23 (4): 106.
- [4] KAWAMOTO S, MUNEKANE H, HATANAKA Y, et al. How GNSS earth observation network system (GEONET) in japan contributes to geohazards mitigation [C] // AGU Fall Meeting Abstracts, 2014.
- [5] International GNSS Service. RINEX: The Receiver Independent Exchange Format Version 2.11 [EB/OL]. 2007. <https://files.igs.org/pub/data/format/rinex211.txt>
- [6] International GNSS Service. RINEX: The Receiver Independent Exchange Format Version 3.05 [EB/OL]. 2020. <https://files.igs.org/pub/data/format/rinex305.pdf>
- [7] International GNSS Service. RINEX: The Receiver Inde-

- pendent Exchange Format Version 4.02 [EB/OL]. 2024. https://files.igs.org/pub/data/format/rinex_4.02.pdf
- [8] LI Y. Analysis of GAMIT/GLOBK in high-precision GNSS data processing for crustal deformation [J]. *Earthquake Research Advances*, 2021, 1 (3): 100028.
- [9] MAO X, ARNOLD D, GIRARDIN V, et al. Dynamic GPS-based LEO orbit determination with 1 cm precision using the bernese GNSS software [J]. *Advances in Space Research*, 2021, 67 (2): 788 - 805.
- [10] BERTIGER W, BAR-SEVER Y, DORSEY A, et al. GipsyX/RTGx, a new tool set for space geodetic operations and research [J]. *Advances in Space Research*, 2020, 66 (3): 469 - 489.
- [11] ESTEY L H, MEERTENS C M. TEQC: the multi-purpose toolkit for GPS/GLONASS data [J]. *GPS Solutions*, 1999, 3 (1): 42 - 49.
- [12] CHEN X, NISCHAN T, DENG Z, et al. GFZRNX-QC: advanced GNSS data processing and quality control for multi-system observations [C] //EGU General Assembly Conference Abstracts, 2024.
- [13] VACLAVOVIC P, DOUSA J, GYORI G. G-Nut software library-state of development and first results [J]. *Acta Geodyn. Geomater*, 2013, 10 (4): 431 - 436.
- [14] 余文坤, 戴吾蛟, 杨哲. 基于 TEQC 的 GNSS 数据质量分析及预处理软件的设计与实现 [J]. *大地测量与地球动力学*, 2010, 30 (5): 81 - 85.
- [15] 范士杰, 郭际明, 彭秀英. TEQC 在 GPS 数据预处理中的应用与分析 [J]. *测绘信息与工程*, 2004 (2): 33 - 35.
- [16] 丁志鹏, 贺凯飞, 邱立杰, 等. 基于 Anubis 的北斗观测数据质量分析 [J]. *大地测量与地球动力学*, 2022, 42 (2): 143 - 147.
- [17] 康朝虎, 刘宁, 田永瑞, 等. Anubis 在 GNSS 数据质量可视化分析中的应用 [J]. *导航定位学报*, 2018, 6 (4): 105 - 111.
- [18] 白天阳, 兰孝奇, 陈浩. 基于 RTKLIB 预处理的 GNSS 数据分析 [J]. *地理空间信息*, 2019, 17 (2): 34 - 37.
- [19] 陈佳清. 利用 TEQC 及 RTKLIB 软件分析 GPS 数据质量 [J]. *导航定位学报*, 2016, 4 (1): 126 - 130.
- [20] 吕浩. 基于分布式计算技术的 GNSS 数据处理方法研究 [D]. 郑州: 解放军信息工程大学, 2015.
- [21] 李孟园. GNSS 观测数据质量智能化评估技术研究 [D]. 西安: 长安大学, 2023.
- [22] ZHAO Q, SUN B, DAI Z, et al. Real-time detection and repair of cycle slips in triple-frequency GNSS measurements [J]. *GPS Solutions*, 2015, 19 (3): 381 - 391.
- [23] VACLAVOVIC P, DOUSA J. Anubis-a tool for quality check of multi-GNSS observation and navigation data [C] //4th International Colloquium Scientific and Fundamental Aspects of the Galileo Programme, 2013.
- II 算法在航迹规划中应用 [J]. *计算机工程与应用*, 2020, 56 (3): 246 - 251.
- [19] ASUNCION A, NEWMAN D. UCI machine learning repository [EB/OL]. 2007.
- [20] 初蓓, 李占山, 张梦林, 等. 基于森林优化特征选择算法的改进研究 [J]. *软件学报*, 2018, 29 (9): 2547 - 2558.
- [21] A fast and elitist multiobjective genetic algorithm: NSGA-II [J]. *IEEE Transactions on Evolutionary Computation*, 2002, 6 (2): 182 - 197.
- [22] ZITZLER E, LAUMANN M, THIELE L. SPEA2: improving the strength pareto evolutionary algorithm [J]. *TIK report*, 2001, 103.
- [23] HU Y, ZHANG Y, GONG D. Multiobjective particle swarm optimization for feature selection with fuzzy cost [J]. *IEEE Transactions on Cybernetics*, 2020, 51 (2): 874 - 888.
- [24] ZHANG Q, LI H. MOEA/D: a multiobjective evolutionary algorithm based on decomposition [J]. *IEEE Transactions on Evolutionary Computation*, 2007, 11 (6): 712 - 731.

(上接第 207 页)

- [12] ZHANG Y, CHENG S, SHI Y, et al. Cost-sensitive feature selection using two-archive multi-objective artificial bee colony algorithm [J]. *Expert Systems with Applications*, 2019, 137: 46 - 58.
- [13] 肖迪, 葛启承, 林锦国, 等. 一种双种群遗传粒子群算法及在 SMB 优化中的应用 [J]. *南京理工大学学报*, 2012, 36 (1): 31 - 36.
- [14] 孙静. 多目标优化的多标签特征选择方法研究 [D]. 哈尔滨: 哈尔滨工程大学, 2022.
- [15] SCHOTT J R. Fault tolerant design using single and multicriteria genetic algorithm optimization [D]. Massachusetts Institute of Technology, 1995.
- [16] ZHOU A, JIN Y, ZHANG Q, et al. Combining model-based and genetics-based offspring generation for multi-objective optimization using a convergence criterion [C] //2006 IEEE International Conference on Evolutionary Computation, IEEE, 2006: 892 - 899.
- [17] 吴亚楠, 张剑. 一种基于 DQN 的改进 NSGA-II 算法 [J]. *舰船电子工程*, 2023, 43 (4): 29 - 33.
- [18] 封硕, 郑宝娟, 陈文兴, 等. 支持强化学习 RNSGA-