

梁毅, 刘东福生, 刘斌城, 等. 一种临近空间多源环境数据融合系统实现方法[J]. 智能计算机与应用, 2025, 15(11): 94-100.
DOI: 10.20169/j.issn.2095-2163.24030101

一种临近空间多源环境数据融合系统实现方法

梁毅, 刘东福生, 刘斌城, 秦筱械, 宋玮, 汤晶, 林靖

(航天科工海鹰集团有限公司, 北京 100070)

摘要: 随着临近空间科学探索和应用的推进, 其价值愈发凸显, 目前临近空间环境探测主要包含卫星、雷达、火箭、气球探测等手段, 但针对临近空间环境数据的处理研究对象主要为一种探测手段数据, 未对多源环境探测数据进行融合, 影响了环境探测准确性。为破解单一数据源的局限性, 设计了基于国产化技术支撑海量环境数据管理的一种临近空间多源环境数据融合系统, 可实现临近空间多源环境数据采集、数据存储、数据处理、数据融合、数据展示等功能。基于该系统可支撑完成临近空间全球环境特性分析、区域环境特性分析和卫星、雷达数据对比分析, 验证了系统的有效性和实用性。

关键词: 临近空间; 国产化; 卫星数据; 雷达数据; 数据融合

中图分类号: TM344.1

文献标志码: A

文章编号: 2095-2163(2025)11-0094-07

A implementation method for near space multi source environmental data fusion system

LIANG Yi, LIU Dongfusheng, LIU Bincheng, QIN Xiaowei, SONG Wei, TANG Jing, LIN Jing

(Haiying Group Co., Ltd., China Aerospace Science and Industry Corporation, Beijing 100070, China)

Abstract: With the advancement of exploration and application in near space, its value has become increasingly prominent. Currently, near space environment detection mainly includes satellite, radar, rocket, balloon detection and other means. However, the research object for processing near space environment data is mainly one detection method data, and the fusion of multi-source environmental detection data has not been carried out, which affects the accuracy of environmental detection. To overcome the limitations of a single data source and achieve systematic engineering applications, a near space multi-source environmental data fusion system based on domestic technology supporting massive environmental data management has been designed. Based on this system, functions such as near space multi-source environmental data collection, storage, processing, fusion, and display can be achieved. Based on this system, it can support the analysis of global and regional environmental characteristics in near space, as well as the comparative analysis of satellite and radar data, verifying the effectiveness and practicality of the system.

Key words: near space; localization; satellite data; radar data; data fusion

0 引言

临近空间是指高度位于 20~100 km 之间的区域, 被称为“近空间”、“亚轨道”或“空天过渡区”^[1]。在这一高度, 传统飞机遵循的空气动力学、卫星遵循的轨道动力学均难以适用。由于临近空间独特的空间环境特性, 使得临近空间飞行器相对于飞机和卫星相比具有显著优势, 临近空间飞行器与飞机相比具有持续工作时长、覆盖范围广、飞行自由且安全可靠等优势; 与卫星相比具有研制、发射和使用成本低、机动性好、对地观测分辨率高等优势。

临近空间逐渐成为一个国家空天商业开发和军事应用竞争的战略空间^[2]。因此, 为支撑临近空间开发利用, 对临近空间大气环境进行研究尤为重要。

随着临近空间应用不断发展, 临近空间环境研究进入了新阶段。邓潘^[3]构建了 532 nm 的单接收口径瑞利激光雷达系统, 得到了观测地临近空间高度 (20~40 km) 的大气密度、压力和温度廓线, 将瑞利激光雷达数据与观测地球载探空数据、风云三号 (FY-3C) 卫星数据、NRMSISE-00 大气模式和美国标准数据进行对比, 验证了激光雷达探测性能及算法的可靠性, 但主要还是偏重于激光雷达数据处理;

作者简介: 梁毅 (1991—), 男, 硕士, 工程师, 主要研究方向: 时空大数据处理与融合技术。

通信作者: 刘东福生 (1987—), 男, 硕士, 高级工程师, 主要研究方向: 信息系统架构, 大数据处理与融合技术。Email: ld19871103@126.com。

收稿日期: 2024-03-01

郝晓静等^[4]比较研究了昆明站 2 台不同频率流星雷达(37.5MHz、53.1MHz)与中频雷达的风场,完成了流星雷达和中频雷达数据处理,并实现 3 台雷达经纬向风速差异定量分析,但主要还是偏重于雷达一种探测方式数据处理,且探测范围具有局部特性,未能实现全球区域临近空间特性分析;冯超等^[5]研究了探空火箭探测手段在临近空间气象保障中的应用,进行了站点布局设计、探测方案制定和临近空间大气环境模型,但未与其他探测数据进行融合处理,也未进行工程应用;王梓皓等^[6]基于临近空间环境的探测需求,开展了高空气球平台的临近空间环境探测系统设计,包括飞行平台系统设计和传感器的设计与选型,通过飞行试验获取了温度、大气压、风速、风向等环境参数,但未与其他探测数据进行融合处理,探测空域范围具有一定局限性;介阳阳等^[7]利用 TIMED 卫星搭载的 SABER 传感器 v2.0 Level 2A 数据对临近空间大气环境参量进行了分析,包括大气温度、大气压力、密度和臭氧等常用大气环境参量,并分析其随不同高度、纬度和时间的变化特点,但主要偏重于理论研究,未进行系统设计;宋平^[8]基于气象卫星、气象火箭和探空气球探测,利用对比分析等方法对临近空间大气环境要素的分布规律和变化特征进行了分析,但未进行数据融合系统设计,

主要侧重于环境参数规律分析。

综上,目前临近空间环境研究主要侧重于数据处理的理论研究,且对单一数据源数据处理研究较为普遍,未进行海量多源临近空间环境数据处理的工程化应用。为此,本文基于临近空间多源环境数据采集、存储、组织和管理需求,设计了一种临近空间多源环境数据融合系统。

1 系统设计

1.1 总体设计

临近空间多源环境数据融合系统总体架构如图 1 所示。临近空间多源环境数据融合系统分为 4 个层级,按照从下往上的支撑关系,分别是基础设施层、数据服务层、算法服务层、展现层。

- (1)基础设施层:为系统提供计算、存储、网络和安全等资源支持;
- (2)数据服务层:提供数据存储、组织、管理等服务,实现多源临近空间环境数据高效存储和管理;
- (3)算法服务层:提供基于机器学习和深度学习算法服务,对多源环境数据进行处理和融合;
- (4)展现层:提供用户界面和交互设计,将算法服务层的结果以可视化的形式呈现给用户,包括图表、报表、界面和 API 等。

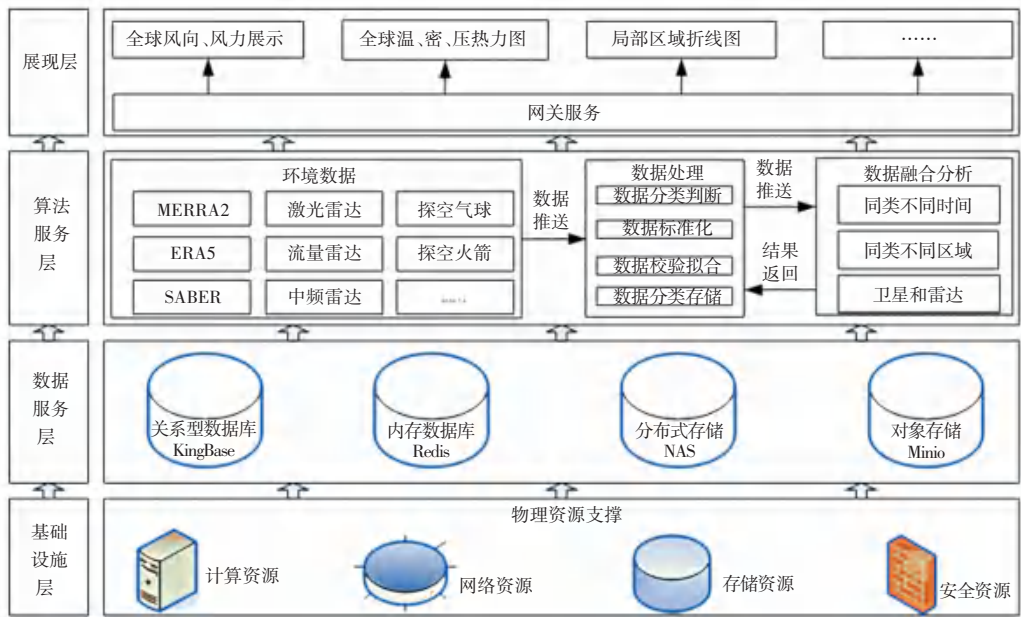


图 1 临近空间多源环境数据融合系统总体架构

Fig. 1 Overall architecture diagram of multi-source environmental data fusion system in near space

1.2 系统详细设计

1.2.1 基础设施层

基础设施层包括 4 种资源：

- (1)计算资源:包括 CPU、GPU、内存等资源,用于运行各种应用程序和服务,操作系统为国产化中科方德操作系统 SVS2.16.2;

- (2)网络资源:包括网络设备和网络带宽,用于连接不同的计算资源和用户;
- (3)存储资源:包括硬盘、闪存和其他存储设备,用于存储各种数据和应用程序;
- (4)安全资源:包括安全设备、安全协议和安全策略,用于保护基础设施、数据和应用程序免受各种安全威胁。

1.2.2 数据服务层

数据服务层提供数据存储、组织、管理等服务,不同种类数据采用不同存储介质进行存储,使得系统可高效存储与检索输出:

- (1)关系型数据库采用国产化人大金仓数据库 KingbaseV8.7,用于存储结构化数据,包括系统管理数据、临近空间环境数据元数据信息、日志数据等^[9-10]。
- 为加强日志审计,将系统登录日志数据和操作日志数据单独存储,因此临近空间多源环境数据融合系统结构化数据采用两套主从集群,分别存储临近空间多源环境数据元数据、系统管理数据以及登录日志数据和操作日志数据。关系型数据库集群设计图如图 2 所示。

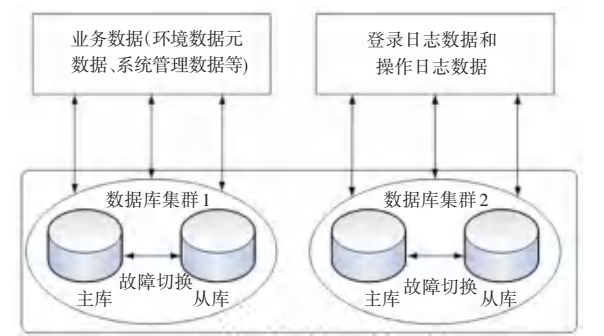


图 2 关系型数据库集群设计

- (2)内存数据库采用开源内存数据库 Redis,用于存储快速读写数据、缓存数据,实现高并发、快速展示等应用场景。
- 内存数据库采用 3 个节点组成的主从复制的结构,充分保障系统的安全性和可靠性。内存数据库集群设计图如图 3 所示。
- (3)对象存储服务采用开源对象存储 MinIO,用于存储临近空间环境原始数据、中间结果数据、应用数据等非结构化文件数据^[11]。
- 对象存储服务采用联邦模式进行架构设计。该模式为可将多个 MinIO 分布式集群在逻辑上组成一个联邦,对外以一个整体提供服务,并提供统一的命名空间,可实现存储容量扩充,实现非结构化数据的存储可靠性和安全性,联邦模式对象存储集群设计

如图 4 所示。

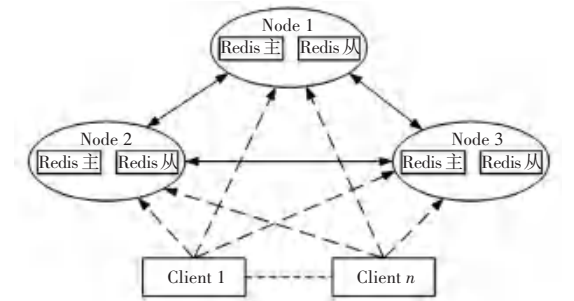


图 3 内存数据库集群技术设计

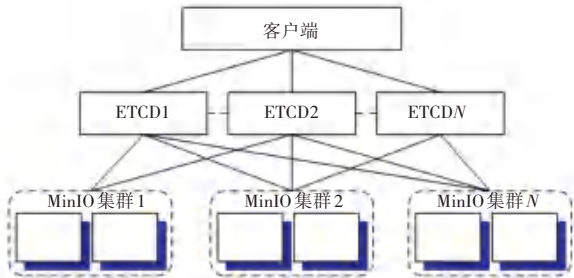


图 4 联邦模式对象存储集群设计

1.2.3 算法服务层

- 算法服务层采用 SpringCloud 微服务架构,实现基于模型算法的数据处理、分析和融合服务,系统微服务架构设计如图 5 所示。
- 数据处理流程总体上包含如下步骤:
- 1)数据模型建立
- 对临近空间多源环境数据进行定义,包括数据判断条件、元数据参数、数据处理方式等,对不同的临近空间环境数据,包括 Merra2、Era5、Saber、激光雷达、流星雷达、中频雷达等环境探测数据,可扩展的 Json 格式建立数据模型,以实现动态描述采集的数据。
- 2)数据标准化
- 为便于临近空间多源环境数据统一处理,对临近空间环境数据进行异常值校验、缺失值拟合、数据空间插值、高度气压转换,以实现数据标准化。
- 3)数据分类存储
- 为便于不同格式数据高效管理,将原始数据、标准化数据等非结构化数据存储到对象存储中,对元数据信息、管理数据等结构化数据存储到关系型数据库中。数据分类存储可提升系统的数据管理能力,实现系统高扩展性和安全性。
- 4)数据融合分析
- 为提升临近空间多源环境数据精度和准确度以

及应用价值,将标准化的临近空间多源环境数据进行同类不同时间、同类不同区域、卫星和雷达数据等融合分析,并将融合分析结果形成标准 API 供展现层进行数据调用,完成可视化展示。

1.2.4 展现层

基于 Vue、Cesium 等技术进行前端微服务化,完成数据服务层数据内容以及算法服务层中算法分析和计算的结果多维可视化。

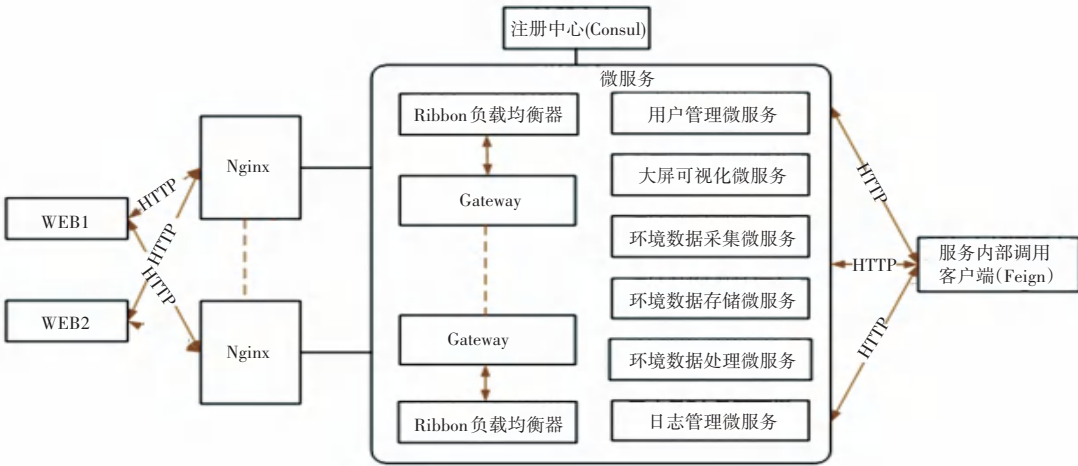


图 5 系统微服务架构设计

Fig. 5 Design of system microservices architecture

2 数据处理方法

由于系统应用在内部局域网络中,临近空间多源环境数据主要通过光盘拷贝方式导入,系统自动监测内部网络服务器指定目录下的新增数据文件,完成自动数据采集和上传,再进行数据处理。

2.1 卫星数据处理方法

为完成临近空间海量卫星数据管理以及应用,实现卫星数据(主要为小文件)自动化采集、处理、存储和管理,需对卫星数据进行预处理。

1) 卫星数据判断: 建立的数据模型,结合卫星数据文件类型和卫星文件的信息,完成卫星数据类型区分。

2) 数据标准化: 不同卫星文件的计量单位有差异,因此需要同一单位,例如温度统一用 K 表示,大气密度统一用 kg/m^3 ; 卫星数据分层级探测,存在数据缺失、数据不完整的情况,要对部分缺失的数据进行拟合和校验,来实现数据的完备性。

3) 数据存储: 为了统一管理标准化的卫星数据,对处理后的卫星数据重新分类命名,并将结果文件数据存储在对对象存储 MinIO 中,元数据信息存储在关系型数据库中,实现卫星数据存储和管理。

2.2 雷达数据处理方法

为完成临近空间海量雷达数据管理以及应用,实现雷达数据(主要为小文件)自动化采集、处理、存储和管理,需要对雷达数据进行预处理。

1) 雷达数据判断: 通过建立的数据模型,结合雷达数据文件类型和雷达数据文件里的信息,完成雷达数据类型区分。

2) 数据标准化: 不同雷达数据文件的计量单位有差异,因此需要统一单位,例如温度统一用 K 表示,大气密度统一用 kg/m^3 。

3) 数据校验拟合: 因为雷达探测仪自身的局限性,探测存在数据缺失、数据不完整的情况,要对部分缺失的数据进行拟合和校验,实现数据的完备性。主要包括以下检查内容:

(1) 参量检查

对激光雷达原始数据中的时间、高度、参量信息进行检查,剔除格式错误或信息不全的数据,检查每个探测廓线的高度信息是否是递增,剔除高度跳变点以上部分数据;对原始数据中的缺测值进行识别,将缺测数据统一标记为 NaN。

(2) 极值检查

对于温度数据,控制温度的范围为 $100\sim 300\text{ K}$,不在此范围内的温度观测值判定无效;对于密度数据,控制密度的范围为 $10^{-6}\sim 10^{-0.5}\text{ kg}/\text{m}^3$,不在此范围内的密度观测值判定无效,标记为 NaN。

(3) 时间一致性检查

原始数据不同大气廓线,可能存在不同的高度分辨率和时间分辨率,通过插值将原始探测廓线的高度分辨率统一为温度和密度廓线中最大高度分辨率,通过对温度和密度分别进行小时平均统一时间分辨率,

以温度廓线的时间为基准,匹配相同时间的密度廓线,对于时间未匹配成功的廓线予以剔除。

(4)物理一致性检查

基于时间一致性检查后的密度和温度数据,利用理想气体状态方程计算出气压廓线。气压必须随高度的减小而单调增加,不符合要求的高度以上部分数据判定无效,同时温度、密度参量在高度也一并判定无效。

4)数据存储:将雷达数据文件存储于对象存储 MinIO 中,雷达数据元数据信息存储在关系型数据库中,实现雷达数据存储和管理。

2.3 数据融合与分析方法

通过对卫星数据处理和雷达数据处理,完成了多源环境数据的标准化,在对象存储中存储卫星数据的原始数据、标准化中间数据,其格式为标准格式 .npy,在关系型数据库中存储卫星数据原始数据和标准化中间数据的元数据信息,同时存储了雷达数据的原始数据、标准化中间数据,在关系型数据库中存储雷达数据原始数据和标准化中间数据的元数据信息。

对卫星数据建立日期、类型、层级等索引结构,针对雷达数据建立日期、类型、经度、纬度、高度等时空索引结构,通过统一标准化的索引结构,支持快速完成数据查找、对比和分析。按照数据趋势,检测数据异常情况,实现数据融合使用与相互对照,提升数据质量,实现数据精度和准确度提高。

3 环境特性分析方法

3.1 全球环境特性分析

卫星数据支持从宏观、局部层面对全球临近空间环境特性分析,展现方式为三维、二维彩色图形,达到全面直观的了解临近空间环境特性,全球临近空间环境特性分析如图 6 所示。支持选择不同高度,采用不同颜色以三维球方式展示全球环境特性,可准确反映出临近空间环境特性区域分布差异。

3.2 区域环境特性分析

根据雷达不同时期的数据,可生成折线图、等值线图,用于同比和环比的对比分析。数据可视化可以帮助用户更直观地了解雷达数据在不同时间段的变化趋势和变化幅度,发现其中的规律和趋势,实现临近空间区域环境特性分析。

2021 年均值温度折线图对比如图 7 所示,反映出 2021 年极值、均值与 USSA-1976 大气模型随高度的温度趋势吻合;2021 年 6 月份温度数据极值分

布对比如图 8 所示,可见 6 月份临近空间随高度的温度振荡变化比较大,同值振荡变化也比较大,在 43 km 高度处,极大值相对于 USSA-1976 大气模型最大振幅达到 8.3%。

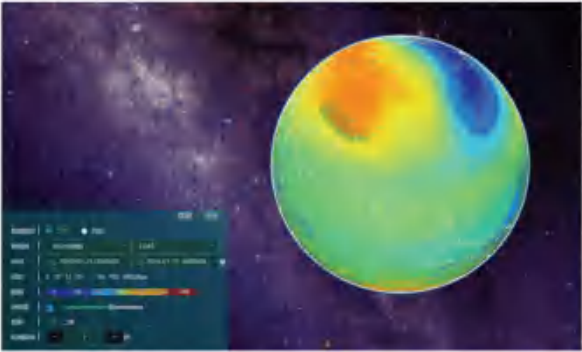


图 6 全球临近空间环境特性分析(三维温度图)

Fig. 6 Analysis of global near space environmental characteristics (3D temperature map)

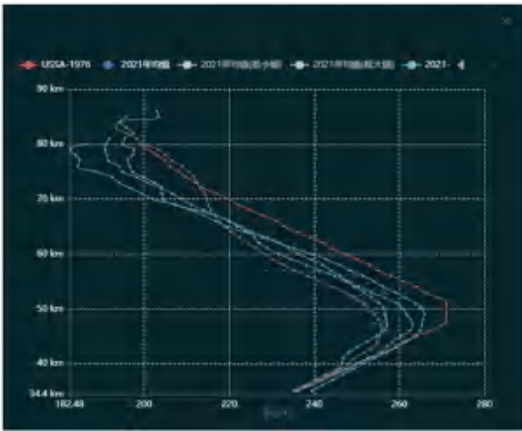


图 7 2021 年均值温度折线对比图

Fig. 7 2021 mean temperature line comparison chart

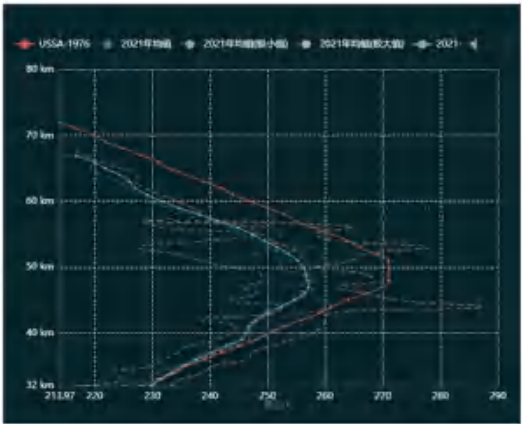


图 8 2021 年 6 月份温度数据极值分布对比图

Fig. 8 June 2021 temperature extremes distribution comparison chart

多月份温度、密度等值线对比如图 9 所示,直

观反映出多月份探测站点的温度、密度等值线分布情况。

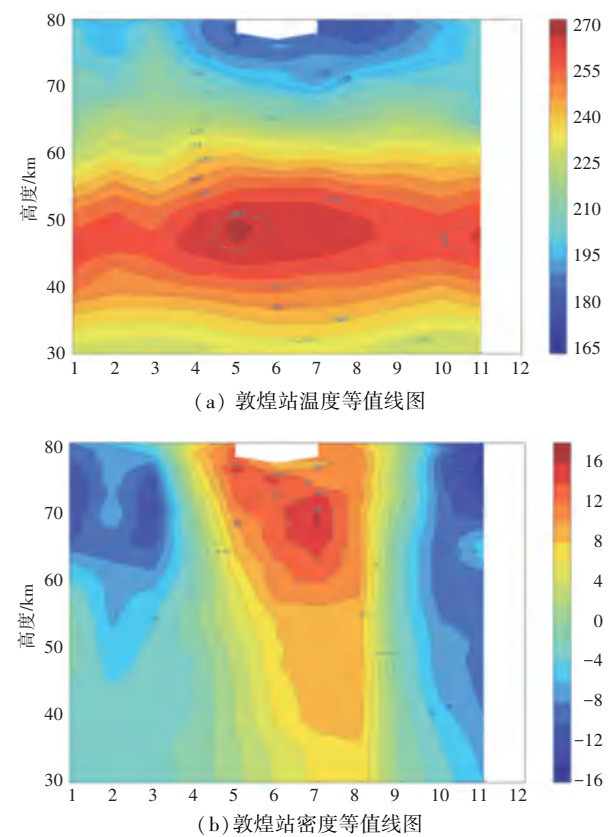


图 9 多月份温度、密度等值线图对比

Fig. 9 Comparison of temperature and density contour maps for multiple months

通过对雷达数据多维分析和可视化,可以更好地了解雷达数据在时间和空间上的变化情况,从而更好地掌握气象变化的趋势和规律,预测未来的气象变化趋势,为相关领域决策提供参考和支持。

3.3 卫星和雷达数据对比分析

在临近空间多源环境数据融合系统中,选择需要分析的雷达数据日期,根据站点经纬度,查找该日期下与站点经纬度相近的卫星数据,完成数据对比分析,如图 10 所示。图 10(a) 为温度数据对比分析,横坐标为日平均温度,纵坐标为高度,卫星探测温度和站点探测温度廓线实现卫星和雷达探测温度数据对比,可以看出该站点高度为 30~60 km 范围内卫星所测临近空间温度数据与雷达所测临近空间温度数据偏差较小,数据准确度较高。图 10(b) 为密度数据对比分析,横坐标为日平均密度采用对数坐标轴,纵坐标为高度,卫星探测密度和站点雷达探测密度廓线实现卫星和雷达探测密度数据对比,可以看出该站点卫星所测临近空间密度数据与雷达所测密度数据在 30~70 km 范围内高度重合,数据准

确性较好,准确度可达 99%,可直接使用两种数据。

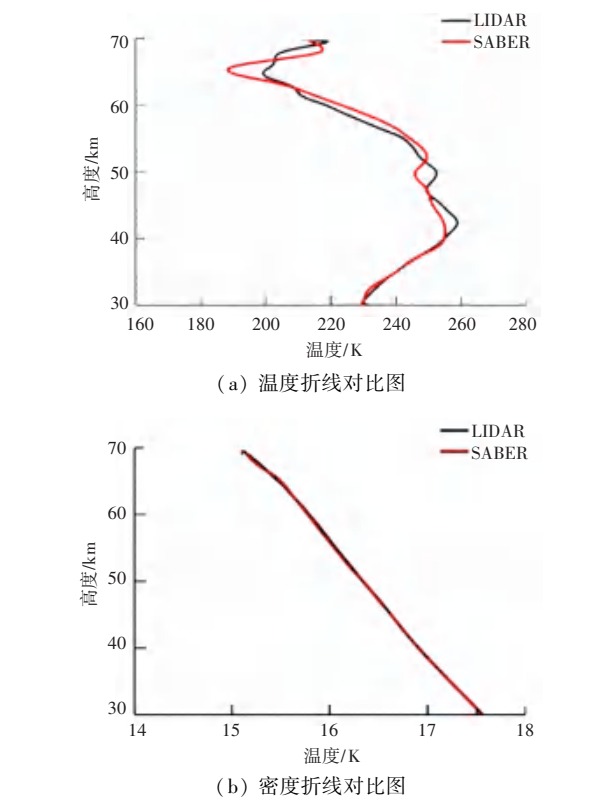


图 10 卫星数据和激光雷达温度、密度折线图对比

Fig. 10 Comparison of temperature and density line plots between satellite data and LiDAR

4 结束语

面向临近空间卫星、雷达等多源环境数据,本文设计了一种临近空间多源环境数据融合系统,基于该系统重点对卫星、雷达数据进行了不同的处理,并实现了数据的融合与分析,以及效果展示。

(1) 基于时空大数据技术、先进可视化技术,建立了分层的临近空间多源环境数据融合系统。该系统采用国产化操作系统、国产化数据库以及开源技术,支持海量临近空间多源环境数据采集、存储、管理,具有较高安全性和扩充性,截止目前系统已存储数据达 20.8 T。

(2) 随着数据种类日益增多,临近空间环境探测数据在系统中采用分别建立数据模型的方法,实现数据种类扩充,方便数据处理和管理;针对卫星、雷达、探空气球、探空火箭等临近空间环境数据已建立 20 种数据模型;将数据整合为统一格式的中间数据,可更好地应对不断增长的数据量和多样化的应用场景,支撑不断变化的用户需求。

(3) 系统支持自动完成卫星、雷达等探测数据融合分析,卫星数据适合空间维度分析临近空间全

球环境特性,雷达数据适合时间维度分析临近空间区域环境特性,但对于特定位置、日期两种数据可以相互验证、对比和分析,提升数据精度和准确度。在 30~70 km 范围高度内,通过对比分析发现卫星探测和雷达探测密度数据高度重合,数据准确度可达 99%。

本次系统设计主要关注的是临近空间历史数据融合处理,未来将基于历史数据进行数值气象预报研究,通过挖掘历史数据预测未来临近空间的环境变化情况,为临近空间应用提供环境参考。

参考文献

[1] 吕达仁,陈泽宇,郭霞,等. 临近空间大气环境研究现状[J]. 力学进展, 2009, 39(6): 674-682.

[2] 马伟,宣益民,韩玉阁. 临近空间热环境分析及低速飞行器的热设计方法[J]. 宇航学报, 2010, 31(5): 1272-1277.

[3] 邓潘. 地基激光雷达对临近空间大气的探测研究[D]. 合肥: 中国科学技术大学, 2018.

[4] 郝晓静,王立斌,马玉,等. 流星雷达与中频雷达探测 70~100 km 临近空间水平风场对比分析[J]. 空天技术, 2023(1): 43-52.

[5] 冯超,邵胜利,王芳栋,等. 探空火箭在临近空间气象保障中的应用研究[J]. 测控技术, 2018, 37(5): 16-19.

[6] 王梓皓,姜毅,黄宛宁,等. 基于气球平台的临近空间环境原位探测技术研究[J]. 空天技术, 2023(4): 70-79.

[7] 介阳阳,赵国庆,叶君好,等. 基于 SABER 卫星数据的临近空间大气参量分析[J]. 气象水文海洋仪器, 2018, 35(4): 108-114.

[8] 宋平. 基于卫星、火箭和气球探测资料的我国临近空间大气环境特征分析[D]. 长沙: 国防科技大学, 2022.

[9] 郑善双. 人大金仓: 从传统数据库到新兴大数据[J]. 软件和集成电路, 2017(6): 82-85.

[10] 罗华群. 金仓数据库 KingbaseES 在政府电子政务领域中应用[J]. 电子技术与软件工程, 2019(6): 145-147.

[11] 邱宇. 基于 MinIO 分布式存储的微服务模块开发应用[J]. 互联网周刊, 2023(22): 38-40.