

多源协同定量遥感产品生产系统

——2018年中国地理学会地理大数据计算环境“优秀实用案例”

柳钦火¹, 仲波^{1*}, 唐娉², 张宏海³, 李宏益², 吴善龙¹, 辛晓洲¹,
李静¹, 贾立¹, 单小军², 张正², 闻建光¹, 杜永明¹, 李丽¹,
杨爱霞¹, 历华¹, 胡光成¹, 赵静¹, 张海龙¹, 余珊珊¹, 窦宝成¹, 吴俊君¹

1. 中国科学院遥感与数字地球研究所遥感科学国家重点实验室, 北京 100101;
2. 中国科学院遥感与数字地球研究所遥感卫星工程实验室, 北京 100101;
3. 中国科学院计算机网络信息中心, 北京 100190

摘要: 国家高技术研究发展计划(863计划)重大项目“星机地综合定量遥感系统与应用示范”制定了通过卫星组网的方式获取多源遥感数据研发高时空分辨率高精度定量遥感产品的思路来支撑行业应用。项目一期第四课题“多尺度遥感数据按需快速处理与定量遥感产品生成关键技术”则负责多源遥感数据进行归一化处理和多源协同定量遥感产品反演的任务。在本课题的支持下, 研究团队突破了多源遥感协同高精度定量反演与多尺度定量遥感产品生产关键技术, 形成了全球陆表综合观测的定量遥感产品技术体系; 在集成数据处理与定量产品生产算法的基础上, 研发并试运行了多源协同定量遥感产品生产系统, 形成更高时间分辨率的全球陆表综合观测多尺度定量遥感产品生产能力。近年来, 利用所研发的“多源协同定量遥感产品生产系统”在全球、中国-东盟、“一带一路”及重点试验区4个不同尺度上完成了遥感数据的归一化处理与20余种定量遥感产品的生产, 数据总量近800TB, 形成了生产能力。具有全球生产能力的定量遥感产品包括: 1 km AOD、1 km 地表反射率、1 km 地表反照率、1 km BRDF、1 km 地表温度、1 km 地表发射率、1 km NDVI、1 km EVI、1 km FVC、1 km LAI、1 km FPAR、1 km NPP、5 km DSR、5 km LSR、5 km PAR等15种。

关键词: 多源协同; 定量遥感; 归一化处理; 产品生产系统

DOI: 10.3974/geodp.2018.03.04

1 多源协同定量遥感产品生产系统(MuSyQ)

多源协同定量遥感产品生产系统(MuSyQ)是在攻克多源遥感协同高精度定量反演与多尺度定量遥感产品生产关键技术的基础上, 建立全球陆表综合观测的定量遥感产品技术体系, 并根据该体系研建的按需流程的多尺度遥感数据按需快速处理与定量遥感产品生产原型系统, 初步形成全球陆表综合观测多尺度定量遥感产品生产能力。

收稿日期: 2018-07-20; 修订日期: 2018-09-17; 出版日期: 2018-09-25

基金项目: 中华人民共和国科学技术部(2018YFA0605504)

*通讯作者: 仲波 L-4528-2016, 中国科学院遥感与数字地球研究所遥感科学国家重点实验, zhongbo@radi.ac.cn
作者ID: 柳钦火 S-1647-2016, 仲波 L-4528-2016, 唐娉 S-3703-2018, 张宏海 S-9762-2018, 李宏益 S-9247-2018, 吴善龙 S-1572-2016, 辛晓洲 S-3681-2016, 李静 R-7298-2016, 贾立 I-1457-2014, 单小军 S-1589-2016, 张正 S-3669-2018, 闻建光 S-9821-2018, 杜永明 S-9953-2018, 李丽 S-4533-2016, 杨爱 R-9028-2016, 历华 S-3644-2018, 胡光成 L-6160-2016, 赵静 A-3272-2017, 张海龙 L-6168-2016, 余珊珊 S-4050-2016, 窦宝成 S-3642-2018, 吴俊君 S-3712-2018

引用格式: 柳钦火, 仲波, 唐娉等. 多源协同定量遥感产品生产系统——2018年中国地理学会地理大数据计算环境“优秀实用案例”[J]. 全球变化数据学报, 2018, 2(3): 271-278. DOI: 10.3974/geodp.2018.03.04.

1.1 数据组构情况

大数据主要来源于卫星遥感图像，包括全球 1 km、5 km、25 km 和全国 30 m 分辨率的多源多尺度卫星遥感数据，主要包括：

数据集 30-300 m：主要传感器包括 HJ-1/CCD 和 Landsat 系列卫星数据产品，近期还将包括 GF-1/WFV (16 m) 数据；在将来，根据系统数据的需求，会进一步增加 CEBRS/WFI 和 ZY3/mux 等数据，用以填补数据空缺，使得数据在时间和空间上更加连续；还包括 HJ-1/IRS 和 HIS 等 100-300 m 等数据。

数据集 1 km：主要传感器包括 MODIS、AVHRR、MERSI、VIRR 等中低分辨率光学传感器数据；在历史数据的使用上，还可以使用 MERIS 数据；近期还将加入 VIIRS 数据。

数据集 5 km (25 km、300 km)：主要包括 MTSAT、MSG、FY2、GOES 等静止卫星数据，以及分辨率较低的微波和重力卫星数据。

系统总共收集和整理了 2010-2015 年共 6 年的全球卫星遥感数据和中国及东盟区域 30 m 一年的卫星遥感数据，总量超过了 800 TB；系统生产了全球 6 年 16 种定量遥感产品。

1.2 数据计算能力

1.2.1 计算机服务器与网络设施

图 1 展示了本系统的软件拓扑结构，其中自主研发的分布式计算组件 MyClouds 搭建了弹性可动态扩展的遥感大数据分布式计算环境，是系统调度硬件资源的核心软件组件。

图 2 展示了本系统的硬件拓扑结构，其中平台管理和提供服务的底层资源，主要来自中心平台集群和各创新单元接入的资源（计算资源、存储资源、数据资源、网络资源等）。

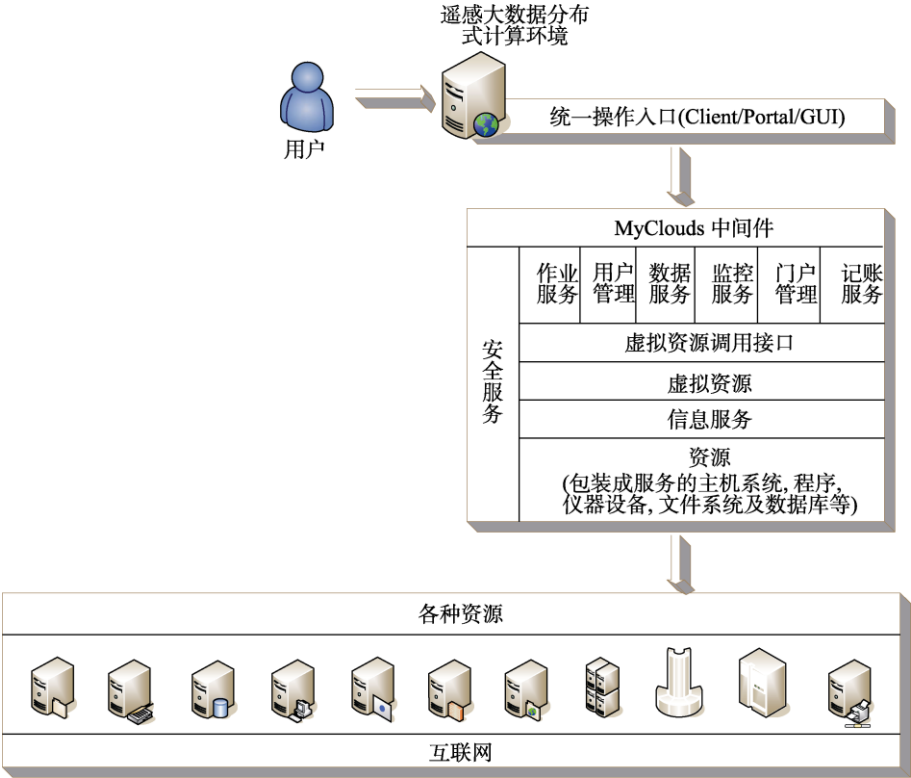


图 1 遥感大数据分布式计算环境-软件拓扑

这些资源将分为两类，独有资源（各创新单元独立管理和使用的资源，只对本创新单元的用户提供服务）和共享资源（各创新单元共享出来的资源，面向所有用户提供服务）。系统主要硬件采用刀片服务器、大内存服务器构成的集群计算系统，配备自主研发的分布式计算组件与分布式数据管理组件及调度软件，可满足 PB 级影像数据高速处理的需求及超大容量的存储和快速访问的要求；集群内部采用 InfiniBand 网络，集群与客户机之间采用千兆网络。

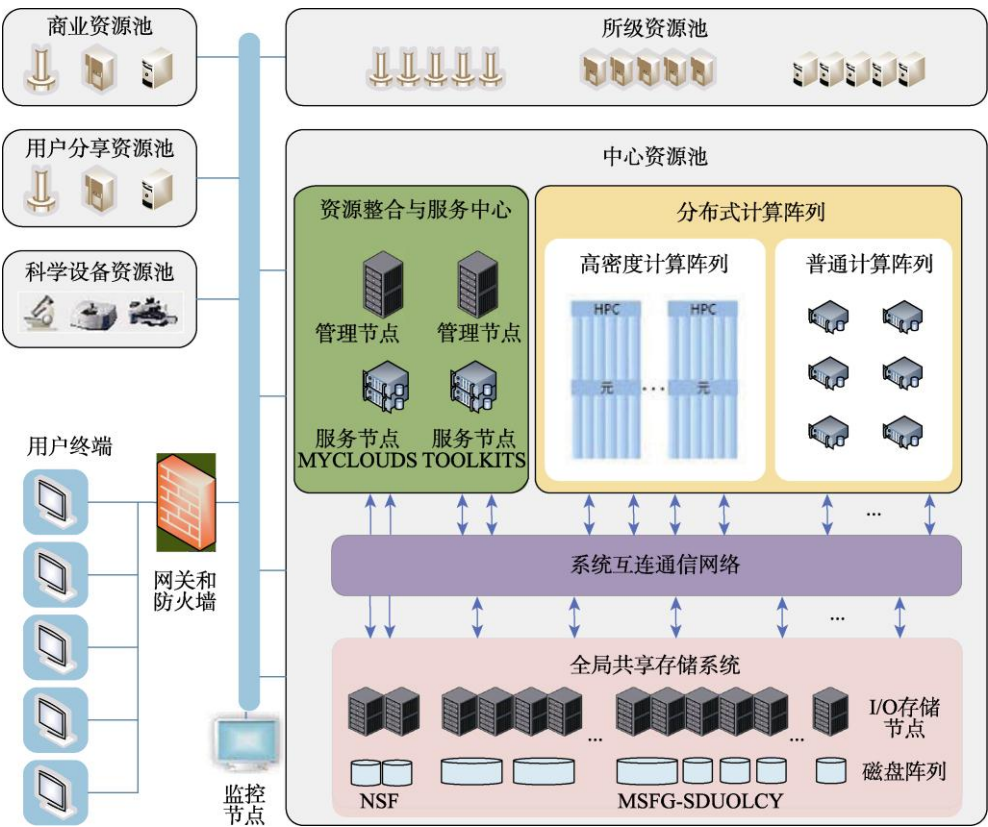


图 2 遥感大数据分布式计算环境-硬件拓扑

1.2.2 模型系统与软件系统

MuSyQ 系统主要包括数据管理、运行管理、归一化数据处理、定量遥感产品生产 4 个分系统以及 1 个前端用户交互界面，软件系统的架构如图 3 所示。系统中的一切数据以及产品生产流程由数据管理分系统进行管理，数据管理分系统为运行管理分系统的任务解析、归一化的数据处理和定量遥感产品生产提供数据流支撑。运行管理分系统负责将产品生产流程转化为计算机可执行的流程脚本，并对生产流程进行调度和监控。归一化数据处理分系统定义了归一化产品的元数据规格、存储结构以及数据格式，集成了 30 m、1 km、5 km 等多个尺度的几何校正、辐射校正、光谱归一化与大气校正等归一化处理模型，采用数据驱动的生产运行模式，当原始数据入库时，自动驱动归一化处理流程，生成归一化产品。定量遥感产品生产分系统定义定量产品的元数据规格、存储结构以及数据格式，集成了辐射收支、植被、冰雪和水热通量共 4 大类 26 个反演模型，通过用户交互界面生成产品生产订单，然后由运

行管理分系统负责解析订单并调度和监控其生产过程。4 个分系统间可以独立部署，通过网络进行互联互通，定量产品生产分系统具有横向扩展能力，内部通过高速局域网进行连接。

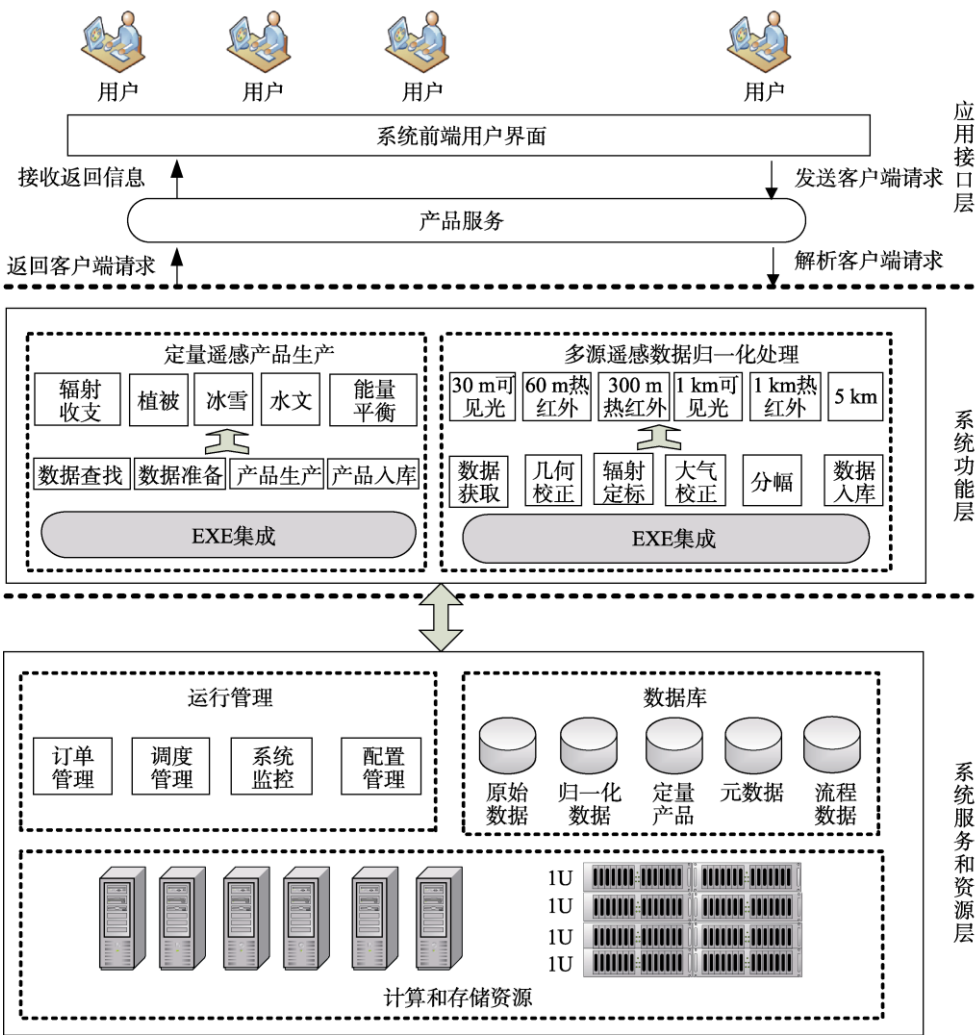


图 3 系统整体架构图

1.2.3 系统数据处理能力

目前系统能够处理多种国内外数据（表 1）及 20 种以上定量产品（表 2），系统还汇集了 2010–2015 年共 6 年的全球卫星遥感数据以及 2014 年中国及东盟区域 30 m 数据，数据包括 MODIS、FY3A/3B/3C MERSI/VIRR、MSG、MTSAT、GOES、FY2E 和 HJ 数据，原始数据总量达 358 TB 左右；目前系统已经生产了全球 6 年 16 种定量遥感产品，处理得到的标准数据和定量产品数据总量达 450 TB 左右。在产品生产过程中，以目前系统硬件配置，处理全球一年的 MODIS、MERSI、VIRR、MSG、MTSAT、GOES 和 FY2 共 10 个传感器的数据，从数据预处理到定量产品生产需要 1 周时间；处理一年全国 30 m 分辨率数据从数据预处理到定量产品生产需要约 2 周时间。

表 1 预处理数据列表

编号	传感器类型	空间分辨率（m）	编号	传感器类型	空间分辨率（m）
1	HJ1A/HJ1B	30	10	GOES11、13、15	5,000
2	TM/ETM	30	11	MTSAT1、2	5,000
3	HJ1BIRS	300	12	葵花 8	5,000
4	Terra/aqua	250、500、1,000	13	MSG2、3	5,000
5	NOAA-AVHRR	1,100	14	FY2E	5,000
6	FY3A MERSI/VIRR	1,000	15	GF1 WfV	16
7	FY3B MERSI/VIRR	1,000	16	GF4	50
8	FY3C MERSI/VIRR	1,000	17	GF1 PMS	8
9	MODIS05、06、07、35	1,000	18	GF2 PMS	5

表 2 定量产品列表

编号	产品名	时间分辨率	编号	产品名	时间分辨率
1	30 m 气溶胶光学厚度	瞬时	12	1 km 植被指数	5 天
2	30 m 植被覆盖度	10 天	13	1 km 反照率	10 天
3	30 m 植被指数	10 天	14	1 km BRDF	10 天
4	30 m 叶面积指数	10 天	15	1 km 发射率	5 天
5	30 m 光合有效辐射吸收比例	10 天	16	1 km 地表温度	瞬时
6	30 m 地表反射率	10 天	17	1 km 光合有效辐射吸收比例	5 天
7	30 m 反照率	10 天	18	1 km 植被净初级生产力	5 天
8	1 km 气溶胶光学厚度	瞬时	19	5 km 下行短波辐射和 PAR	3 小时、1 天
9	1 km 叶面积指数	5 天	20	5 km 下行长波辐射	3 小时、1 天
10	1 km 植被覆盖度	5 天	21	5 km 地表温度	瞬时
11	1 km 地表反射率	5 天	22	300 m 地表温度产品	瞬时

1.2.4 数据计算能力

系统集成的中心常备通用计算资源达 500 TFlops，在线存储能力达 2.0 PB，并且通过 MyClouds 组件可无缝对接中国科技云与其他资源，包括其他研究所资源、实验室资源及企业资源，如阿里云、亚马逊云资源等，可随时动态扩展计算资源与存储资源。

1.2.5 网站

全球变化科学研究数据出版系统：<http://www.geodoi.ac.cn/WebCn/Default.aspx>。

全球生态环境遥感监测年报：<http://www.chinageoss.org/geoarc/2017/>。

1.3 通过系统完成的主要科学发现与科学成果

（1）系统性地发展了多源遥感数据协同反演理论，将定量遥感产品从基于单一遥感数据的 8/16 天分辨率提高到了多源协同的 5 天，且空间覆盖的完整性也得到了提高；

（2）首次形成了 10 余种卫星遥感数据的归一化处理流程，尤其是国产卫星数据归一化处理中的关键技术问题，实现了多源遥感数据协同使用；

（3）利用多源遥感数据生产了全球 6 年的辐射收支、植被、水热通量产品，支撑了国家遥感中心《全球生态环境遥感监测年报》的编写。

（4）自主研发的分布式计算组件 MyClouds 首次研发了针对遥感数据处理的分布式计

算环境,具有轻量级、稳定实用、安全可靠的特性,提供了丰富的工具,方便用户使用,可满足PB级影像数据高速处理的需求及超大容量的存储和快速访问的要求。

2 与当前国内外同类系统的综合比较

2.1 与国际同类计算环境的比较

国际上与MuSyQ系统类似的系统包括美国EOS/MODIS的数据处理、产品生产与发布系统^[1]、欧空局的MERIS处理系统^[2];这两个系统都是针对单一传感器的系统,遥感数据种类少、尺度单一、数据量小、算法和模型复杂度也比较低。计算环境也是基于单一计算服务器,计算的负载量小。

MuSyQ系统将多个传感器多尺度(30 m、1 km、5 km等)的全球数据按照适宜的剖分原则统一组织、管理,采用统一的数据格式读写库和跨尺度、跨投影的数据转换适配器,系统具备多分辨率数据独立处理及跨尺度协同处理的能力,集成了26个遥感定量产品算法模型,实现了多源协同定量遥感产品的自动化生产。相比类似系统,MuSyQ系统在复杂的产品生产工作流程中实现了无冗余节点产品生产,生产效率更高、数据存储组织变形更小、精度更高;通过数据结构和并行优化,针对单景/剖分面片数据密集型计算,计算效率更高。

MuSyQ系统中采用了MyClouds分布式计算环境,国际上与MyClouds分布式计算环境类似的系统包括美国的XSEDE^[3],XSEDE主要是针对HPC的网格计算环境,采用的是针对HPC的二级调度,没有全局的作业调度系统,而遥感大数据分布式计算环境,不但可集成HPC集群,还可集成普通的服务器及云主机,有自动集群生成软件环境,整个环境更经济、更灵活、更专业,对资源利用的更充分。

2.2 与国内同类计算环境的比较

国内与MuSyQ系统类似的系统也主要是针对不同传感器的数据处理与分发系统,包括中国资源卫星应用中心的中高分辨率数据分发系统、国家气象局的风云卫星处理与分发系统。这些系统一般只具有L1B产品生产的能力,并不能对数据进行辐射和几何的归一化;风云系统也具有定量遥感产品的生产能力,但基于单一数据生产的产品数量、产品时空覆盖范围及精度等都有待提高。

3 实用性、先进性和创新点

(1)首次完成了30余种多源遥感数据和遥感产品的归一化处理,形成了能够支持多源多尺度遥感数据协同反演的归一化数据集。

多源及多尺度遥感数据协同使用的难点首先体现在这些数据的一致性较差,由于数据的不一致性增加了噪声,往往导致协同反演失败。本课题发展了包括几何归一化、光谱归一化、交叉辐射定标、大气校正及多尺度数据分幅等关键技术,可以完成30 m、300 m、1 km、5 km及25 km 5个尺度共计30余种遥感数据和遥感产品的归一化处理,形成具有较高一致性的多源遥感数据集,为多源数据的协同反演奠定了基础,在全球属于首次。

(2)研发了“多源协同定量遥感产品生产系统”,形成了10余种1 km(5天合成)和5 km(3小时)全球产品及10种30 m和300 m(10天到1个月合成)中国及周边区域产

品快速生产能力。

“多源协同定量遥感产品生产系统 (MuSyQ)”集成了多源多尺度遥感数据归一化处理技术和多源协同定量遥感产品生产算法,可以对现阶段主流的 30 m-5 km 分辨率遥感数据进行归一化处理,并能够生产分辨率为 1-5 km 的近 20 种全球定量遥感产品和 10 种分辨率为 30-300 m 的中国及周边区域定量遥感产品,大部分定量遥感产品都利用多源遥感数据协同生产的算法;在数据从处理能力上,这是世界上唯一可以同时处理 10 种以上多源遥感数据的系统;从先进性的角度考虑,该系统中所集成的定量遥感产品生产算法大部分都利用了多源遥感数据协同生产算法,将 1 km 分辨率产品的时间分辨率提高到了 5 天,将 30 m 分辨率产品的时间分辨率最好可以提高到 10 天,对于陆表的变化有了频次更高的观测,对于行业应用、科学研究等具有重要的意义,推动了定量遥感产品的生产从单一遥感数据向多源遥感数据协同转变。

(3) 将 FY-2、FY-3、HJ-1 等国产数据大规模用于不同尺度序列化的定量产品生产中,实现了利用国产数据为中国及周边区域提供多种 30 m 和 300 m 高时间分辨率 (10 天到 1 个月合成) 定量产品快速生产的能力。

(4) 促进遥感技术与超算技术的融合发展

针对海量遥感数据的处理,引入了遥感大数据分布式计算环境,利用自主研发的分布式计算组件 MyClouds 搭建了弹性可动态扩展的遥感大数据分布式计算环境,提供了有效的资源共享与调度机制,用户可随时将拥有的资源共享出来,也可随时退出来,用户可优先使用本地资源,只有在资源不足时,系统才会调度共享资源;配备自主研发的分布式计算组件与分布式数据管理组件,可满足 PB 级影像数据高速处理的需求及超大容量的存储和快速访问的要求,利用该平台的分布式计算能力实现了全球及重点区域产品的快速生成;推动了分布式计算技术在遥感数据处理与定量遥感产品生产方面的应用,有助于推动遥感大数据领域的发展。

4 推广应用

4.1 直接用户

系统的直接用户包括科研院所、行业应用部门和国家行政机构等。除此以外,随着国家“一带一路”战略规划的不断延伸,系统所生产的产品可以服务于国家企业在国外项目开展过程中对于资源、环境方面信息的获取,也可以将所生产的产品和系统本身在“一带一路”沿线国家进行推广和应用。

4.2 应用前景

(1) 推进定量遥感产品在科研院所、行业部门及国际行政机构的应用

利用 MuSyQ 系统所生产的定量遥感产品已经在相关的科研院所、行业应用部门和国家行政机构等进行了应用,并获得了应用证明。所生产的产品连续三年应用到“全球生态环境遥感监测年度报告”中,发表相关数据集近 30 套,截至 2018 年 6 月 17 日,报告数据集下载总量达到 124.95 TB。MuSyQ 系统支撑了《中国可持续发展遥感监测报告 (2016)》,报告在 2017 年 6 月发布,20 余家媒体记者出席了新闻发布会,之后人民日报、人民日报海外版、新华社、解放军报、光明日报、中国日报等 20 余个全国性大媒体以及 30 余地方

媒体对此给予了报道, 36家网媒进行了广泛转载, 其中, 中央级中字头媒体8家。

(2) MuSyQ 系统具有市场化潜力, 促进遥感技术的产业化发展

MuSyQ 具备 10 余种多源遥感数据处理及 20 余种定量遥感产品生产的能力, 可以降低定量遥感应用的门槛, 为地方政府、公司企业和公众提供服务。在降低定量遥感应用门槛的同时必然提升其市场份额, 进一步促进遥感产业化发展。近期, 行业部门已经对定量遥感产品生产系统提出了强烈的需求, 并开始部署相关的建设任务。系统本身和系统中的数据处理技术和多源协同定量遥感产品生产技术都可以直接服务于这些建设项目, 具有很好的产业化前景。

5 科学发现成果

重大科学发现包括:

(1) 系统性地发展了多源遥感数据协同反演理论, 将定量遥感产品从基于单一遥感数据的 8/16 天分辨率提高到了多源协同的 5 天, 且空间覆盖的完整性也得到了提高;

(2) 首次形成了 10 余种卫星遥感数据的归一化处理流程, 尤其是国产卫星数据归一化处理中的关键技术问题, 实现了多源遥感数据协同使用;

(3) 利用多源遥感数据生产了全球 6 年的辐射收支、植被、水热通量产品, 支撑了国家遥感中心《全球生态环境遥感监测年报》的编写。

参考文献

- [1] Masuoka, E., Tilmes, C., Devine, N., *et al.* Evolution of the MODIS science data processing system [C]. Sydney: Geoscience and Remote Sensing Symposium, 2001: 1454–1457.
- [2] MERIS web site [OL]. <https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/envisat/instruments/Meris>.
- [3] Lifka, D., Foster, I., Mehringer, S., *et al.* XSEDE cloud survey report [R]. XSEDE Cloud Integration Investigation Team, 2013. <http://www.cac.cornell.edu/technologies/XSEDECloudSurveyReport.pdf>.
- [4] 中国科学院. 中国可持续发展遥感监测报告(2016) [R]. 经济日报, 2017 年 6 月 21.