

特色专题

遥感多维数据格式研究进展

张立福^{1,2}, 张赛^{1,3}, 孙雪剑^{1,2*}, 赵淑馨^{1,3}, 黄长平^{1,2}, 高宇^{1,3}, 童庆禧^{1,2}

摘要 综述了遥感多维数据格式(特别是多维时空谱数据)的国内外发展情况、最新研究成果、当前技术难点以及未来发展方向。介绍了中国科学院空天信息创新研究院在国际上首次提出的多维数据格式及其多维遥感数据综合与表征的理论和体系,促进了国内在数据组织方面的发展,并对国际研究产生了积极影响。随着遥感技术的快速发展,多维时空谱遥感数据不断涌现。旨在综述遥感多维数据格式的国内外发展情况,阐述最新研究成果,剖析当前技术难点,并对未来发展方向进行展望,为相关领域研究人员提供全面的参考。

关键词 遥感多维数据格式; 多维数据集; 时空谱数据; 数据组织; 理论技术体系

遥感是全球变化研究最重要的技术手段之一,已从军事和科研扩展到广泛的民用领域,提供实时、大范围的地理信息。随着技术的进步,遥感数据的分辨率、覆盖范围在不断提升,通过遥感卫星长时间连续地观测,积累了大量长时间序列、多分辨率、多光谱系列数据。例如, Landsat TM 和 MODIS 遥感观测数据已经成为全球变化研究重要的遥感数据源。Landsat 长时间序列数据被广泛应用于森林扰动和恢复的趋势变化监测^[1]、森林扰动历史数据自动化重建^[2]、地表森林生物量变化定量分析^[3],以及土地覆盖变化监测^[4-5]等研究; MODIS 长时间序列数据被大量应用于农作物物候监测^[6-7]等研究。目前用于时间序列分析的遥感数据集通常是以一种常用格式的文件为基本

存储单元^[8]。多维数据的获取与应用成为研究热点,对遥感数据的组织和存储提出了更高要求。

目前常见的遥感数据格式包括分层数据格式-地球观测系统(hierarchical data format-earth observing system, HDF-EOS)、网络通用数据格式(network common data form, NetCDF)、地理参考标签图像文件格式(georeferenced tagged image file format, GeoTIFF)以及遥感商用软件自带格式(如 ENVI 和 ERDAS)。目前缺乏完全由中国原创且与国际完全不同体系的遥感数据格式,并且现有格式在一定程度上存在局限性。在传统研究中,使用三维立方体模型存储长时序遥感数据会将不同时相的数据单独存储,难以对其时间维度实现高效存储、快速检索及深度分析,从而很难充分探究不同地物随时间变化的特点。在此背景下,中国科学院空天信息创新研究院张立福团队基于遥感多时相、多空间、多光谱等数据特点,在国际上首次提出时空谱多维数据格式(multi-dimensional dataset, MDD),为遥感数据的组织和存储带来了新的突破,提高了长时序遥感数据的利用价值。为便于多

1. 中国科学院空天信息创新研究院遥感科学国家重点实验室, 北京 100101
2. 中国科学院空天信息创新研究院遥感卫星应用国家工程研究中心, 北京 100101
3. 中国科学院大学资源与环境学院, 北京 100101

收稿日期: 2025-02-25; 修回日期: 2025-04-08

基金项目: 深地国家科技重大专项(2024ZD1002100); 国家自然科学基金重点项目(41830108)

作者简介: 张立福, 研究员, 研究方向为高光谱遥感机理及多学科应用, 电子信箱: zhanglf@aircas.ac.cn; 孙雪剑(通信作者), 副研究员, 研究方向为高光谱数据融合, 电子信箱: sunxj201494@aircas.ac.cn

引用格式: 张立福, 张赛, 孙雪剑, 等. 遥感多维数据格式研究进展[J]. 科技导报, 2025, 43(18): 67-76; doi:10.3981/j.issn.1000-7857.2025.02.00227

维时-空-谱数据的构建读取和分析,本研究配套研发并出版了多维分析(multi-dimensional data analysis, MDA)软件模块,该模块集成于多维遥感分析(multi-dimensional analysis of remote sensing, MARS)软件中。

本文围绕遥感多维数据格式展开论述,探讨其国内外发展情况、最新研究成果和应用、业内知名专家的评价、技术难点以及未来发展方向,旨在将这一创新成果传递给更广泛的读者和科技爱好者,共同推动遥感科学与技术的发展和应用。

1 研究意义

多维数据格式通过全方位地记录并一体化组织地球表面的各种信息,包括不同时间的变化、不同空间位置的特征以及光谱所反映的物质成分等,可以构建更完整、精确的地球系统模型,有助于深入研究地球的大气、陆地、海洋等各圈层之间的相互作用和物质能量循环。同时,多维数据格式解决了现有数据格式以及传统时间序列变化检测存在的部分问题。

1) 解决数据格式兼容性问题。

随着遥感技术的发展,多种卫星平台和传感器不断涌现,各自采用不同的数据格式。例如,Landsat 系列卫星与 Sentinel 系列卫星的数据格式就存在差异。在多源数据融合时,这种格式差异使得数据难以直接整合和共享。不同数据格式在数据组织方式、编码规则以及元数据描述等方面各不相同。随着新的遥感应用需求的出现,不断有新的数据格式被提出或修改,进一步加剧了格式兼容性问题的复杂性。因此,实现对多源长时间序列遥感数据的统一管理,对科学研究是极为重要的。

2) 传统遥感时间序列变化检测仍存在问题。

传统的遥感时间序列变化检测存在以下问题:(1) 在时间维上,限于可显式表达的、具有物理意义的物候特征(如生长季开始时间、生长期长度等),对时序特征的挖掘较浅,同时对各个时间点的一系列离散图像数据分别进行处理,输出多个不同时间点的离散光谱立方体数据,仅针对单一像元的时间序列数据进行分析,忽略了像元邻域的空间信息;(2) 在空间维上,前述的算法大多是以单个像元作为研究对象,仅

对单个像元的时间序列曲线进行分析,没有考虑像元之间的空间相关性,部分融合空间特征的算法对空间信息的利用依然很少^[9];究其原因,主要是遥感时序数据包含了时、空、谱 3 个维度的信息,然而对其存储和处理的方式仍然是将其分解为多个三维数据来进行操作,仍是一系列不同时相离散的三维立方体数据,没有对遥感时序数据的四维信息进行统一的管理和分析^[10]。

2 国内外研究进展

2.1 国外发展现状

国际上,遥感数据格式种类繁多,如 HDF-EOS、NetCDF、GeoTIFF 等,广泛应用于不同遥感平台和传感器数据。国际上常见的遥感数据格式虽然在各自的应用领域有一定的优势,但也存在局限性。这些格式在数据组织和存储方面,没有对不同应用场景进行详细的划分,导致在一些特定的应用中,难以满足针对性的需求。

2.1.1 HDF-EOS

HDF-EOS 数据格式是美国伊利诺伊大学国家超级计算应用中心(National Center for Super Computing Applications, NCSA)于 1987 年研制开发的一种软件和函数库,主要用来存储由不同计算机平台产生的各种类型科学数据,适用于多种计算机平台,易于扩展。HDF-EOS 专门用于处理 EOS 产品,使用标准 HDF 数据类型定义了点、条带、栅格 3 种特殊数据类型,并引入了元数据(metadata)。其核心存储原理围绕分层数据管理、地理空间数据类型标准化和元数据自描述性展开。在存储结构上,HDF-EOS 采用树状分层模型,将数据组织为“组”(group)和“数据集”(dataset)的嵌套层级。

HDF-EOS 在遥感数据处理中展现的核心优势在于其多源数据整合能力与地理空间标准化支持。对于卫星遥感任务(如 Landsat、Terra/Aqua),单一过境周期可能产生数 10 个波段的光谱数据、多个角度的观测几何参数,以及大气校正所需的辅助数据。HDF-EOS 通过分层存储将这些数据封装在单一文件中,避免了传统分散存储导致的文件管理混乱。HDF 不断发展,已被广泛应用于环境科学、地球科学、航空、海

洋、生物等许多领域,虽然 HDF-EOS 具有自我描述、多样性、灵活性、可扩展性和独立性等特点,可提供系统宽搜索服务方式并有效存储地理定位数据,但它并非完全由中国原创,在与国际体系的差异方面,其起源于美国,中国在该格式的发展中缺乏主导权。

尽管 HDF-EOS 功能强大,但其在应用中仍面临显著挑战。首先是复杂性带来的技术门槛,用户需同时掌握 HDF5 数据模型、EOS 扩展类型定义和地理编码规则。存储效率问题在大规模遥感场景中凸显,强制嵌入的元数据(尤其是 XML 格式的全局元数据)可能导致文件体积膨胀,而 HDF5 的块状存储机制(chunking)在未合理配置时,会使高频访问小范围数据的效率低下。

2.1.2 NetCDF

NetCDF 是由美国大学大气研究协会的 Unidata 项目科学家针对科学数据的特点开发的,是一种面向数组型数据的科学数据格式,可以有效存储多维度、多变量的遥感数据,并支持数据的压缩、子集提取等操作。NetCDF 采用分层结构组织数据,其核心由文件头和数据体 2 部分构成。文件头定义了数据的维度(dimensions)、变量(variables)和属性(attributes),形成逻辑网络以描述数据关系。NetCDF 以维度为纽带整合时空数据,这种结构天然适配遥感数据的时空特性。它广泛应用于大气科学、水文、海洋学、环境模拟、地球物理等诸多领域。

NetCDF 能够存储多维数组数据,如全球不同地理位置气象站在不同时间点的气象要素(温度、湿度、风速等),这种格式可以方便地表示和存储气象数据的时空分布特性。但是 NetCDF 格式通常依赖于复杂的数据结构,如树或链表来组织信息,并且建立在三维立方体模型上,难以用于长时间的数据序列分析。其复杂的数据模型和长时序处理瓶颈仍制约着超大规模遥感应用。

2.1.3 GeoTIFF

GeoTIFF 是在 TIFF(tagged image file format)基础上扩展而来的。它在 TIFF 的文件头(header)和图像数据之间添加了地理参考信息,其存储原理围绕标签化元数据和多波段图像封装展开。文件头中包含了图像的基本信息,如尺寸、颜色模式等,而地理参考

信息则包括了图像的坐标系统、地理变换参数等内容。GeoTIFF 可以存储单波段或多波段的遥感图像数据,通过其中的地理参考信息,使用者能够方便地将图像数据与实际地理空间位置关联起来。

在遥感领域,GeoTIFF 的核心优势在于地理空间直读性与广泛兼容性。例如, Sentinel-2 卫星的 L2A 级地表反射率产品以 GeoTIFF 格式分发,每个波段(如 B2 蓝光、B8 近红外)存储为独立文件,并通过仿射变换参数实现 10 m、20 m、60 m 多分辨率数据的快速定位。但是在处理复杂的多维时空谱遥感数据时,GeoTIFF 具有一些局限性。例如,对于包含多个时间序列、多种光谱波段和不同空间分辨率的数据,其组织方式可能会比较繁琐。虽然 GeoTIFF 允许通过自定义标签添加元数据(如传感器参数、大气校正系数),但其标签体系缺乏结构化层次。对三维(如大气垂直剖面)或四维数据(时间×高度×纬度×经度)缺乏原生支持。例如,全球气候模型输出的温度场(含 10 个垂直层)需拆分为 10 个独立的 GeoTIFF 文件。同时,由于 GeoTIFF 主要侧重于图像数据和地理参考信息的存储,对于添加新的数据类型(如传感器的其他辅助数据)的灵活性相对较低。

2.2 国内发展现状

中国现有的常见遥感数据格式在原创性以及与国际体系差异方面存在一定的局限性,这也促使中国科学家积极探索和研发具有完全自主知识产权且与国际不同体系的遥感数据格式。中国在遥感数据处理与应用中对 HDF5 格式的优化改进,主要集中在元数据扩展、存储结构优化、性能提升和工具链适配等方面,以满足国产卫星数据的特殊需求并提升数据处理效率。例如,在 HDF5 文件中添加符合中国遥感数据规范的元数据字段,包括卫星标识(如 GF-7)、传感器类型(如 PMS 全色多光谱传感器)、数据生产信息(如中国资源卫星应用中心(CRESDA)、处理时间)以及辐射定标参数、几何校正系数等关键参数。

遥感数据具有多样性和复杂性,包括不同的分辨率、光谱波段、投影方式等。因此,设计能够兼容各种类型遥感数据的存储格式是一个巨大挑战。此外,遥感数据量巨大,如何高效地利用存储空间,避免浪费,是研发遥感存储格式时需要重点考虑的问题。

3 最新研究进展

在遥感多维数据格式方面,中国取得了重大的创新突破与丰硕成果。历经 10 余年潜心研究,在国际上首次提出多维数据格式 MDD,构建了多维遥感数据综合与表征的理论和技術体系,促进了国内在数据组织方面研究的发展,并获授权 PCT 发明专利,实现了中国在遥感数据格式领域专利的突破。目前,MDD 成为国家对地观测科学数据中心、全球变化数据出版系统认可的数据格式;研发了国际首个“遥感多维数据格式互操作分析软件系统”,在全球变化数据出版系统出版。该系统在“全球变化科学研究数据出版系统”正式出版^[1]。目前该软件系统已经更新至 MARS 2.4.1.0。同时,研究成果在多个行业用户单位进行广泛推广和应用,产生较好的社会效益和经济效益。此外,牵头编写的团体标准《遥感时空谱多维数据格式》被中国地理信息产业协会批准发布实施,成为国家对地观测科学数据中心、全球变化数据出版系统认可的数据格式。这一标准的制定,为中国遥感数据的组织和存储提供统一规范,推动行业的发展。

3.1 多维数据的组织与存储

由于不同传感器数据,其波段设置、光谱响应函

数、空间几何成像等不同,难以直接应用于时序分析,因此需要进行数据的重构:(1) 根据应用需求或目标,获取多源遥感数据;(2) 进行标准化预处理,即投影转换、几何纠正、拼接、裁剪等。在构建长时序遥感数据时,由于原始遥感数据不可避免地会受到天气、云及阴影等的影响,造成一些随机噪声。为了准确提取时序特征,一般需要对其进行插值、滤波和融合重构。常用的插值方法有:二次样本插值、三次样条插值、等间距样条插值等。常用的滤波方法有:Savitzky-Golay 滤波、小波变换、惠克特滤波等。

在数据层对多维数据集重构完成后,虽然在成像结果上保持一致,但是在时间维度上仍然是离散的,还需要将多维数据集在时间维度上进行整合。针对长时序遥感数据的组织和存储问题,MDD 数据集将多维数据的空间、时间和光谱构建成一个四维数据集,以此来实现对长时序数据的统一管理。

多维数据格式的头文件中记录了用于表达和解析影像数据体文件的所有元数据信息(表 1)。同时,多维数据结构可以对已生成的文件进行增加或者删减图像数据等操作。MDD 数据体文件是存储影像数据的文件,采用二进制的字节流。

表 1 头文件中包含的信息

字段	含义
Samples	数据每个波段所包含的列数
Lines	数据每个波段所包含的行数
Bands	数据所包含的波段数
Time	数据所包含的时间数
Header Offset	文件中存在的嵌入式头信息的字节数。读取文件时,会跳过这些字节
File Type	文件类型,如MDD Standard
Data Type	数据表示的类型: 1=8位字节; 2=16位有符号整数; 3=32位有符号长整数; 4=32位浮点数; 5=6位双精度浮点数; 6=2×32位复数,双重精度的实数对; 9=2×64位双精度复数,实数对双精度; 12=16位无符号整数; 13=32位无符号长整数; 14=64位有符号长整型; 15=64位无符号长整数
Interleave	说明数据存储所使用的结构,包括 TSB、TSP、TIB、TIP和TIS
Sensor Type	传感器类型,如 Landsat TM、SPOT、MODIS 等
Byte Order	字节顺序,指占内存多于一个字节类型的数据在内存中的存放顺序,Byte order=0表示先存储低字节(Least significant bit); byte order=1表示先存储高字节(Most significant bit)
Map Info	将文件的信息按照投影名称、参考像素位置(文件坐标)、经度、纬度、像素大小、投影区域和坐标系的顺序列出
Coordinate System String	坐标系字符串字段。列出了用于地理坐标系或投影坐标系的参数
Band Names	数据文件中各个波段的名称
Time Names	数据中各个文件获取时间

类似于传统遥感数据的存储, MDD 数据依照存储方式的差异, 可分为以下几种不同的类别: 时序波段连续存储(temporal sequential in band, TSB)、时序像元连续存储(temporal sequential in pixel, TSP)、波段时序交错存储(temporal interleaved by band, TIB)、像元时序交错存储(temporal interleaved by pixel, TIP)和光谱时序交错存储(temporal interleaved by spectrum, TIS)共 5 种存储方式, 分别如图 1~图 5 所示。

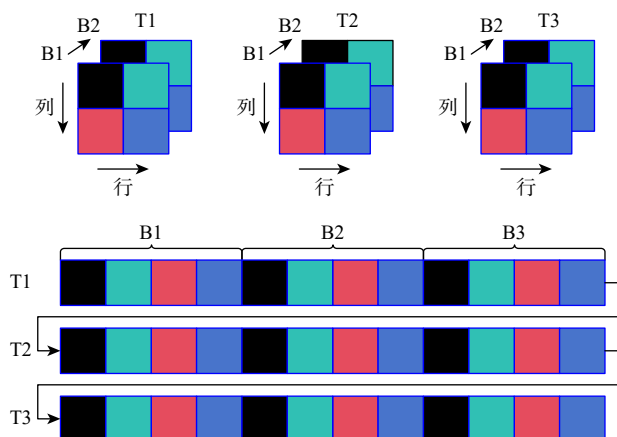


图 1 数据存储的 TSB 格式结构

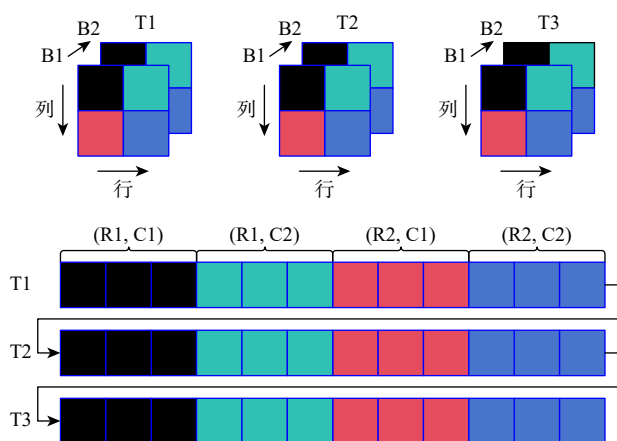


图 2 数据存储的 TSP 格式结构

在 TSB 的格式中, 从宏观角度上看, 是由一组组不同时序的光谱数据排列而成。在微观角度上看, 在每组光谱数据中, 先后顺序是依照波段序列进行的。最后, 不同波段内数据的布局是从行到列。这种组织一次将所有波段放在一起, 确保了整个波段的数据在存储空间中的连续性。因此, TSB 数据结构适用于对一个时间上或若干个波段的数据进行空间维度的处理, 可应用于提取一个或多个时间所有波段组成的

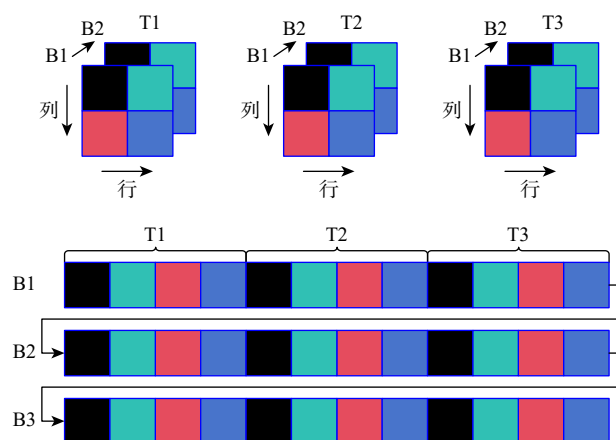


图 3 数据存储的 TIB 格式结构

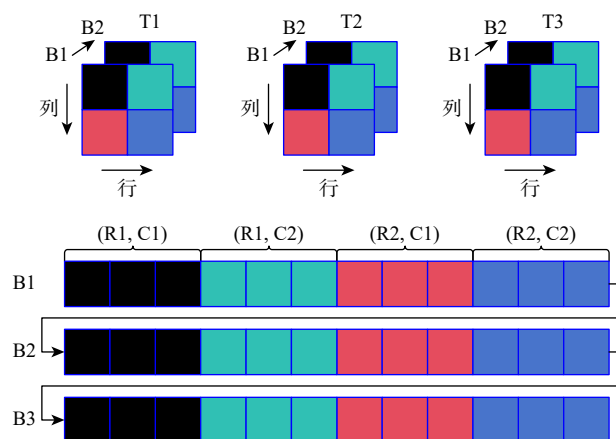


图 4 数据存储的 TIP 格式结构

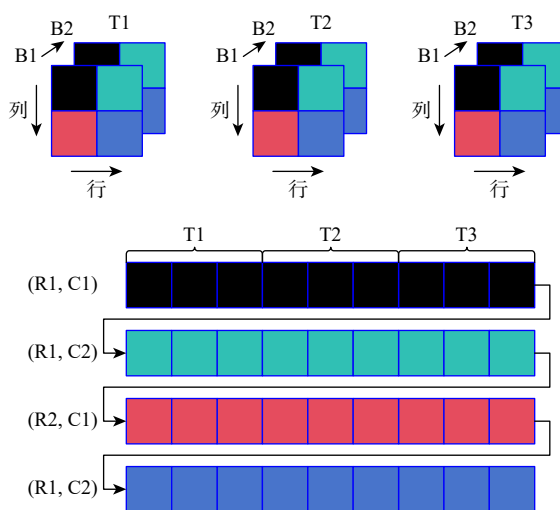


图 5 数据存储的 TIS 格式结构

光谱立方体数据, 方便进行光谱邻域运算和空间域滤波等。

与 TSB 相似, TSP 格式所有的波段也按光谱顺序排列。但与 TSB 不同的是, 像素首先按光谱顺序排

列,然后以 TSP 的格式从行到列的顺序排列。它汇集了像素的光谱数据,保证了光谱数据在存储空间中的连续性。因此,TSP 数据结构适用于对一个时间的光谱数据进行操作,可应用于提取一个像元或者一片区域的光谱曲线或者对不同时间的影像进行光谱特征化。

TIB 格式中将每个波段的整个时间序列按时间顺序放在一起,然后将每个波段的时间序列数据按光谱顺序排列。此外,每个波段中的所有像素都从一行排列到另一列。TIB 保证了存储空间中空间数据的连续性,同时保证了每个频段的时间的连续性。因此,TIB 数据结构适用于提取某个波段的时间序列立方体数据,可应用于提取一个波段的时间序列立方体,可以进行光谱时间序列分析或针对某个波段选择 3 个时间的数据假彩色合成显示。

TIP 格式将每个波段的每个像素的整个时间序列按时间顺序放在一起,然后在每个波段中,将像素的时间序列从行到列排列,最后将所有波段的数据按光谱顺序排列。TIP 格式将所有像素的时间谱放在一起,确保了存储空间中时间谱数据的连续性。它适用于对像元的时间谱进行处理与分析,可应用于提取一个像元或者一片区域在某个波段的时谱曲线,在时间维度进行平滑和滤波处理,对时谱曲线进行拟合,以及进行预测分析等。为了研究时空谱数据,探究不同农作物的植被指数在时间序列上的特点,本研究采用了 TIP 的结构来构建 MDD 数据集。

TIS 格式将每个像素的时间谱的整个时间序列按时间顺序组合在一起,然后在空间维度上从行到列排列出时间序列。TIS 保证了频谱的连续性,同时,它将频谱的时间序列排列在一起,因此,TIS 数据结构适用于提取像元光谱曲线的时间序列数据,可应用于提取某一个像元在一个时间范围内的所有光谱曲线并对这些曲线进行三维可视化,并分析该像元的光谱随着时间变化的特征^[8]。

MDD 数据格式具有以下特点。(1) 灵活性: MDD 对多维图像数据的组织有 5 种存储结构,5 种存储结构之间可以相互转换,用户可以根据数据分析的需求灵活选择对应的数据组织结构。(2) 多维性: MDD 是一种多维数据格式,允许用户将遥感数据时间、空间、光谱维度的信息集成到同一个数据集中。(3) 可扩展性: MDD 数据集,可根据需要,将新获取的图像

数据添加到已生成的 MDD 文件中,实现时间维度数据信息的扩展。(4) 完整性: MDD 数据除了包含遥感栅格图像数据本身之外,还附带描述该数据的所有属性的全面信息,包含时间、大小、地理投影等。

3.2 MDD 的计算机配套系统(MARS 软件)

为了使用户理解和应用.mdd 数据格式,研发了可用于.mdd 数据输入、存储、分析和输出的计算机软件模块,命名为 MDA 多维分析计算机软件模块,该模块集成于遥感多维数据格式互操作分析软件系统(简称 MARS 1.0)中。

MDA 是基于 MDD 多维数据格式开发的集多维数据构建、多维数据分析和多维数据可视化于一体的遥感数据时空谱多维分析软件模块,该软件模块包括 MDD 数据构建(MDD Builder)及打开(MDD Open)、格式转换(MDD Conversion)、数据导出(MDD Export)和数据运算(MDD Math)等功能,其具体功能如图 6 所示。

目前,MDA 模块已经集成了对原始 MODIS、原始 Landsat 和经处理为 ENVI 标准数据格式的遥感数据的 MDD 数据构建。利用 MDA 将原始 Landsat 数据构建为 MDD 数据的总体流程如图 7 所示。

3.3 时谱的定义

长时序遥感数据融合了时间、空间和光谱总共 3 个维度的信息特征。现有研究主要聚焦于空间和光谱维度的分析,尚未形成完整的概念体系来系统描述这类多维数据。传统遥感时序数据的存储和处理方式通常采用离散时间点管理,本质上仍停留在三维数据处理层面,未能实现对长时序数据四维特征的有机整合与深度挖掘。针对这一研究空白,张立福团队在继承前人研究成果的基础上创新性地提出了“时谱”理论^[10]。该理论借鉴光谱概念,将地物目标在不同时相的遥感特征序列定义为时谱特征。如图 8 所示,特定地物在时间序列上的时谱特征值变化轨迹被称作时谱曲线。

时谱特征是指从时谱曲线中提取的统计或时序特征(如最大值、最小值、均值、方差等),用于表征单个像元在某一波段上的时间动态变化规律及其光谱特性。多波段时谱曲线由多条单波段时谱曲线组合而成,共同描述同一地物在不同波段上的时间变化趋势。多波段时谱特征则是从多波段时谱曲线中提取

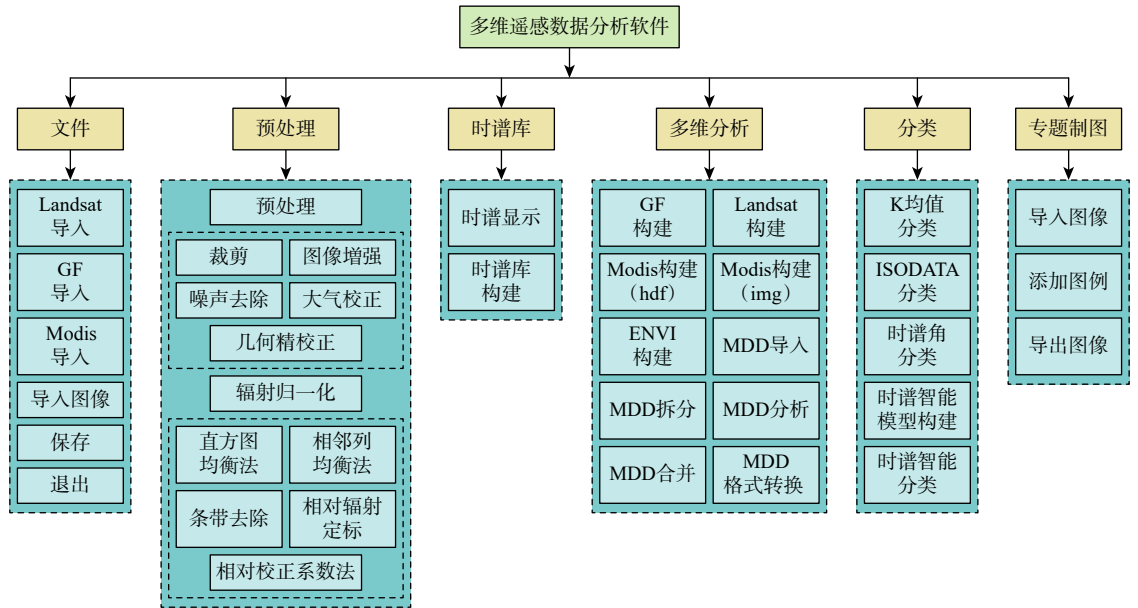


图 6 MARS 软件功能介绍

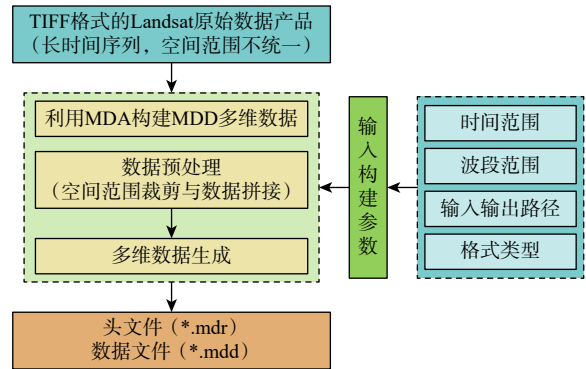


图 7 利用 MDA 模块以 Landsat 为数据源构建 MDD 的总体流程

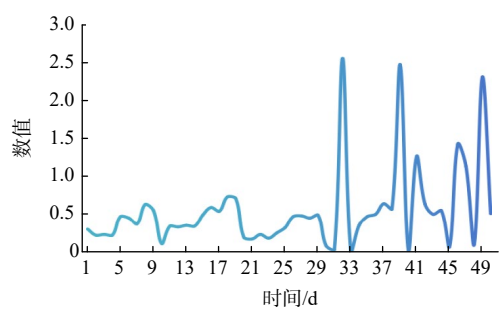


图 8 植被指数时谱曲线示例

的复合特征,能够更全面地刻画地物的时空-光谱联合变化模式。

3.4 遥感时谱理论的应用

在生态环境领域,遥感时间序列数据中包含了自然以及人为因素导致的森林组成、结构、功能变化以及其差异信息,能够有效提升对森林演替过程以及发

展态势的理解能力,基于时空谱信息的森林扰动检测具有更高的敏感度和精准性^[12]。同时,通过充分利用时空谱数据在时间维、空间维、光谱维技术的优势,在生态保护修复工程需求的基础上可形成星空地协同作用的技术框架^[13],有效提升生态保护修复的效果。

传统的农作物精细分类是基于单时期的遥感影像提取农作物的种植信息,但是这种方法在研究区的地物类型较为复杂时,分类效果会显著降低。因此,基于单一遥感影像的农作物精细分类存在不足。徐权^[14]选取了 GI、SRI、NDVI 和 NGRDI 4 种常用的植被指数作为光谱特征空间进行分析,构建了时间序列的无人机遥感影像,相比于单一时期无人机遥感影像的总体分类精度和 Kappa 系数,分别提高了 26.65% 和 0.3496,有效地解决了单一时期影像的弊端,进一步提高了农作物的分类精度和效率。

地表覆盖变化检测方法,通过对土地变化检测以及调查农作物种植情况,不仅可以为国家的农业政策及时提供资料,同时也可以使土地利用更加充分,促进人与自然之间的和谐关系。通过遥感手段实现土地变化检测,不同于单一对某个时相的数据进行处理和分析,这要求同时多个时相数据的基础上进行综合而统一的分析。区别于传统的三维数据集,在原有的三维数据集基础上增加了时间的维度,变成了四维数据。时谱的出现使地物有了一种不同的方式来表现其特点,在土地种植类型判断的问题上就有了别样

的方法和思路。基于时空谱的耕地类型分析也具有全物候的特点,可以很好地消除季节因素引起的伪变化。Lin 等^[15]提出了基于时谱特征的地表覆盖变化检

测方法,如图 9 所示。结果表明所提出方法的有效性和稳定性,在不影响定量分析遗漏率的情况下,该方法的错误检测显著减少。

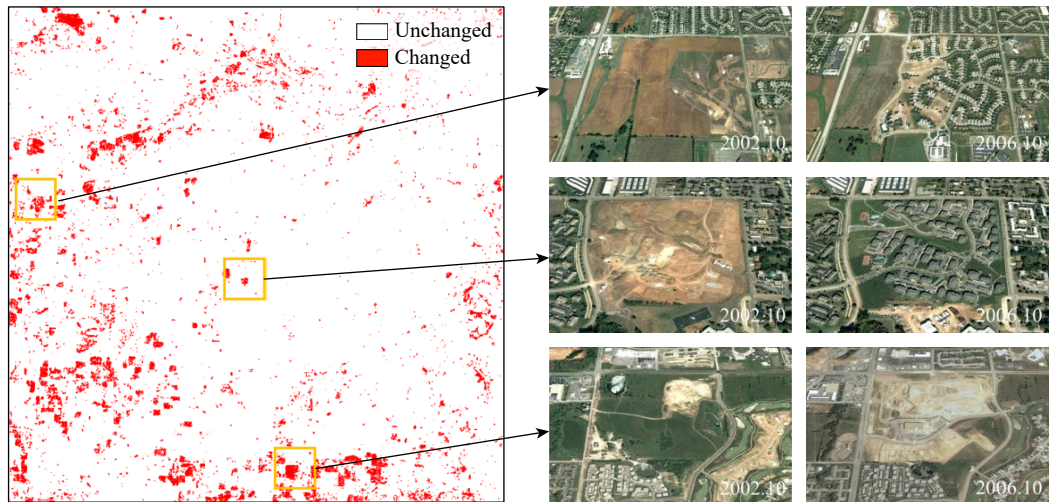


图 9 基于时空谱特征的变化检测方法

4 技术优势

遥感数据来源多样,格式各异,传统存储方式难以有效管理和处理。MDD 格式通过一体化存储时间、空间、光谱维度的信息,解决了数据格式不统一的问题,提高了数据管理的便捷性和效率。这种数据一体化组织及格式研究打破了传统学科界限。地理科学、生态学、气象学、地质学等多学科都依赖于遥感数据获取研究对象的信息。统一的数据格式和一体化组织方式使得不同学科的数据能够更好地共享和交互,促进了跨学科的研究合作。

多维时空谱数据一体化组织减少了数据整合过程中的冗余操作和数据转换的复杂性。传统的数据处理往往需要在不同的数据格式和存储结构之间频繁切换,耗费大量计算资源和时间。新的数据格式可将时空谱信息有机结合,在数据读取、预处理、特征提取等环节实现更高效的算法设计和计算流程优化。

统一数据格式是实现遥感数据广泛共享和互操作的基础。在国际合作、不同研究机构和企业之间的数据交流中,标准化的数据格式能够确保数据的无缝对接和正确解读。例如,全球范围内的环境监测项目,各国的遥感数据可按照统一的数据格式进行整合和共享,便于对全球环境变化进行协同研究和联合评估,避免因数据格式差异导致数据隔阂和信息孤岛

现象,从而提升整个遥感数据应用领域的协同创新能力。

5 未来展望

MDD 的提出加强了与国际组织和其他国家的合作,推动了遥感数据格式的标准化和互操作性,为遥感技术的国际交流与合作提供了新机遇。在环境保护领域,利用多维数据分析系统 MARS,可以进行水体悬浮物浓度指数的计算,监测水体污染和环境质量。通过对长时间序列遥感影像的分析,及时发现环境变化趋势,为环境保护部门制定政策和采取措施提供科学依据。在城市规划领域,建筑指数(building index, BI)可用于提取城市地区的建筑信息,反映城市化程度和用地利用情况。其帮助城市规划者合理布局城市空间,提高土地利用效率,改善城市居住环境。未来通过拓展 MDD 数据格式在农业、气象、环境监测、城市规划等领域的应用,可以推动遥感技术与 GIS、物联网等技术的融合,提升数据服务水平和应用效果。

MDD 格式为遥感分析带来效率和精度的显著提高,并在多个行业用户单位进行广泛推广和应用,产生较好的社会效益和经济效益。例如,中国科学院院

士傅伯杰团队^[16]完成的全球表层土壤水旬度数据集(RSSM, 2003—2018年)采取MDD数据格式存储,每年1个数据文件,共16个数据文件;赵恒谦等^[17]则基于MDD数据格式生产了空间分辨率约463 m,以16 d为间隔的2018—2021年间每年23个时相的中国陆域范围植被指数UNVI产品数据集。遥感学界对MDD的科学价值及其广阔的应用前景给予了充分的认可。2023年度国家最高科学技术奖获得者李德仁院士评价该成果为“这是国际遥感领域的理论和技术创新”;美国马里兰大学教授、IEEE Fellow Chein-I Chang教授评价“他们的研究突破了时间-空间-光谱信息融合的瓶颈,影响巨大”;IEEE Fellow、原IEEE TGRS主编Antonio J. Plaza评价“他们在多源遥感数据时空谱融合算法方面取得了世界领先地位,推动了该领域多维数据分析软件和技术的发展”,“这是一个非常重要的且备受期待的贡献,将对该领域的研究产生重大影响”。

此外,研发的国际首个“遥感多维数据格式互操作分析软件系统”,在全球变化数据出版系统出版后,被多个行业用户单位广泛应用。例如,在地质矿产勘探中,岩矿识别指数通过对遥感数据进行数学处理,帮助区分不同的岩矿类型,为寻找矿产资源提供有力支持;在森林防火和灾害监测中,火点指数用于检测火灾和火点,通过分析红外波段数据,可以迅速发现火灾灾情,为紧急响应和救援提供及时信息。

6 结论

MDD格式的提出促进了国内在遥感数据存储和组织方面研究的发展,并获得了国际PCT专利授权,有助于推动遥感技术的标准化和国际化进程,增强国际交流与合作。MDD格式及配套软件为遥感数据的时空谱多维联合分析提供了支持,显著提升了数据处理效率和准确性,有助于深入挖掘遥感数据中的信息,为科学研究和应用提供更有力的支持。未来,随着技术的不断创新和应用领域的不断拓展,遥感多维数据格式将发挥更加重要的作用,为人类社会发展和科学研究提供有力支持。近期,Google DeepMind提出的Alpha Earth Foundations(AEF),通过整合10余个不同来源的地球观测数据与文本数据,构建统一的

64维嵌入表征场,实现了跨模态、跨时空的语义一致性的数据融合^[18]。AEF似乎展示出了无限的应用潜力,但仍存在一些问题,例如其不同维度的信息并不具有特定可解释的物理意义。而张立福团队提出的MDD格式将多源异构数据融合后仍具有实际的物理意义,时间-空间-光谱的四维数据存储格式在实际问题分析中仍具有更强的可解释性。

参考文献(References)

- [1] Kennedy R E, Yang Z Q, Cohen W B. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr: Temporal segmentation algorithms[J]. *Remote Sensing of Environment*, 2010, 114(12): 2897–2910.
- [2] Huang C Q, Goward S N, Masek J G, et al. An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks[J]. *Remote Sensing of Environment*, 2010, 114(1): 183–198.
- [3] Powell S L, Cohen W B, Healey S P, et al. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches[J]. *Remote Sensing of Environment*, 2010, 114(5): 1053–1068.
- [4] Margono B A, Turubanova S, Zhuravleva I, et al. Mapping and monitoring deforestation and forest degradation in Sumatra (Indonesia) using Landsat time series data sets from 1990 to 2010[J]. *Environmental Research Letters*, 2012, 7(3): 034010.
- [5] Townsend P A, Helmers D P, Kingdon C C, et al. Changes in the extent of surface mining and reclamation in the Central Appalachians detected using a 1976–2006 Landsat time series[J]. *Remote Sensing of Environment*, 2009, 113(1): 62–72.
- [6] Galford G L, Mustard J F, Melillo J, et al. Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil[J]. *Remote Sensing of Environment*, 2008, 112(2): 576–587.
- [7] Sakamoto T, Yokozawa M, Toritani H, et al. A crop phenology detection method using time-series MODIS data[J]. *Remote Sensing of Environment*, 2005, 96(3/4): 366–374.
- [8] 张立福, 孙雪剑, 张霞, 等. 时空谱多维数据格式(MDD)结构与计算机配套系统[J]. *全球变化数据学报(中英文)*, 2017, 1(2): 121–135.
- [9] Li X C, Gong P, Liang L. A 30-year (1984–2013) record of annual urban dynamics of Beijing city derived from Landsat data[J]. *Remote Sensing of Environment*, 2015, 166: 78–90.

- [10] 张立福, 王飒, 刘华亮, 等. 从光谱到时谱: 遥感时间序列变化检测研究进展[J]. 武汉大学学报(信息科学版), 2021, 46(4): 451–468.
- [11] Zhang L F, Sun X J, Zhang X, et al. Inter-operational tool for temporal-spatial data analysis in Multi-Dimension Data Format (. mdd) [J/OL]. Global Change Research Data Publishing & Repository, 2017. doi: [10.3974/geodb.2017.02.20.V1](https://doi.org/10.3974/geodb.2017.02.20.V1).
- [12] 吴伶, 刘湘南, 刘美玲, 等. 融合遥感时间序列时空谱信息的森林扰动检测与归因研究进展[J]. 遥感学报, 2024, 28(3): 558–575.
- [13] 张丹丹, 肖晨超, 魏丹丹, 等. 基于时空谱的生态保护修复工程碳汇遥感监测技术框架研究[J]. 航天返回与遥感, 2022, 43(6): 119–128.
- [14] 徐权. 基于时间序列无人机影像的高精度农作物分类方法研究[D]. 石河子: 石河子大学, 2022.
- [15] Lin Y K, Zhang L F, Wang N. A new time series change detection method for landsat land use and land cover change[C]//Proceedings of 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp). IEEE, 2019.
- [16] 陈永喆, 冯晓明, 傅伯杰. 全球表层土壤水旬度数据集(RSSM, 2003–2018)[J/DB/OL]. 全球变化数据仓储电子杂志(中英文), 2021. <https://doi.org/10.3974/geodb.2021.02.01.V1>.
- [17] 赵恒谦, 刘轩琦, 张立福, 等. 中国陆域植被指数UNVI多维数据产品(2018—2021)[J]. 全球变化数据学报(中英文), 2022, 6(4): 645–655.
- [18] 秦其明. AlphaEarth Foundations: 遥感基础大模型的潜力与挑战[J]. 地球信息科学学报, 2025, 27(10): 2283–2290.

Research progress on multidimensional data formats in remote sensing

ZHANG Lifu^{1,2}, ZHANG Sai^{1,3}, SUN Xuejian^{1,2*}, ZHAO Shuxin^{1,3}, HUANG Changping^{1,2}, GAO Yu^{1,3}, TONG Qingxi^{1,2}

1. State Key Laboratory of Remote Sensing Science, Institute of Aerospace Information Innovation, Chinese Academy of Sciences, Beijing 100101, China
2. National Engineering Research Center for Remote Sensing Satellite Applications, Institute of Aerospace Information Innovation, Chinese Academy of Sciences, Beijing 100101, China
3. School of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100101, China

Abstract This article provides an overview of the domestic and international development of remote sensing multidimensional data formats, especially multidimensional spatiotemporal spectral data, the latest research results, current technical difficulties, and future development directions. This paper mainly introduces the multi-dimensional data format MDD (Multi Dimensional Dataset) proposed by the Aerospace Information Innovation Institute of the Chinese Academy of Sciences for the first time in the world, as well as the theory and technical system of multi-dimensional remote sensing data synthesis and representation, which fills the original gap in data organization in China and has a positive impact on international research. With the rapid development of remote sensing technology, multidimensional spatiotemporal spectral remote sensing data continues to emerge. This article aims to provide an overview of the domestic and international development of remote sensing multidimensional data formats, expound on the latest research achievements, analyze current technical difficulties, and look forward to future development directions, providing comprehensive references for researchers in related fields.

Keywords remote sensing multidimensional data format; MDD; spatiotemporal spectral data; data organization; theoretical and technical system ●



(责任编辑 赵庆圆)