2017年度“清华——浪潮地球系统科学青年人才奖”推荐书

被推荐人姓名：陆春松　专业技术职务：教授　专业：大气物理学与大气环境

出生年月：[redacted]

工作单位：南京信息工程大学

电子信箱：[redacted]　电话：[redacted]

我郑重推荐 陆春松 作为“清华——浪潮地球系统科学青年人才奖”的候选人。

推荐人签名：徐祥德

推荐人单位公章：

2017年11月27日

推荐人姓名：徐祥德　专业技术职务：研究员　专业：天气动力学

工作单位：中国气象科学研究院

电子信箱：[redacted]　电话：[redacted]

推荐身份：☐中国科学院院士　■中国工程院院士　☐国家和部委重点实验室主任

☐重点高等院校校长　☐长江学者　☐重点科研院所院（所）长

☐杰出青年基金获得者　☐千人计划学者　☐国家重大科技项目首席科学家

请简要说明被推荐人的原创性科学或技术成果的内容、重大科学意义、对促进科学技术发展所起的作用及产生的影响（限2000字以内，纸面不够可加页），相关证明材料和论文原文以附件方式提供。

陆春松成果总结

云物理学是大气科学的基础学科，对降水、气候变化等科学研究至关重要。陆春松专攻于云物理及其参数化方面的研究，取得了丰硕的成果。目前发表论文37篇，在《Geophys Res Lett》、《J Geophys Res》、《J Atmos Sci》等刊物上发表SCI论文26篇，第一作者/通讯作者SCI论文14篇。

2015年，因陆春松在云物理及其参数化方面取得的成绩，被美国地球物理学会授予霍尔顿青年科学家奖（Holton Junior Scientist Award）。陆春松是该年度唯一的获奖人，也是首次由华人获得该奖项。美国地球物理学会大气科学部主任William K. M. Lau教授在颁奖词中指出：**对流夹卷、湍流混合及其与云雾微物理相互作用是富有挑战性的重要问题，陆春松的研究目标明确、创造力强，具有独到之处**。

2016年，陆春松荣获了美国美华海洋大气学会的宇翔青年学者奖。他还曾获得**美国布鲁克海文国家实验室的突出贡献奖、ESI高被引论文奖、涂长望青年气象科技奖、谢义炳青年气象科技奖**、《Atmos Res》杂志高引用率论文奖、美国地球物理学会优秀学生论文奖、美国能源部科学办公室学生海报奖、江苏省优秀博士论文奖等。**2016年，被选为国际云降水物理委员会（ICCP）委员（共27名委员，中国有2名）**。

近五年来，陆春松主持国家级项目3项、省部级项目2项，并**担任国家重点研发计划第一课题负责人**（清华大学牵头），**研究内容都与改进气候模式中云物理过程的参数化方案有关**。他还获得了4个人才项目，包括**江苏省杰青**、“六大人才高峰”高层次人才（B类）、“333工程”中青年科学技术带头人、“青蓝工程”优秀青年骨干教师。

主要学术成绩概述：

1. 关于积云中夹卷率的研究及其参数化

积云参数化方案中夹卷率的数值存在很大的不确定性，影响模式对降水、台风、热带大气季节内振荡等的模拟效果。因此，开展夹卷率影响因子及其参数化方面的研究具有重要的科学意义和应用价值。

陆春松研建的夹卷率计算新方法不确定性比传统方法小约30%，精度高。利用该新方法的优势，揭示了夹卷率的影响因子及夹卷率对云微物理量的影响，并**研建了适用于不同气候模式分辨率的夹卷率的参数化方案。**

以上内容发表8篇SCI论文，其中第一作者/通讯作者论文6篇。以第一作者发表在**《J Atmos Sci》上的论文入选为ESI高被引论文。**

荷兰代尔夫特理工大学的Steven Böing在《J Atmos Sci》上指出，用飞机观测资料估算夹卷率的概率密度分布函数**是非常困难的，但Lu et al. (2012)应用新方法，成功**获得了概率密度分布函数。**美国密歇根科技大学的Raymond Shaw在《Science》上引用本研究内容，用于阐述他们观测结果的理论意义。**美国国家强风暴实验室的David Turner、以色列魏茨曼科学研究学院的Ilan Koren、美国地球物理流体动力学实验室（GFDL）的Leo J. Donner、美国斯克里普斯海洋研究所的Guang Zhang、美国西北太平洋国家实验室的B. Schmid、美国布鲁克海文国家实验室的Satoshi Endo等学者在《Atmos Chem Phys》等刊物上引用本研究结果，**强调了夹卷率计算新方法的先进性和参数化方案的可靠性，并用于证明他们观测/模拟结果是合理的。**

该内容在2015、2016年美国地球物理学会年会上以口头报告的形式展示。该研究得到了美国能源部布鲁克海文国家实验室**刘延刚高级研究员的充分肯定：该工作为利用环境探空资料和地基微波辐射计的含水量资料，进行夹卷率垂直廓线的常规例行监测打下基础。**美国地球物理学会大气科学部主任William K. M. Lau教授在**霍尔顿青年科学家奖的颁奖词**中指出：**该工作对气候模式中积云参数化方案的改进有很大的推动作用，从而更好地模拟大气对流过程，包括对流的日变化及Madden-Julian震荡（MJO）**。

2. 关于低云中混合蒸发机制的研究及其参数化

气候模式中云与气候变化之间的反馈存在很大的不确定性，低云中的混合蒸发过程是导致该不确定性的一个重要原因。因此，需要深入理解云中的混合蒸发机制，并建立参数化方案。

陆春松定义了一个新的无量纲数（过渡尺度数），揭示了积云和层积云中占主导的混合蒸发机制，阐明了混合蒸发机制对云微物理量的影响，揭示了混合蒸

发机制的尺度依赖性；在此基础上，定义了均匀混合百分比，**构建了混合蒸发机制的参数化方案，**该研究成果将有助于提高模式对云-气候之间反馈过程的模拟能力。

以上内容发表9篇SCI论文，其中6篇为第一作者。**成果被写入美国气象学会气象专著《Meteorological Monographs》“大气辐射观测项目：前20年”中（第22章，由Graham Feingold和Allison McComiskey撰写）。**

美国密歇根科技大学的Raymond Shaw**在《Science》上引用本研究成果，用于验证他们关于云滴谱的观测结果。**美国国家大气研究中心的Wojciech Grabowski、美国迈阿密大学的Bruce Albrecht、德国马普研究所的Bipin Kumar、加拿大麦吉尔大学的Paloma Borque、以色列耶路撒冷希伯莱大学的Mark Pinsky、加拿大环境与气候变化部的Alexei Korolev等学者在《Bull Amer Meteor Soc》等刊物上引用本研究内容，**充分肯定了新定义的物理量在混合蒸发机制研究中的理论意义和实用性，并认为新研建的参数化方案将促进气溶胶间接效应的评估。**

韩国延世大学大气科学系主任Seong Soo Yum教授认为该研究内容新颖，对于发展云中夹卷过程的参数化具有非常重要的意义，2015年11月，**邀请陆春松到该校访问，并做特邀报告。美国能源部布鲁克海文国家实验室也对该研究成果给予充分认可，授予陆春松突出贡献奖。**

2017 年度“清华——浪潮地球系统科学青年人才奖”推荐书

被推荐人姓名: 马 艳 专业技术职务: 副研究员 专业: 信号与信息处理

出生年月: [redacted]

工作单位: 中国科学院遥感与数字地球研究所

电子信箱: [redacted] 电话: [redacted]

我郑重推荐 马艳 作为“清华——浪潮地球系统科学青年人才奖”的候选人。

推荐人签名: 郭华东

推荐人单位公章: 中国科学院遥感与数字地球研究所 110000000098036

2017 年 12 月 6 日

推荐人姓名: 郭华东 专业技术职务: 研究员 专业: 雷达遥感、数字地球

工作单位: 中国科学院遥感与数字地球研究所

电子信箱: [redacted] 电话: [redacted]

推荐身份: √中国科学院院士 □中国工程院院士 □国家和部委重点实验室主任

□重点高等院校校长 □长江学者 □重点科研院所院（所）长

□杰出青年基金获得者 □千人计划学者 √国家重大科技项目首席科学家

请简要说明被推荐人的原创性科学或技术成果的内容、重大科学意义、对促进科学技术发展所起的作用及产生的影响（限 2000 字以内，纸面不够可加页），相关证明材料和论文原文以附件方式提供。

作为重要的地球系统科学数据之一，大规模多源、多时相时空遥感数据的处理与分析，对全球资源环境变化及可持续发展问题的研究具有重要意义。围绕海量遥感数据处理与分析中，全球时空遥感数据组织、计算模型及系统平台等方面科学问题，被推荐人马艳完成了以下原创性工作：

原创性成果一：海量时空遥感数据组织与计算模型

海量遥感数据的多源、非结构化、高维时空特性与线性、多层次的分布式计算机存储结构的不匹配，导致海量时空遥感数据及处理模型在现代高性能计算架构上可实现性不足。针对于此，被推荐人提出了基于泛型算子的海量全球时空遥感数据组织与计算模型，包括：（1）**基于泛型算子的遥感数据并行处理方法**:通过引入泛型算子对核心算法进行抽象表示，并对遥感数据处理的典型并行模式、公共算法及计算流程进行高度抽象与统一泛型模板化表示，实现遥感数据处理的隐性自动并行化实现。试验结果表明基于该计算模型的遥感数据处理实现效率提高约 5-6 倍，自动并行化系统损失控制在 5%以内，解决了当前大规模遥感数据并行效率低下的难题。（2）**基于多填充曲线的分布式时空遥感数据组织模型**：突破单一数据布局模式局限，提出访问模式感知的外存多副本数据布局优化模型，采用多空间填充曲线实现高维时空遥感数据在线性外存结构上的动态数据组织与映射，以同时适应多种非规则遥感数据访问模式；同时基于单边消息通讯技术，构建具有全局视图的分布式遥感数据内存组织模型，解决了高维、非结构化时空遥感数据在线性、多层分布内存单元的组织与表达难题，实验显示基于该模型的 I/O 加速给遥感算法带来 15%～20%性能优化。

成果产出与评议：该研究发表顶级 SCI 期刊论文 3 篇[1,6,7]，获得发明专利 4 项[11~14]。其中 JCR 二区论文 2 篇（TPDS,CCPE 等）[1][7]，乔治梅森大学教授引用该研究为“遥感数据处理高性能计算的最新进展”；1 篇 Cluster’12（录取率 28.86%）[6]，同行认为 “系统实现具有创新性”。

科学意义及影响：该研究解决了海量多源、高维、非结构化时空遥感数据在多层、分布式高性能计算机结构上的数据组织表达与计算难题，并显著提高了数据存储与处理性能。研究成果已成功应用于我国首套基于并行集群的 “BJ-1”小卫星地面预处理系统，载人航天“天宫一号”卫星数据处理系统；还应用于“发改委”遥感卫星国家工程实验室及“863”重大项目-跨区域遥感综合监测系统，并取得良好效果。

原创性成果二：基于多数据中心的全球时空遥感大数据处理方法

遥感数据的跨多中心分布、信息产品服务的多样需求给传统基于静态算法流程的遥感数据处理带来了极大的资源优化与并行处理挑战。针对于此，被推荐人提出基于多数据中心的全球时空遥感数据分布式处理方法，包括：**（1）遥感大数据“数据密集处理”问题的量化指标模型**：通过梳理现代卫星数据中心遥感数据处理的全流程，综合分析影响处理效率的数据量、算法复杂度、数据吞吐率等相关因素，创新性地提出了遥感大数据“数据密集处理”问题的量化指标模型，定量化刻画了遥感大数据处理效率的指标因子，为指导遥感数据处理算法及系统架构优化建立了理论基础。**（2）基于多卫星数据中心的分布式遥感数据处理方法**：基于上述量化模型，建立多卫星数据中心资源优化模型，对多数据中心的数据、计算等资源进行最优调度，提高卫星数据中心的资源利用率，解决分布式资源优化难题；进而提出基于异构卫星数据中心的分布式遥感数据处理方法，在支持数据中心自治的同时，基于多中心资源优化模型及云计算技术，实现数据驱动的全球时空遥感数据处理流程动态构建与并行处理，解决分布式多中心架构的遥感数据并行处理难题。

成果产出与评议：该研究发表顶级 SCI 期刊论文 5 篇[3~5,8,10]，其中 JCR 一/二区 (Info. Sci ，FGCS)3 篇[3~5],[3]为 ESI 热点论文，引用 107 次。加拿大Ryerson 大学等学者在引用中评价该研究为“地球空间大数据的一个典型代表”，希腊国立科技大学称该研究“为前沿地学空间大数据框架创造了机会”。

科学意义及影响：该研究首次解决了多卫星数据中心架构的分布式资源优化与遥感数据并行处理难题，并为遥感数据处理及系统优化提供了理论模型基础。研究成果已应用于“863”重大专项-星机地综合定量遥感系统与应用示范中，连接资源、海洋、遥感地面站等 6 个国内主要卫星中心，实现了全球定量遥感时空产品的快速在线生产。

原创性成果三：基于大规模关联任务调度的分布式流域生态水文过程模拟

现有大规模遥感应用中，大量关联子任务间的数据依赖关系导致并行处理效率低下，数据规模难以扩展等问题。针对此，被推荐人提出了基于任务树的关联遥感任务优化调度方法，通过构建任务树模型，引入基于关键路径和状态队列的 DAG 调度算法及多任务包计算（many-task），进行关联任务控制与最优化调度。该研究获得国家自然基金资助，在小时级完成了亚洲范围高分辨率影像的快速自动镶嵌，性能提高了近 20 倍；在分布式流域生态水文模型 DisVOM 上实现了二十万量级任务并行的鄱阳湖地区精细水文过程模拟。实验表明该方法可突破数据及计算规模限制，获得较好性能。相关研究发表顶级 SCI 论文（TPDS、Supecomput）2 篇[2,9]，获得发明专利 1 项[15]。同行评议该研究为“突破大规模遥感数据处理可扩展限制”的一个“很好的想法”。

相关证明材料和论文原文（以附件方式提供）

● 代表性论文

[1] Lizhe Wang, ***Yan Ma****, Albert Zomaya, Rajiv Ranjan Dan Chen, "A Parallel File System with Application-aware Data Layout Policies in Digital Earth," ***IEEE Trans. Parallel . Syst. (TPDS)***, vol. 26, no. 6, pp. 1497-1508, 2015.(CCF“A 类”, IF:**2.17**,,引用：**28**) ***(Highlight article From the June 2015 Issue)***

[2] ***Yan Ma***, Lizhe Wang, Albert Y. Zomaya, Dan Chen, Rajiv Ranjan, "Task-Tree based Large-Scale Mosaicking for Remote Sensed Imageries with Dynamic DAG Scheduling," ***IEEE Transactions on Parallel and Distributed Systems***, 20 Nov. 2013.（IF: **2.17**,CCF“A 类”,引用：**41**）

[3] ***Yan Ma,*** Haiping Wu, Lizhe Wang, Bormin Huang, Rajiv Ranjan, Albert Zomaya, Wei Jie, “Remote sensing big data computing: Challenges and opportunities”, ***Future Generation Computer Systems(FGCS),*** 51:47 – 60, 2015 （JCR 二区，IF:2.64, 引用：**107**）

[4] ***Yan Ma***, Lizhe Wang, Peng Liu, and Rajiv Ranjan. Towards building a data-intensive index for big data computing – A case study of remote sensing data processing. ***Information Sciences,*** 319:171–188, 2015. (IF：4.038，JCR 一区，引用：**35**)

[5] LZ. Wang, ***Y. Ma****, J. Yan, V. Chang, AY Zomaya; pipscloud: High performance cloud computingfor remote sensing big data management and processing. ***Future Gener. Comput.Syst.(FGCS)***, 78 (2018), pp. 353–368 (JCR 二区，IF:**2.64,** 引用：**13)**

[6] ***Ma Y***, Wang L, Liu D, et al. Generic Parallel Programming for Massive Remote Sensing Data Processing. ***Cluster Computing (CLUSTER)***, 2012 IEEE International Conference on. IEEE, 2012: 420-428. (20125015790095)

[7] ***Yan Ma***, Lizhe Wang, Dingsheng Liu, Tao Yuan, Peng Liu, Wanfeng Zhang: Distributed data structure templates for data-intensive remote sensing applications. ***Concurrency and Computation: Practice and Experience***.25(12): 1784-1797 (2013)（引用：**27**）

[8] Zhang, J., Yan, J., ***Ma, Y****. et al. Infrastructures and services for remote sensing data production management across multiple satellite data centers. ***Cluster Comput*** (2016) 19: 1243.

[9] Shasha Yue,***Yan Ma****,Lajiao Chen,Yuzhu Wang,Weijing Song. Dynamic DAG scheduling for many-task computing of distributed eco-hydrological model. ***J Supercomput*** (2017)

[10] Lizhe Wang, Jie Tao, ***Yan Ma***, Samee U. Khan, Joanna Kolodziej, and Dan Chen. Software design and implementation for mapreduce across distributed data centers. ***International Journal of Applied Mathematics & Information Sciences***, 7(1):85–90, 2013.

● 发明专利

[11] **马艳**,王力哲,刘定生,刘鹏. 面向数据密集型遥感图像处理的泛型编程的框架编程方法. 中国, 103019671A. 2012-10-08. （已授权发明专利）

[12] **马艳**,刘定生,李景山,赵灵军,刘鹏,王妍颖.面向遥感图像处理算法的快速并行化编程模板方法. 中国, 101799762A(P). 2010-08-11 （已授权发明专利）

[13] 赵灵军,刘定生,李景山,**马艳**,汤灿恩. 遥感影像处理系统中处理功能可快速扩展的构建方法. 中国, 101814029A(P). 2010-08-25（已授权发明专利）

[14] 刘定生，李国庆，章文毅，李景山，赵灵军，**马艳**，王建等.多卫星遥感数据一体化并行地面预处理系统. 中国, 101315424[P]. 2008-12-03. （已授权发明专利）

[15] **马艳**,王力哲,刘定生,刘鹏,刘志文. 一种基于任务树的大尺度遥感影像并行镶嵌方法. 中国, 103150148A. 2013-03-06

A Parallel File System with Application-Aware Data Layout Policies for Massive Remote Sensing Image Processing in Digital Earth

Lizhe Wang, *Senior Member, IEEE*, Yan Ma, *Member, IEEE*, Albert Y. Zomaya, *Fellow, IEEE*, Rajiv Ranjan, *Member, IEEE*, and Dan Chen

Abstract—Remote sensing applications in Digital Earth are overwhelmed with vast quantities of remote sensing (RS) image data. The intolerable I/O burden introduced by the massive amounts of RS data and the irregular RS data access patterns has made the traditional cluster based parallel I/O systems no longer applicable. We propose a RS data object-based parallel file system for remote sensing applications and implement it with the OrangeFS file system. It provides application-aware data layout policies, together with RS data object based data I/O interfaces, for efficient support of various data access patterns of RS applications from the server side. With the prior knowledge of the desired RS data access patterns, HPGFS could offer relevant space-filling curves to organize the sliced 3-D data bricks and distribute them over I/O servers. In this way, data layouts consistent with expected data access patterns could be created to explore data locality and achieve performance improvement. Moreover, the multi-band RS data with complex structured geographical metadata could be accessed and managed as a single data object. Through experiments on remote sensing applications with different access patterns, we have achieved performance improvement of about 30 percent for I/O and 20 percent overall.

Index Terms—Parallel file system, parallel I/O, data-intensive computing, digital earth, big data, remote sensing image processing

1 INTRODUCTION

DIGITAL Earth [1], a global information model of earth's surface, plays a significant role in addressing challenges in global resources and environmental change. The digital earth approach is to exploit regional to global covered remote sensing data for processing. With recent advances in remote sensing technology, the latest-generation space-borne sensors are continuously producing massive quantities of high-dimensional remote sensing (RS) image data in Digital Earth at a rate of several terabytes per day [2], [3]. Meanwhile, remote sensing data processing for digital earth is commonly recognized as typical data-intensive applications. The large-scale RS applications like continental mosaicking for Global Rain Forest Mapping [4] normally require processing thousands of imageries that adds up to nearly terabytes of data.

Incorporation of distributed cluster-based high-performance computing (HPC) with remote sensing applications in Digital Earth is an effective solution for addressing these computational challenges introduced by massive data. However, the performance gap between I/O and computing is gradually widening by further growth of computing capability, especially in cluster scenarios with thousands of cores. In this case, the data processing has to wait for several CPU cycles for data accessing [5]. Withal, most of the remote sensing applications are significantly challenged with the specific complex data access patterns that result from different correlations between computation and data. These complex data access patterns perform intensive small non-contiguous or irregular I/O over multiple data files and vary across different applications. The data I/O then turns out to be an expensive burden for remote sensing applications in digital earth. Remote sensing image data sets, in particular, are characterized by multi-dimensional image data structure and abundant geographical metadata. Consequently, offering direct and efficient storing and accessing of these special structured RS data sets in traditional file system could be rather cumbersome, although it remains paramount.

The intolerable I/O burden introduced by massive quantities of RS data and the complex data access patterns has rendered traditional I/O systems no longer applicable. Employing a parallel file system with Parallel I/O [6], [7] is an effective solution to tackle the I/O challenges encountered in remote sensing applications. The I/O parallelism is commonly achieved by block-wise striping of large data across a great number of storage blocks with a fixed stripe size. This method allows applications to access data from multiple I/O servers or storage concurrently. However, despite the high I/O parallelism, parallel file systems continue to suffer from the relatively poor performance of non-contiguous I/O patterns. The reason for this is that most of the mainstream parallel file systems seldom directly support these data accesses patterns properly [8]. Furthermore, although the standard data striping method

- *L. Wang and Y. Ma are with the Institute of Remote Sensing and Digital Earth, Chinese Academy of Science, Beijing, China 100094. E-mail: lizhe.wang@gmail.com, yanma@ceode.ac.cn.*
- *A.Y. Zomaya is with the School of Information Technologies, the University of Sydney, Australia.*
- *R. Ranjan is with ICT/CSIRO, Australia.*
- *D. Chen is with the School of Computer, Wuhan University, Wuhan, Hubei, China.*

Manuscript received 21 Dec. 2013; revised 21 Mar. 2014; accepted 19 Apr. 2014. Date of publication 2 June 2014; date of current version 8 May 2015. Recommended for acceptance by J. Chen.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPDS.2014.2322362

Task-Tree Based Large-Scale Mosaicking for Massive Remote Sensed Imageries with Dynamic DAG Scheduling

Yan Ma, *Member, IEEE*, Lizhe Wang, *Senior Member, IEEE*, Albert Y. Zomaya, *Fellow, IEEE*, Dan Chen, and Rajiv Ranjan, *Member, IEEE*

Abstract—Remote sensed imagery mosaicking at large scale has been receiving increasing attentions in regional to global research. However, when scaling to large areas, image mosaicking becomes extremely challenging for the dependency relationships among a large collection of tasks which give rise to ordering constraint, the demand of significant processing capabilities and also the difficulties inherent in organizing these enormous tasks and RS image data. We propose a task-tree based mosaicking for remote sensed imageries at large scale with dynamic DAG scheduling. It expresses large scale mosaicking as a data-driven task tree with minimal height. And also a critical path based dynamical DAG scheduling solution with status queue named CPDS-SQ is provided to offer an optimized schedule on multi-core cluster with minimal completion time. All the individual dependent tasks are run by a core parallel mosaicking program implemented with MPI to perform mosaicking on different pairs of images. Eventually, an effective but easier approach is offered to improve the large-scale processing capability by decoupling the dependence relationships among tasks from the complex parallel processing procedure. Through experiments on large-scale mosaicking, we confirmed that our approach were efficient and scalable.

Index Terms—Remote sensing image processing, parallel computing, DAG scheduling, data-intensive computing, big data computing

1 Introduction

In recent years the demands for accurate and up-to-date [1] environmental information at regional to global scale are increasing dramatically. The spaceborne sensors which conduct global observations [2] of the earth surface continuously have provided a unique and quantitative way for global monitoring. The efforts to improve the large scale capabilities of mosaicking [3] have received particular attention so to solve the problems of land use or climate changes [4], [5]. The Global Rain Forest Mapping (GRFM [5]) project has produced several semi-continental mosaics using some 13000 scenes of JERS-1 L-band SAR images. The United States Geological Survey (USGS) team has created Antarctica [6] mosaic from over 1,000 ETM+ scenes.

Mosaicking [7] of remote sensed (RS) imageries usually stitches a large collection of scenes with overlapping regions into a geometric alignment, radiometric balanced seamless one. Subsequently, a continuous view of the entire big area could be formed. The sheer volume of national-scale mosaic will be hundreds of gigabytes, and the continental mosaic could even exceed one terabyte [6]. Extremely large amounts of data need to be processed. Generally, mosaicking for remote sensed imageries mainly consists of re-projection, registration, radiometric balancing, seam line extraction and blending. This processing chain with multi-stage [7] is rather compute-intensive and also quite time-consuming. Generating a regional mosaic with a dozen of scenes would take couple of hours. When scale to large region, image mosaicking becomes extremely challenging due to the vast volume of image data, multiple processing steps with higher complexity, and also the intricate adjacent relationships among large collection of scenes with overlapping regions. Particularly, some time-critical applications like disaster assessment even require near real-time processing capabilities.

The enormous computation introduced by large-scale image mosaicking requires significant processing capabilities, which have far, outpace a single computer. Appling multi-core cluster based high-performance computing (HPC) in large-scale mosaicking is an effective solution to address these computational challenges. The parallelism is normally achieved by recursively breaking down large-scale processing issue into smaller tasks responsible for regional mosaics. In this way, certain tasks dealing with larger regions wait for the regional mosaics generated by much smaller tasks as input to be available. However, in the mosaicking scenario showed here, procedure consists of several complex processing stages, vast adjacent scenes with overlapping regions, and the tasks performing on these scenes subject to ordering constraints. The complexities inherent in organizing large numbers of "messy"

- *Y. Ma and L. Wang are with the Institute of Remote Sensing and Digital Earth Chinese Academy of Science, Beijing 100094, China. E-mail: Lizhe.Wang@gmail.com.*
- *A.Y. Zomaya is with the School of Information Technologies, the University of Sydney, Australia.*
- *D. Chen is with the School of Computer, China University of Geosciences. E-mail: Danjj43@gmail.com.*
- *R. Ranjan is with ICT/CSIRO, Australia.*

Manuscript received 12 July 2013; revised 30 Sept. 2013; accepted 10 Oct. 2013. Date of publication 22 Oct. 2013; date of current version 16 July 2014. Recommended for acceptance by J. Chen.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPDS.2013.272

Future Generation Computer Systems 51 (2015) 47–60

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Remote sensing big data computing: Challenges and opportunities

Yan Ma[a], Haiping Wu[b], Lizhe Wang[a,*], Bormin Huang[c], Rajiv Ranjan[d], Albert Zomaya[e], Wei Jie[f]

[a] *Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, PR China*
[b] *Ministry of Education Key Laboratory for Earth System Modeling and Center for Earth System Science, Tsinghua University, PR China*
[c] *Space Science and Engineering Center, University of Wisconsin-Madison, USA*
[d] *Computational Informatics, CSIRO, Australia*
[e] *School of Information Technologies, University of Sydney, Australia*
[f] *School of Computing and Technology, West London University, UK*

HIGHLIGHTS

- This paper identifies the properties and features of remote sensing big data.
- This paper reviews the stat-of-the-arts of remote sensing big data computing.
- This paper discusses the "data-intensive computing" issues in remote sensing big data processing.

ARTICLE INFO

Article history:
Received 30 July 2014
Received in revised form
19 September 2014
Accepted 25 October 2014
Available online 10 November 2014

Keywords:
Remote sensing data processing
Big data
Data-intensive computing

ABSTRACT

As we have entered an era of high resolution earth observation, the RS data are undergoing an explosive growth. The proliferation of data also give rise to the increasing complexity of RS data, like the diversity and higher dimensionality characteristic of the data. RS data are regarded as RS "Big Data". Fortunately, we are witness the coming technological leapfrogging. In this paper, we give a brief overview on the Big Data and data-intensive problems, including the analysis of RS Big Data, Big Data challenges, current techniques and works for processing RS Big Data.

1. Introduction

The recent advances in remote sensing (RS) and computer techniques give birth to the explosive growth of remote sensing (RS) data. Momentary, the observation data streaming from the spacecrafts in current active missions of NASA would approximately be 1.73 GB gigabytes [1]. The RS data gathered by a single satellite data center are dramatically increasing by several terabytes per day [2]. According to the statistics of OGC [3], the global archived observation data would probably exceed one Exabyte. Especially, the advent of the high-resolution earth observation era (EOS-4) has also led to the high dimensionality of the RS image data. Remote sensing data are recognized as "Big Data" in some certain sense [4]. Meanwhile, the accurate and up-to-date information provided by the continual global earth observation are revolutionizing the way that earth system is interpreted. The large-scale environmental monitoring and researches [5–7] are exploiting regional to global covered multi-temporal and multi-sensor RS data for processing. Obviously, large remote sensing applications overwhelmed with massive remote sensing data are regarded as typical data-intensive issues [8].

The unprecedented proliferation of data has posed significant challenges in managing, processing and interpreting these RS "Big Data". Great efforts have been towards the incorporation of the high-performance computing (HPC) paradigm in RS applications [9–11]. These HPC-based approaches have became the most dominant yet efficient way for addressing the enormous computational requirements introduced by massive RS data. However, the computation capability available is no longer the rate-limiting factor when in contrast to the traditional compute-intensive issues. Despite of the huge processing power, the cluster-based HPC systems still remain considerably challenging with the data-intensive issues [12] emerged in RS applications. There involves a number of

* Corresponding author.
E-mail address: Lizhe.Wang@gmail.com (L. Wang).

http://dx.doi.org/10.1016/j.future.2014.10.029
0167-739X/

Information Sciences 319 (2015) 171–188

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Towards building a data-intensive index for big data computing – A case study of Remote Sensing data processing

Yan Ma [a], Lizhe Wang [a,*], Peng Liu [a], Rajiv Ranjan [b]

[a] Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, PR China
[b] Computational Informatics, CSIRO, Australia

ARTICLE INFO

Article history:
Received 20 June 2014
Received in revised form 2 October 2014
Accepted 5 October 2014
Available online 25 October 2014

Keywords:
Big data
Parallel computing
Data-intensive computing
Remote Sensing data processing

ABSTRACT

With the recent advances in Remote Sensing (RS) techniques, continuous Earth Observation is generating tremendous volume of RS data. The proliferation of RS data is revolutionizing the way in which RS data are processed and understood. Data with higher dimensionality, as well as the increasing requirement for real-time processing capabilities, have also given rise to the challenging issue of "Data-Intensive (DI) Computing". However, how to properly identify and qualify the DI issue remains a significant problem that is worth exploring. DI computing is a complex issue. While the huge data volume may be one of the reasons for this, some other factors could also be important. In this paper, we propose an empirical model (DI_{RS}) of DI index to estimate RS applications. DI_{RS} here is a novel empirical model (DI_{RS}) that could quantify the DI issues in RS data processing with a normalized DI index. Through experimental analysis of the typical algorithms across the whole RS data processing flow, we identify the key factors that affect the DI issues mostly. Finally, combined with the empirical knowledge of domain experts, we formulate DI_{RS} model to describe the correlations between the key factors and DI index. By virtue of experimental validation on more selected RS applications, we have found that DI_{RS} model is an easy but promising approach.

1. Introduction

The advent of the high-resolution global Earth Observation era is revolutionizing the way in which Remote Sensing (RS) data are generated, processed, and analyzed. The latest generation of space-borne sensors are generating nearly continuous streams of massive RS imageries. These data streams are transmitted through downlink channels at a rate of several gigabits per second. The amount of RS data acquired by a single satellite data center is dramatically increasing by several TB per day [25,12]. The advances in the spatial, temporal, and spectral resolution of sensors also lead to the high dimensionality of RS imagery pixels. Meanwhile, large-scale RS applications [10,31,27,9,19,30] are exploiting multi-temporal RS data with regional to global coverage as input for processing. Obviously, Remote Sensing applications are overwhelmed with the amount of high-dimensional RS data.

With the proliferation of data, Remote Sensing data processing turns out to be extremely challenging. Issues to be considered include efficient management and rapid processing of massive RS data, pixels of high dimensionality, and especially various multi-staged algorithms of RS applications with higher complexity and intensive irregular data access patterns [21].

* Corresponding author.
E-mail address: Lizhe.Wang@gmail.com (L. Wang).

http://dx.doi.org/10.1016/j.ins.2014.10.006
0020-0255/

Future Generation Computer Systems 78 (2018) 353–368

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

pipsCloud: High performance cloud computing for remote sensing big data management and processing

Lizhe Wang [a,b], Yan Ma [b,*], Jining Yan [b], Victor Chang [c], Albert Y. Zomaya [d]

[a] School of Computer Science, China University of Geoscience, Wuhan 430074, PR China
[b] Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, PR China
[c] Xi'an Jiaotong-Liverpool University, Suzhou, PR China
[d] School of Information Technologies, University of Sydney, Australia

HIGHLIGHTS

- A Cloud-enabled HPC platform for large-scale RS applications.
- Hilbert-R^+ Tree based data indexing for optimal RS big data indexing.
- Collaborative large-scale RS workflow processing across data centers.
- Cloud-enabled virtual HPC environment with VMs and bare-metal provisioning.

ARTICLE INFO

Article history:
Received 29 February 2016
Received in revised form
25 May 2016
Accepted 12 June 2016
Available online 23 August 2016

Keywords:
Big data
Cloud computing
Data-intensive computing
High performance computing
Remote sensing

ABSTRACT

Massive, large-region coverage, multi-temporal, multi-spectral remote sensing (RS) datasets are employed widely due to the increasing requirements for accurate and up-to-date information about resources and the environment for regional and global monitoring. In general, RS data processing involves a complex multi-stage processing sequence, which comprises several independent processing steps according to the type of RS application. RS data processing for regional environmental and disaster monitoring is recognized as being computationally intensive and data intensive.

We propose pipsCloud to address these issues in an efficient manner, which combines recent cloud computing and HPC techniques to obtain a large-scale RS data processing system that is suitable for on-demand real-time services. Due to the ubiquity, elasticity, and high-level transparency of the cloud computing model, massive RS data management and data processing for dynamic environmental monitoring can all be performed on the cloud via Web interfaces. A Hilbert-R^+-based data indexing method is employed for the optimal querying and access of RS images, RS data products, and interim data. In the core platform beneath the cloud services, we provide a parallel file system for massive high-dimensional RS data, as well as interfaces for accessing irregular RS data to improve data locality and optimize the I/O performance. Moreover, we use an adaptive RS data analysis workflow management system for on-demand workflow construction and the collaborative processing of a distributed complex chain of RS data, e.g., for forest fire detection, mineral resources detection, and coastline monitoring. Our experimental analysis demonstrated the efficiency of the pipsCloud platform.

1. Introduction

Remarkable advances in high-resolution Earth observation techniques have led to explosive increases in the volume and rate of remote sensing (RS) data [1]. The latest generation space-borne sensors are capable of generating continuous streams of observation data at a rate of several gigabytes per second [2] almost every hour, every day, and every year. The current volume of globally archived observation data probably exceeds one exabyte according to statistics in an OGC report [3]. The volume of RS data acquired by a regular satellite data center has increased by several terabytes per day, especially with high-resolution missions [4]. The use of high-resolution satellites with higher spatial, spectral, and temporal resolution of data inevitably leads to higher pixel

* Corresponding author.
E-mail address: mayan@radi.ac.cn (Y. Ma).

http://dx.doi.org/10.1016/j.future.2016.06.009

Generic Parallel Programming for Massive Remote Sensing Data Processing

Yan Ma[1], Lizhe Wang[1], Dingsheng Liu[1], Peng Liu[1], Jun Wang[2] and Jie Tao[3]

[1]Center for Earth Observation and Digital Earth, Chinese Academy of Science, P. R. China.

[2] Department of Electrical Engineering & Computer Science, University of Central Florida Orlando, United States

[3]Steinbuch Center for Computing, Karlsruhe Institute of Technology, Germany

***Abstract*—Remote Sensing (RS) data processing is characterized by massive remote sensing images and increasing amount of algorithms of higher complexity. Parallel programming for data-intensive applications like massive remote sensing image processing on parallel systems is bound to be especially trivial and challenging. We propose a C++ template mechanism enabled generic parallel programming skeleton for these remote sensing applications in high performance clusters. It provides both programming templates for distributed RS data and generic parallel skeletons for RS algorithms. Through one-side communication primitives provided by MPI, the distributed RS data template could provide a global view of the big RS data whose sliced data blocks are scattered among the distributed memory of cluster nodes. Moreover, by data serialization and RMA (Remote Memory Access), the data templates could also offer a simple and effective way to distribute and communicate massive remote sensing data with complex data structures. Furthermore, the generic parallel skeletons implement the recurring patterns of computation, performance optimization and pass the user-defined sequential functions as parameters of templates for type genericity. With the implemented skeletons, Developers without extensive parallel computing technologies can implement efficient parallel remote sensing programs without concerning for parallel computing details. Through experiments on remote sensing applications, we confirmed that our templates were productive and efficient.**

***Key word*: parallel programming; generic programming; data-intensive computing; remote sensing image processing**

I. INTRODUCTION

Recently the amount of remote sensing data continuously acquired by satellite and airborne sensors has been dramatically increased [1]: typically several terabytes remote sensing data are acquired per day and the volume of single dataset is several gigabytes. Extremely massive data are required to be processed and analyzed daily. The remote sensing datasets characterized by multi-band image data structures and related geographical information used for image processing are usually organized in various formats, thus making it rather complicated and difficult for algorithms to load and reside remote sensing data. With the proliferation of data, remote sensing data processing becomes extremely challenging because of the massive image data, the pixels of higher dimensionality, and various algorithms with higher complexity, together with complicated data access modes resulted from different correlations between algorithms and remote sensing data. In particular, many time-critical applications like disaster monitoring even require real-time or near real-time processing capabilities.

The enormous computational requirements introduced by the unprecedented massive data and various complicated algorithms especially many time-critical applications have far outpaced the computing capability of single computer. It thus is an effective solution to address these computational challenges to incorporate cluster based high-performance computing (HPC) models in remote sensing image processing[2]. However, parallel cluster systems are characterized by increasing scale and multilevel parallel hierarchy. To develop efficient codes for parallel remote sensing algorithms usually involves handling data slicing and distribution, task partition, message passing among distributed memory spaces and shared memory management for multicores, synchronization and communication with low-level APIs. Furthermore, it still remains a big challenge for parallel systems to load massive datasets into memory before processing. It is also quite complicated and inefficient to exchange among cluster nodes the datasets with multi-band images and complex structured geographical information. Therefore, parallel programming on multicore clusters for processing massive remote sensing data is complicated, difficult and error-prone.

Generic parallel algorithms are abstract and recurring patterns lifting from many concrete parallel programs and conceal parallel details as skeletons. The approach of generic parallel computing relies on type genericity to resolve polymorphism at compile time, so as to reduce the runtime overhead while providing readability of high-level. To properly solve the aforementioned problems of processing massive RS data in HPC clusters, we propose the RS-GPPS (Generic Parallel Programming Skeletons) for massive remote sensing data processing applications. The RS-GPPS is enabled by the template class mechanism and work on top of MPI. We focus on so-called programming templates, which are

978-0-7695-4807-4/12 $26.00 © 2012 IEEE
DOI 10.1109/CLUSTER.2012.51

IEEE computer society

CONCURRENCY AND COMPUTATION: PRACTICE AND EXPERIENCE
Concurrency Computat.: Pract. Exper. 2012
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/cpe.2965

SPECIAL ISSUE PAPER

Distributed data structure templates for data-intensive remote sensing applications

Yan Ma[1,2,3], Lizhe Wang[1,*,†], Dingsheng Liu[1], Tao Yuan[1,3], Peng Liu[1] and Wanfeng Zhang[1,3]

[1]*Center for Earth Observation and Digital Earth, Chinese Academy of Sciences, China*
[2]*Institute of Electronics, Chinese Academy of Sciences, China*
[3]*Graduate University of Chinese Academy of Sciences, China*

SUMMARY

The remotely sensed images continuously acquired by satellite and airborne sensors are increasing dramatically. Remote sensing applications are overwhelmed with tons of remote sensing data with complex data structures. Efficient programming in parallel systems for data-intensive applications like massive remote sensing data processing will be a challenge. We propose a generic data-structure oriented programming template to support massive remote sensing data processing in high-performance clusters. These templates provide distributed abstractions for large remote sensing image data with complex data structure and allow these distributed data to be accessed as a global one. Through data serialization and one-sided message passing primitives provided by message passing interface, the distributed remote sensing data template whose sliced data blocks are scattered among nodes could offer a simple and effective way to distribute and communicate massive remote sensing data. Efficient parallel input/output directly to and from the distributed data structure will also be offered to address the input/output bottleneck caused by massive image data. Developers can take the advantage of our templates to program efficient parallel remote sensing algorithms without dealing with data slicing and communication through low-level message passing interface APIs. Through experiments on remote sensing applications, we confirmed that our templates were productive and efficient. Copyright © 2012 John Wiley & Sons, Ltd.

Received 12 June 2012; Revised 6 August 2012; Accepted 31 October 2012

KEY WORDS: parallel programming; generic programming; data-intensive computing; remote sensing image processing

1. INTRODUCTION

With the rapid development of remote sensing technologies, the amount of remote sensing images has been dramatically increasing by terabytes per day, and the sheer volume of single satellite image is several gigabytes. Huge quantities of data are required to be processed and analyzed daily. The remote sensing image datasets are characterized by multi-band image data structure which would lead to the pixel's multi-dimensionality and the closely related geographical information which are always involved in image processing. To make the situation even more complicated, the datasets are usually organized in various formats, thus making it rather complicated and difficult for algorithms to load and store remote sensing data into local memory. With the proliferation of data, remote

*Correspondence to: Correspondence to: Lizhe Wang, Center for Earth Observation and Digital Earth, Chinese Academy of Sciences, China.
†E-mail: lizhe.wang@gmail.com

Noname manuscript No.
(will be inserted by the editor)

Dynamic DAG scheduling for many-task computing of distributed eco-hydrological model

Shasha Yue · Yan Ma* · Lajiao Chen · Yuzhu Wang · Weijing Song

Received: date / Accepted: date

Abstract The computing of distributed hydrological model at large scale is increasingly characterized by data-intensive and computation-intensive, especially for the multi-process coupling model. Parallel computing is one effective approach to cope with this situation. The easily extensible fine-grained parallelization method can substantially improve the computing efficiency. Base on many-task computing, we proposed a parallel scheme that the whole computing of the distributed hydrological model is split into tremendous amount of small subtasks which are directly dispatched into the cluster by traditional local resource managers (LRMs). The task-splitting method, the single task model and the representation of dependencies between tasks are also proposed. In order to efficiently schedule so many tasks, a dynamic DAG scheduling method based on critical path and depth is provided. The management of intermediate file, the control strategy for LMRs and the fault recovery is also introduced to deal with the problems encountered in the actual parallel implementation process. The parallel scheme is tested with an optimality-based

Shasha Yue
Institute of Remote Sensing and Digital Earth, Chinese Academy of Science, Beijing 100094, P. R. China
Tel.: +86-010-82178082
E-mail: shashayue@hotmail.com

Yan Ma
Institute of Remote Sensing and Digital Earth, Chinese Academy of Science, Beijing 100094, P. R. China
Tel.: +86-010-82178080
E-mail: mayan@radi.ac.cn

Lajiao Chen, Yuzhu Wang
Institute of Remote Sensing and Digital Earth, Chinese Academy of Science, Beijing 100094, P. R. China

Weijing Song
School of Computer Science, China University of Geoscience, Wuhan 430074, P. R. China

证书号第1347257号

发明专利证书

发明名称：面向遥感图像处理算法的快速并行化编程模板方法

发　明　人：马艳；刘定生；李景山；赵灵军；刘鹏；王妍颖

专　利　号：ZL 2010 1 0140761.9

专利申请日：2010年04月07日

专利权人：中国科学院对地观测与数字地球科学中心

授权公告日：2014年02月19日

本发明经过本局依照中华人民共和国专利法进行审查，决定授予专利权，颁发本证书并在专利登记簿上予以登记。专利权自授权公告之日起生效。

本专利的专利权期限为二十年，自申请日起算。专利权人应当依照专利法及其实施细则规定缴纳年费。本专利的年费应当在每年04月07日前缴纳。未按照规定缴纳年费的，专利权自应当缴纳年费期满之日起终止。

专利证书记载专利权登记时的法律状况。专利权的转移、质押、无效、终止、恢复和专利权人的姓名或名称、国籍、地址变更等事项记载在专利登记簿上。

局长
申长雨

申长雨

中华人民共和国国家知识产权局
2014年02月19日

北京市海淀区地方税务局 印花税代扣专用章 国家知识产权局

中华人民共和国国家知识产权局

100193
北京市海淀区东北旺西路8号中关村软件园8号楼227B 北京纽乐康知识产权代理事务所（普通合伙）
田磊

发文日：

2016年03月29日

申请号或专利号：201310070436.3　　　　发文序号：2016011901044920

申请人或专利权人：中国科学院对地观测与数字地球科学中心

发明创造名称：一种基于任务树的大尺度遥感影像并行镶嵌方法

授予发明专利权通知书

1.根据专利法第 39 条及实施细则第 54 条的规定，上述发明专利申请经实质审查，没有发现驳回理由，现作出授予专利权的通知。

申请人收到本通知书后，还应当依照办理登记手续通知书的内容办理登记手续。

申请人按期办理登记手续后，国家知识产权局将作出授予专利权的决定，颁发发明专利证书，并予以登记和公告。

期满未办理登记手续的，视为放弃取得专利权的权利。

2.授予专利权的上述发明专利申请是以下列申请文件为基础的：

☐原始申请文件。☐分案申请递交日提交的文件。☒下列申请文件：

申请日提交的说明书摘要、说明书第 1-20 段、摘要附图、说明书附图；

2016 年 1 月 5 日提交的权利要求第 1-3 项。

3.授予专利权的上述发明专利申请的名称：

☒未变更。

☐由__变更为上述名称。

4. ☐申请人于______年______月______日提交专利号为______的"放弃专利权声明"，经审查：

☐进入放弃专利权的程序。

☐未进入放弃专利权的程序。理由是：申请人声明放弃的专利与本发明专利申请不属于相同的发明创造。

5. ☐审查员依职权对申请文件修改如下：

6.在本通知书发出后收到的申请人主动修改的申请文件，不予考虑。

审查员：狄晓斐　　　　审查部门：专利审查协作江苏中心电学发明审查部

联系电话：0512-88995724

(19) 中华人民共和国国家知识产权局

(12) 发明专利

(10) 授权公告号 CN 101814029 B
(45) 授权公告日 2013.11.27

(21) 申请号 201010150738.8

(22) 申请日 2010.04.20

(73) 专利权人 中国科学院对地观测与数字地球科学中心
地址 100000 北京市海淀区北三环西路45号

(72) 发明人 赵灵军 刘定生 李景山 马艳 汤灿恩

(74) 专利代理机构 北京纽乐康知识产权代理事务所 11210
代理人 田磊

(51) Int.Cl.
G06F 9/44 (2006.01)
G01S 7/48 (2006.01)

(56) 对比文件
US 2001/0011366 A1,2001.08.02, 全文.
US 4477880 A,1984.10.16, 全文.
CN 101221243 A,2008.07.16, 全文.

审查员 李娇

权利要求书1页 说明书7页 附图3页

(54) 发明名称

遥感影像处理系统中处理功能可快速扩展的构建方法

(57) 摘要

本发明涉及一种遥感影像处理系统中处理功能可快速扩展的构建方法,包括以下步骤:1)按照指定接口规范的参数文件开发算法程序;2)通过遥感图像处理系统进行可视化的算法程序注册;3)业务功能组合注册;4)动态提交任务与任务调度执行。本发明的有益效果为:在不修改遥感图像处理系统本身的情况下,用户只要按照指定的规范开发自己的算法程序并进行算法注册和组合,就可以扩展新的业务功能;并且用户无需开发遥感图像处理中常用的人工交互获取参数界面,只需要在注册算法模块时指定参数的获取方式,就可以利用系统动态生成的任务提交界面,进行业务功能的处理操作。

CN 101814029 B

(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 101315424 B
(45)授权公告日 2011.06.08

(21)申请号 200810135047.3

(22)申请日 2008.07.29

(73)专利权人 中国科学院对地观测与数字地球科学中心

地址 100086 北京市海淀区北三环西路 45 号

(72)发明人 刘定生 李国庆 章文毅 李景山 赵灵军 马艳 王建 陈甫 黄克颖 杨进 黄鹏 黄方 马广彬 于文洋 张万军

(74)专利代理机构 北京纽乐康知识产权代理事务所 11210

代理人 田磊

(51)Int.Cl.

G01S 7/48(2006.01)

G01S 17/00(2006.01)

(56)对比文件

CN 1831555 A,2006.09.13,

WO 2006/134329 A2,2006.12.21,

CN 1924610 A,2007.03.07,

CN 1831556 A,2006.09.13,

WO 2006/095170 A1,2006.09.14,

WO 2006/064051 A1,2006.06.22,

安兴华，王小鸽，都志辉，刘定生，李国庆．一种适用于机群系统的细粒度遥感图像镶嵌并行算法．《清华大学学报（自然科学版）》.2002,第 42 卷（第 10 期）,1389-1392.

王明远，刘长柱．亚洲一流的遥感卫星地面应用系统．《空间电子技术》.2002,（第 2 期）,49-60,64.

SPOT 项目组．法国遥感卫星 SPOT 数据 1A 级预处理技术．《宇航学报》.1994, 第 15 卷（第 4 期）,71-74,102.

刘蓉蓉，林子瑜．遥感图像的预处理．《吉林师范大学学报（自然科学版）》.2007,（第 4 期）,6-10.

牟胜梅，刘光明．海量遥感图像数据实时预处理系统结构设计．《计算机工程与科学》.2004, 第 26 卷（第 10 期）,41-44,55.

陆松，刘光明．分布共享存储的遥感图像并行预处理系统结构研究．《计算机工程与科学》.2004, 第 26 卷（第 10 期）,56-59.

王坤龙，刘定生，章文毅，李国庆．卫星数据深加工处理流程自动化分析、设计与实现．《遥感技术与应用》.2005, 第 20 卷（第 3 期）,355-360.

谢光益．基于层次数据格式海量数据预处理

（续）

审查员 张静

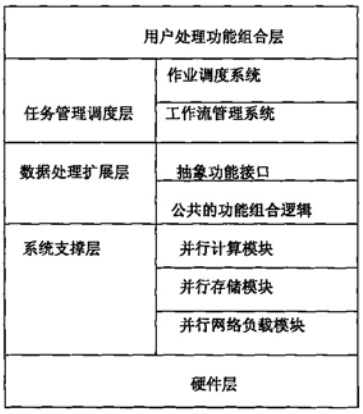
权利要求书 1 页 说明书 3 页 附图 1 页

(54)发明名称

多卫星遥感数据一体化并行地面预处理系统

(57)摘要

本发明涉及一种多卫星遥感数据一体化并行地面预处理系统，包括硬件层、系统支撑层、数据处理扩展层，任务管理调度层和用户处理功能组合层；其中硬件层提供预处理系统的硬件支撑；系统支撑层提供系统软件支撑环境；数据处理扩展层具有多种卫星多种传感器的抽象功能接口和公共的功能组合逻辑；任务管理调度层实现任务调度和流程管理；用户处理功能组合层为用户提供简单易用的流程组合界面。本发明有益效果为：用户在不修改处理程序的情况下便可方便均衡的扩展系统的处理、存储、网络通信能力，提高程序的运行速度；有新的卫星遥感数据进入系统需要处理、归档、产品生产和分发时只需将相应数据的处理模块根据一定的模板加入系统中即可完成功能的扩展。



［转续页］

CN 101315424 B

2017年度“清华——浪潮地球系统科学青年人才奖”推荐书

被推荐人姓名:闵敏　　专业技术职务:副研究员　　专业:大气物理学与大气环境

出生年月:[redacted]

工作单位:国家卫星气象中心

电子信箱:[redacted]　　电话:[redacted]

我郑重推荐闵敏作为“清华—浪潮地球系统科学青年人才奖”的候选人。

推荐人签名:李俊

推荐人单位公章:

2017年11月17日

推荐人姓名:李俊　　专业技术职务:研究员　　专业:大气物理学与大气环境

工作单位:国家卫星气象中心

电子信箱:[redacted]　　电话:[redacted]

推荐身份:☐中国科学院院士　☐中国工程院院士　☐国家和部委重点实验室主任

☐重点高等院校校长　☐长江学者　☐重点科研院所院(所)长

☐杰出青年基金获得者 ☑千人计划学者 ☐国家重大科技项目首席科学家

请简要说明被推荐人的原创性科学或技术成果的内容、重大科学意义、对促进科学技术发展所起的作用及产生的影响(限2000字以内,纸面不够可加页),相关证明材料和论文原文以附件方式提供。

从卫星观测中准确提取定量地球物理参数信息，得到科学数据产品是形成地球系统科学观测数据的重要一步，这些科学数据对于认识和理解地球系统科学问题，数值模式驱动，参数化方案建立和模式检验极为重要。根据对新一代地球静止轨道气象星观测的科学应用需求，闵敏博士设计并主导开发了具有自主知识产权的新一代地球静止轨道气象卫星（简称静止气象卫星，例如日本葵花8号，中国风云四号）科学数据算法软件和数据处理系统平台（FYGAT）。该平台以静止气象卫星遥感理论和反演算法为基础，具有综合集成和测试各类算法，并实时生成多种地球系统科学数据（云，大气温湿度，气溶胶，辐射，地表等）的能力。该系统主要的科学价值和意义体现在：(1) 作为新一代静止气象卫星科学算法研发和数据生成的平台，可以对算法系统和科学数据版本作改进和升级；(2) 是我国新一代风云静止气象卫星-风云四号业务产品生成系统的原型；(3) 可以实现国际上所有新一代静止气象卫星（中美日欧）观测数据的实时处理，得到 30 多种地球系统科学数据产品，且其实时性极大的提高了研究到应用（R2A）的转化效率和科学数据的应用效果；(4) 可以对静止气象卫星历史数据进行再处理和分析，为全球气候环境变化研究提供高质量的长时间序列科学数据集。

自 2011 开始研制，经过 5 年的努力，FYGAT 系统于 2016 年 6 月研制成功，并发布第一版投入实时应用。**最近，FYGAT 系统还被设计成为我国自主研发大科学装置地球模拟器“寰”的重要支撑技术，为该地球系统模式提供各种近实时的气象卫星科学数据产品，用于模式同化，模式性能的检验和评估等。FYGAT 系统在 Linux 系统下采用 Fortran 和 C/C++混编技术实现，并采用多线程技术实现多种产品的并行处理能力。目前，该系统在国家卫星气象中心被部署在 5 台浪潮超级服务器上（共享一套 NAS 架构存储），对葵花 8 号和风云四号气象卫星观测数据进行实时处理，生成 30 多种地球系统科学数据产品。而美国同类应用系统 CSPP-GEO 目前仍然处于研制和测试阶段，尚未投入应用**（见 CSPP-GEO 官方网站在 2017 年 6 月 28 日发布的科学软件应用计划 PPT，其 Version1.0 将于 2017 年年底投入使用，且第一版只提供有限种类科学数据产品)。

目前，FYGAT 已在国家气象中心、公共气象服务中心、广东省、上海市、陕西省等省一级气象部门推广应用，实现了葵花 8 号定量科学产品的实时生成和应用**（见应用证明 6 份）**。此外，FYGAT 最近还被成功的用于我国新一代地球静止气象卫星 FY-4A 在轨测试，及时发现并解决了多个技术问题（例如：云检测异常、气溶胶反演偏差问题、红外通道条带等等）。FYGAT 系统及其所生成地产品目前正在全国气象系统进一步推广和应用，多家国家和省一级气象业务部门都把该平台作为重要的业务或应用系统，FYGAY 实时生成的定量科学数据产品在天气预报和环境气象监测，预测，预警等地球系统科学领域获得广泛应用。

闵敏，男，1982 年 12 月出生，2005 年毕业于南京信息工大学获得学士学位，2010 年毕业于中国科学院大气物理研究所获得博士学位。目前是中国气象局国家卫星气象中心副研究员（2017 年作为研究员被卫星中心推荐到局里），并担任风云三号地面系统科学仿真副主任设计师，风云三/四号云光学和风云三号云检测产品责任人。负责新一代地球静止气象卫星科学数据算法平台的研制，开发风云三号卫星红外定标在轨跟踪算法及成像仪图像空间质量在轨恢复和算法评估，研究卫星资料在云和气溶胶气候评估中的应用。目前，已经在 JGR, IEEE-TGRS 等国内外知名刊物发表科研论文 30 多篇，其中 SCI 论文近 20 篇。2015 年入选中国气象局“青年英才计划”，2016 年获北京大学“谢义炳青年气象科技奖”，2017 年获中国气象学会颁发的“涂长望青年气象科技奖”。

2017年度“清华——浪潮地球系统科学青年人才奖”推荐书

被推荐人姓名: 吴波 专业技术职务: 副研究员 专业: 大气科学

出生年月: [redacted]

工作单位: 中国科学院大气物理研究所

电子信箱: [redacted] 电话: [redacted]

我郑重推荐 吴波 作为“清华——浪潮地球系统科学青年人才奖”的候选人。

推荐人签名:

推荐人单位公章:

2017年12月6日

推荐人姓名: 周天军 专业技术职务: 研究员 专业: 大气科学

工作单位: 中国科学院大气物理研究所

电子信箱: [redacted] 电话: [redacted]

推荐身份: ☐中国科学院院士 ☐中国工程院院士 ☐国家和部委重点实验室主任

☐重点高等院校校长 ☐长江学者 ☐重点科研院所院（所）长

√☐杰出青年基金获得者☐千人计划学者 ☐国家重大科技项目首席科学家

请简要说明被推荐人的原创性科学或技术成果的内容、重大科学意义、对促进科学技术发展所起的作用及产生的影响（限2000字以内，纸面不够可加页），相关证明材料和论文原文以附件方式提供。

吴波副研究员的研究方向是气候模拟，围绕“热带海-气相互作用”和“年代际气候预测”开展了系统的研究，在理论和应用研究两方面都取得了多项创新性成果，成果具有广泛的国际影响力。迄今为止，吴波博士已发表论文52篇，其中第一作者和通讯作者SCI论文14篇（气候领域顶级期刊Journal of Climate 8篇）；论文被SCI引用1034次。两篇第一作者论文SCI引用破百（单篇最高138次）。他完成的基于气候系统模式的年代际预测试验结果和相关论文被政府间气候变化专门委员会（IPCC）第五次评估报告（AR5）第11章引用。先后获得“全国百篇优秀博士学位论文（2011年）”、“中国科学院青年创新促进会会员（2012年）”、“学笃风正创新贡献奖（2016年）”等荣誉。

吴波博士取得了多项在国际上有相当影响力的研究成果。在“海气相互作用和季风”方面，给出了ENSO在不同位相对东亚季风影响的完整图像；提出印度洋洋盆模态强迫东亚-西北太平洋季风的机制及其季节依赖性；指出西北太平洋存在两种典型的大气强迫海洋的机制；发现在ENSO冷、暖位相，西北太平洋大气环流异常的非对称性；提出了西北太平洋异常反气旋产生和维持的新机制。这些研究工作的一个显著特点，就是需要结合数值模式，开展大规模数值实验，结果对理解热带大气动力学和暖池区海气相互作用过程，及其对东亚气候年际变率的影响有重要意义。

在年代际预测方面，吴波博士搭建了国内首个基于耦合气候系统模式的年代际预测系统，并发表了国内第一篇关于年代际预测问题的学术论文；他所构建的海洋同化系统，目前已经应用到中国气象科学研究院和国家气候中心的模式系统中。下面重点介绍吴波基于气候系统模式、通过数值模拟所开展的理论和应用研究成果。

1）提出El Nino影响东亚气候的新机制--“风-湿焓平流强迫机制”；开发了基于耦合气候系统模式的海洋观测数据同化系统，并使用该系统设计数值试验验证了所提出的新机制。

西北太平洋异常反气旋（WNPAC）将El Nino和东亚季风联系起来，是东亚气候季节可预报性的主要来源。对WNPAC相关过程的探索是东亚-西北太平洋季风年际变率研究的核心。现有解释WNPAC在El Nino成熟冬季和之后春季维持的核心理论是经

典的西北太平洋局地“风-蒸发-SST”反馈机制。但是，近年来研究发现了一些新事实，用“风-蒸发-SST”反馈机制无法解释。吴波博士从湿静力能收支诊断分析入手，提出了一种全新的 WNPAC 维持机制——“风-湿焓平流强迫机制”。吴波开发了基于 EnOI-IAU 同化方法和耦合气候系统模式 FGOALS 的海洋同化系统，基于该同化系统开展了一系列数值试验，结果验证了“风-湿焓平流强迫机制”，同时证明了该机制对 El Niño 冬季和春季 WNPAC 的贡献超过 50%，略大于西北太平洋局地海气相互作用的贡献。相关结果以两篇系列论文发表在 J. Climate （Wu et al. 2017a，b）。“风-湿焓平流强迫机制”的证明借助的是中国自己研发的气候系统模式，这一工作彰显了中国气候模式对完善国际热带海气相互作用理论的重要贡献。

El Niño 衰减年的夏季，WNPAC 持续影响东亚一西北太平洋夏季风。2009 年以来许多研究强调，热带印度洋“电容器效应”是维持 WNPAC 的关键机制。吴波使用大气环流模式，设计理想数值试验证明了夏季 WNPAC 是局地 SST 负异常和热带印度洋遥强迫共同作用的结果，其中前者主要在 6 月起作用，而后者则在 7、8 月份起作用。由于梅雨锋出现在初夏，西北太平洋局地 SST 负异常可能对东亚夏季风季节预测更有意义。该结果发表在 Wu et al.（2010, J. Climate），已被 SCI 引用 138 次。

2）基于耦合气候系统模式搭建年代际预测系统，试验结果参加了 CMIP5 国际计划，被 IPCC AR5 引用

年代际预测试验是耦合模式比较计划第五阶段（CMIP5）的核心试验之一，并为 IPCC AR5 近期气候预测提供科学依据。吴波使用“Incremental Analysis Updates”（IAU）同化方案，为耦合气候系统模式 FGOALS-gl 建立了初始化系统，完成了国内首次基于模式的年代际预测试验。相关结果在科学通报中、英文版发表。他基于相似的同化系统，完成了 IAP/LASG 耦合模式 FGOALS-s2 的年代际预测试验，试验结果提交了 CMIP5。他随后完成的第二代年际-年代际预测系统，其初始化部分采用新发展的 EnOI-IAU 海洋数据同化方案（EnOI 是集合最优插值的缩写）。

目前，年代际预测试验虽然对太平洋年代际振荡（PDO）和大西洋多年代际振荡（AMO）具有某些预报技巧，但对东亚季风年代际变化的预测仍然困难。以理解 PDO 和 AMO 如何调制东亚气候、发展“动力-统计”相结合的东亚气候年代际预测系统为目标，他在国际上率先发现将东亚夏季风年代际变率与 PDO 和 AMO 联系起来的是“日本-太平洋遥相关（PJ）”和“环球遥相关（CGT）”型，并指出存在年代际尺度的 CGT 遥相关型。上述成果发表在 Wu and Zhou（2012, Chin. Sci. Bull.）和 Wu et al.（2016, J.

Climate）。Wu and Zhou（2012）及 FGOALS-s2 年代际预测试验结果被 IPCC AR5 第 11 章引用。

综上，我认为吴波副研究员的工作堪称数值模拟和理论研究相结合的典范。他的研究成果，对于提升中国气候模式的国际影响力和加深对季风可预报性理论的理解，起到双重推动作用。因此，我郑重推荐吴波博士参评 2017 年度“清华——浪潮地球系统科学青年人才奖”。

2017年度“清华——浪潮地球系统科学青年人才奖”推荐书

被推荐人姓名:_俞乐______专业技术职务:_副教授____专业:_地球系统观测____

出生年月: [redacted]

工作单位: ____清华大学地球系统科学系________________________________

电子信箱: [redacted] 电话: [redacted]

我郑重推荐_俞乐________作为“清华——浪潮地球系统科学青年人才奖”的候选人。

推荐人签名: 宫鹏

推荐人单位公章: 清华大学地球系统科学系

2017年12月7日

推荐人姓名:__宫鹏_________专业技术职务:__教授______________ 专业:_地球系统观测_______________

工作单位: ___清华大学地球系统科学系

____

电子信箱: [redacted] 电话:

推荐身份: ☐中国科学院院士 ☐中国工程院院士 ■国家和部委重点实验室主任

☐重点高等院校校长 ☐长江学者 ☐重点科研院所院（所）长

■杰出青年基金获得者☐千人计划学者 ☐国家重大科技项目首席科学家

请简要说明被推荐人的原创性科学或技术成果的内容、重大科学意义、对促进科学技术发展所起的作用及产生的影响（限 2000 字以内，纸面不够可加页），相关证明材料和论文原文以附件方式提供。

俞乐博士瞄准全球农地分布制图这一国际研究热点和难点，针对基于多源遥感数据的全球农地制图方法和不确定性、全球制图产品中的尺度效应等科学和技术问题，开展了一系列的原创研究，取得了若干重要的、具有国际影响力的创新性成果。共发表 SCI 论文 55 篇，被 Science, PNAS, Nature Climate Change 等期刊论文引用 958 次，其中他引 639 余次（截止 2017 年 10 月 24 日），有 1 篇文章被 Essential Science Indicators (ESI) 数据库列为"高引用论文"。相关成果包括：

(1) 建立了基于多源数据的全球农地分布遥感制图框架

以全球为研究对象，系统性的建立了全球农地制图框架：包括样本采集、多源遥感数据快速处理 / 集成、结合专题产品和社会统计资料、基于机器学习算法的农地特征识别和提取、不确定性分析等。完成了当前世界上最高空间分辨率(30 米)的全球农地分布图，该套农地产品与以往相比，空间分辨率提高了 1 个数量级。该数据与世界粮农组织(Food and Agriculture Organization of the United Nations，FAO)收集的同年各国国家统计数据的回归系数 R^2=0.96，较同年 FAO 统计数差别仅为 0.02%，在空间分辨率和精度上均取得了重要的进步(Yu et al., 2013a)。

采用多源遥感数据，包括 250 米 MODIS 时间序列，全球 DEM、生物气候数据，土壤水分条件数据等改进了 30 米全球土地覆盖产品中的农地分类精度，使农地类型精度提高 10%(Yu et al., 2013b)。首次实现了集成多源多分辨率遥感数据开展全球 30 米制图工作。开发的多源遥感数据快速处理技术和以分割斑块替代像素作为基本计算单元的策略，节省了上百倍的计算时间，克服了多源多分辨率遥感数据快速处理的难题。

进一步分析了农地制图在空间分布、面积估算和未来情景模拟的不确定性。总结了非洲、南美洲、东南亚和大洋洲与 FAO 统计资料差异较大的区域(多云多雨地区、小田块区(如梯田)、水稻种植区是制图产品面积估算误差的主要区域)，指明了这些地区是将来工作的重点区域(Yu et al., 2013a)。

并对未来四种主要的气候变化情景下(RCP 2.6, RCP 4.5, RCP 6.0 和 RCP 8.5)的土地利用开展了 30 米分辨率下的模拟，分析了农地模拟在时空上的不确定性分布，该数据在空间分辨率(30m)和时间跨度(2010-2100)上均处于世界领先水平(Li, Yu*, et al., 2016)。

(2) 多分辨率全球制图产品研发及应用

分析了土地覆盖制图产品误差的时空分布和尺度效应。提出将验证样本点按空间位置对应到生态地理特征较为一致的生态区进行精度分析，得到误差分布图。这给用户选择根据各自的研究区选择数据提供了依据。还研制了全球土地覆盖多分辨率数据集，分析了全球土地覆盖制图中的由尺度引起的面积估算偏差，通过多尺度分析为模型和应用研究中最优分辨率的选择提供了科学依据。基于此项分析，申请人研制了全球首套多分辨率

(30 米，250 米，500 米，1 千米，5 千米，10 千米，50 千米，100 千米)全球土地覆盖数据（FROM-GLC-Hierarchy），用以满足数据用户的不同分辨率需求(Yu et al., 2014a)。这一工作对地球系统模式和气候模式等研究都很有价值,已经服务于国家气候中心 BCC_AVIM 陆面模式的研究工作(史学丽等, 2016)。

系统分析了当前全球土地覆盖制图领域的进展。提出文献空间化概念并构建了空间化文献数据库，总结了采用遥感数据，尤其是 30 米空间分辨率的数据全球制图热点区和空白区，并提出制图中的输入特征(光谱、纹理等)比分类器对分类结果的影响更大(Yu et al., 2014b)。指出了当前的土地覆盖制图领域对于数据应用的关注还存在巨大不足,如仅有 18%的生物多样性保护的热点区是全球土地覆盖文献的研究热点区域。研究揭示了以生物多样性保护为目的的全球关键陆地生态区范围内的连续观测数据和土地覆盖制图研究的数量和质量都还存在巨大的不足，并提出了多个切实可行的解决方案(Yu et al., 2015)。另一项延续性的工作，开展了全球制图的特征选择研究，提出了新的特征(如植被气候响应特征)在制图中的重要价值(Yu et al., 2016)。

俞乐博士应邀参加了联合国环境规划署(The United Nations Environment Programme, UNEP)、USGS、地球观测组织(Group on Earth Observations, GEO)、资源发展制图区域中心(The Regional Centre for Mapping of Resources for Development, RCMRD)、全球森林与土地覆盖动态观测工作组(Global Observations of Forest and Land Cover Dynamics, GOFC-GOLD)等多个国际机构的专题研讨会。目前正在应邀参与多个国内外土地覆盖遥感制图工作，如在“GEO 非洲土地覆盖工作组”中担任技术咨询专家，为开发新一代非洲土地覆盖数据提供技术支持。并在科技部国家遥感中心组织的“生态环境遥感监测年报”编写中作为主要作者之一，参与非洲和“一带一路”土地覆盖制图工作。