大学生创新训练项目申请书

项目编号

项目名称 基于卫星遥感和机器学习的云参数预测降水研究

项目负责人 李欣艳/陈绎冰 联系电话 **19850078165/15282570351**

所在学院 大气物理学院

学 号 201883300199/201883300371专业班级 大气探测18级1班

指导教师 杨元建

E-mail  yyj1985@nuist.edu.cn

申请日期  **2020.05.06**

起止年月 2020.05 至 2021.05

南京信息工程大学

**填 写 说 明**

1、本申请书所列各项内容均须实事求是，认真填写，表达明确严谨，简明扼要

2、申请人可以是个人，也可为创新团队，首页只填负责人。“项目编号”一栏不填。

3、本申请书为大16开本（A4），左侧装订成册。可网上下载、自行复印或加页，但格式、内容、大小均须与原件一致。

4、请负责人所在学院认真审核。

* 1. 基本情况

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 项目  名称 | | 基于机器学习和卫星遥感云参数的降水反演与预测研究 | | | | | | | | | | | | | |
| 所属  学科 | | 学科一级门： | | | 理学 | | | | | 学科二级类： | | | 大气科学类 | | |
| 申请  金额 | | 3000 元 | | | | | 起止年月 | | 2020 年 5 月 至 2021 年 5 月 | | | | | | |
| 负责人  姓名 | | 李欣艳 | | 性别 | | | 女 | 民族 | 汉 | | | 出生年月 | | | 2000年6月 |
| 学号 | | 201883300199 | | 联系  电话 | | | 宅： 手机:19850078165 | | | | | | | | |
| 指导  教师 | | 杨元建 | | 联系  电话 | | | 宅： 手机: 18055108292 | | | | | | | | |
| 负责人曾经参与科研的情况 | | | 无 | | | | | | | | | | | | |
| 指导教师承担科研课题情况 | | | 目前主要从事卫星遥感及其在天气、气候与环境变化中的应用研究先后主持和参与国家重点研发计划项目专题、国家自然基金、中德国际合作项目、中国气象局气候变化专项等项目10余项，多项成果在气象部门应用和推广。目前正在承担国家重点研发计划项目“全球气象卫星遥感动态监测、分析技术及定量应用方法及平台研究”的专题项目 “卫星遥感云参数和降水研究”。 | | | | | | | | | | | | |
| 指导教师对本项目的支持情况 | | | 指导教师杨元建将对本项目进行全面技术把关与指导，并将从研发专项给予一定的经费资助论文出版费和用于学术交流的会议差旅，确保本项目的顺利实施和高质量完成。 | | | | | | | | | | | | |
| 项  目  组  主  要  成  员 | 姓 名 | | 学号 | | | 专业班级 | | | | | 所在学院 | | | 项目中的分工 | |
| 陈绎冰 | | 201883300371 | | | 大气探测18级1班 | | | | | 大气物理学院 | | | 构建机器学习模型 | |
| 赵文川 | | 201883300252 | | | 大气探测18级1班 | | | | | 大气物理学院 | | | 卫星、高密度自动站数据的收集与处理 | |
| 范荣峰 | | 201883300250 | | | 大气探测18级1班 | | | | | 大气物理学院 | | | 基于数据，分析典型个例和空间分布特征 | |
| 李天依 | | 201883300198 | | | 大气物理18级1班 | | | | | 大气物理学院 | | | 云微物理过程分析，数据可视化处理 | |

* 1. 立项依据（可加页）

|  |
| --- |
| 1. **项目简介**   本项目以中国东部夏季降水为研究对象，利用TRMM PR和VIRS同步探测的降水信息和光谱信息的融合资料，研究对流降水和层云降水的云顶可见光、红外信号、云参数的时空分布特征；通过机器学习（随机森林）建模，实现基于卫星光谱及云参数的降水反演和预测，并给出因子重要性排序，并将该模型推广应用到静止卫星上；同时，对比开展基于静止卫星光谱和高密度的雨量计的机器学习建模研究，旨在提升静止卫星光谱反演、预测和监测降水的能力。   1. **研究目的**   本项目以TRMM PR和VIRS同步观测的融合资料为基础，重点给出层状云和对流云的云参数和降水的相关性分析，建立云参数和降水强度的关系，并利用卫星的降水、光谱资料和地面高密度自动站的雨强观测值，通过两种不同方案构建给出最优卫星云参数的降水预测的机器学习模型，旨在进一步提升新一代静止卫星光谱云参数反演、预测和监测降水的能力，为防灾减灾提供科技支撑。   1. **研究内容**   **基于静止卫星光谱-云参数以及物理量的机器学习建模研究。**选取近几年风云四号卫星（FY-4）、葵花8号(HIMAWARI-8)卫星的降水资料、光谱资料以及地面高密度自动站的降水资料，基于机器学习建模（随机森林）研究云参数（云顶高度、云顶温度、云光学厚度、云粒子有效半径等）与降水的关系。   1. **国、内外研究现状和发展动态**   综合对比目前降水监测特别是对暴雨的监测方法：传统雨量计精度高，但空间分布不均，特别在暴雨多发的山区，往往密度较低，对大范围降水的表征能力较差。地基测雨雷达对降水的有效探测距离多在300 km 之内，故对大范围降水具有很好的监测能力。但是，地基的降水测量方式不能用于海洋上，也不能用于人迹罕至的地方且经常受到回波削弱、回波过冲以及地形山脉产生虚假回波的影响。随着卫星遥感技术不断发展，可见光/红外光谱扫描仪已成为目前最为广泛的星载探测仪器，尤其是静止卫星上的可见光/红外仪器可以提供大范围高时空分辨率的探测资料。利用卫星观测的可见光/红外信号进行降水云和非降水云的辨别，并在此基础上利用统计关系来估算地表降水率，一直都是卫星遥感降水研究中的重要课题（王晨曦等，2010；周毓荃等. 2011；傅云飞等，2011，2012）。然而，**对于静止卫星等多光谱卫星在降水监测方面，面临的一个重要问就是很难在时空同步上来获取光谱信号和降水信息（以往的研究多以地面降水观测资料来验证卫星光谱信号估算的降水），因而时空上均存在较大误差，不能准确发展卫星光谱估算降水的方法。**  国内外关于云特征参数与降水关系的研究，已经有了一些进展。例如，刘健等（2007）研究了FY-1D和NO- AA极轨卫星反演得到的云光学厚度和地面降水数据，研究发现雨量基本与云光学厚度呈正相关关系，即：地面观测不到降水，对应云的光学厚度很小；地面雨量大，对应的光学厚度较大。张杰等（2006）分析了MODIS云参数与地面降水量的关系，研究结果表明，祁连山区产生较大降水的云粒子有效半径在6~12μm之间，云光学厚度在8~20之间。Rosen-feld and Gutman（1944）研究了NOAA卫星反演的云粒子有效半径与降水的关系，提出有效半径大于14μm是云中产生降水的阈值。傅云飞（2014）研究了利用卫星双光谱反射率算法反演的云参数及其在气溶胶-云-降水相互作用以及降水预测方面进行了应用。周毓荃等（2008）利用风云静止卫星的观测资料，融合其他多种观测资料，反演了云顶高度、云顶温度、云光学厚度、云粒子有效半径等近10种云宏微物理特征参数及其与降水的关系。可见，上述结果表明**利用卫星遥感云参数的统计反演降水方面具有重要的科学意义和应用价值。**  然而由于地面降水监测站点在空间上的稀疏，对卫星光谱遥感反演降水带来很大的不确定性和挑战。1997年底美日共同研制的热带测雨（Tropical Rainfall Measuring Mission，简称TRMM）卫星的发射升空，利用其搭载的测雨雷达（Precipitation Radar，简称PR）、微波成像仪（TRMM Microwave Imager，简称TMI）和可见光/红外辐射计（Visible and Infrared Sensor，简称VIRS）等仪器，为研究热带、副热带地区的降水云光谱信号特征及其与降水的关系提供了新的契机。  **首先，TRMM PR能够对降水类型及降水的三维结构（水平和垂直结构）进行识别和捕获。**截止目前，TRMM PR多时空尺度不同类型降水（对流降水和层云降水）的研究方面得到了广泛应用，并取得了丰富的成果(e.g., Liu and Fu, 2001; Kodama et al.,2002; Fu and Liu, 2001, 2003;Schumacher and Houze, 2003,2006; Li et al., 2011; Fu et al., 2006, 2007; Yamamoto et al., 2008; Fu et al.,2010; Yu et al.,2010; Li and Schumacher, 2011；Yang et al., 2014).这些成果对云和降水的微物理、热动力特征以及降水潜热与大气环流方面的研究具有重要的推进作用，特别是为我们东亚降水的三维结构提供了清楚的物理图像(Fu and Liu ,2001,2003,2007; Fu et al, 2003,2006; 傅云飞等，2012; Liu et al.,2013 )。  **其次，同一平台多种仪器的准时空同步观测的降水信息与云光谱信号，可以有效、准确建立降水信号和光谱信号的关系。**Inoue and Aonashi（2000）最早对东亚锋区降水云和非降水云的可见光/红外信号差异进行了研究。Liu和Fu(2008)等使用TRMM PR和VIRS融合观测资料，对热带及副热带地区降水云和非降水云特征进行了统计分析，首次揭示了降水云与非降水云在ISCCP9种云型中出现的频率分布，并指出了其云分类方案的命名欠缺合理性。Liu和Fu(2010)还研究了东亚锋面系统和台风系统中云降水与非降水云在可见光/红外通道信号上的差异，并在此基础上构建了多套降水云识别方案。Xu et al.（2009）借助1998-2007年的TRMM资料，研究了中国南部、台湾和南海地区梅雨季的降水分布特征和降水系统的中尺度结构。傅云飞等（2011）通过融合PR和VIRS十年探测资料，利用PR对降水云的直接识别，分析了夏季热带和副热带地区对流、层状降水云光谱信号的气候分布特征。Fu（2014）、傅云飞（2014）利用双光谱反射率算法对TRMM VIRS信号进行反演得到的云参数，并对中国东部降水云参数和降水的关系进行了分区研究，揭示了区域差异性特征。可见，TRMM PR和VIRS的综合探测能力优势明显且广泛应用并成果斐然。  随机森林（Random Forest）是2001年由Leo Breiman和Culter Adele开发的一种数据挖掘方法，是一种现代分类与回归的机器学习技术，同时也是一种组合式的自学习技术(Breiman et al.2001)[3]。其优越性体现在同等运算率下的高预测精度，以及相较于传统的统计方法，对非线性的数据关系有更好的拟合效果(Dietterich T G.et al.2000)[4]，并且能够进行变量重要性分析，对比神经网络和支持向量机等其他暗箱方法在分析变量关系上存在优势(Esteban J.et al.2019)。近年来，国内外基于卫星遥感和机器学习的气象资料研究正快速发展。例如，陆会（2019）进行了基于葵花8号(HIMAWARI-8)卫星遥感数据的大雾识别研究，主要通过两类算法对大雾识别进行研究，一类算法是基于传统的机器学习方法的大雾的识别，二类算法是基于深度学习的大雾识别。由于新一代静止卫星（风四和葵花八）在空间分辨率的大幅提高（2-4km），因此急需提升卫星光谱反演降水能力，利用随机森林的机器学习训练和建立卫星光谱和云参数以及降水的关系模型，进行降水的反演和预测，将对降水监测与预报方面的研究、防灾与减灾具有重要的科学意义和社会价值。   1. **创新点与项目特色**   对于静止、极轨等多光谱卫星在降水监测方面，面临的一个重要问题就是很难在时空同步上来获取光谱信号和降水信息（以往的研究多以地面降水观测资料来验证卫星光谱信号估算的降水），因而时空上均存在较大误差，不能准确发展卫星光谱估算降水的方法。因此本项目基于卫星遥感和机器学习的云参数和降水强度关系模型研究的创新主要为以下两点：  **1、卫星和探空反演的各类云参数数值和地面高密度自动站雨强观测值的时空匹配，解决了降水信息和卫星光谱信号的准时空同步性，这是研究方法和思路的创新；**  **2、基于机器学习，建立云参数与降水强度间的关系模型，结合高密度自动站雨量计，一方面将TRMM机器学习模型拓展到静止卫星上；另一方面，基于高密度雨量数据和静止卫星进行机器学习的再建模，这是研究内容上的创新，也是本项目的集成创新之处。**   1. **技术路线、拟解决的问题及预期成果**   **1、技术路线**  本项目基于卫星遥感和机器学习两种方案，同时展开研究，建立可靠的云参数和降水的关系，以提高对降水的预测。  该方案通过卫星遥感，以TRMM PR和VIRS融合观测资料为基础，研究云参数和降水强度的关系。整个方案的技术路线如下7个步骤：  **1）卫星资料整编**。整理出的1998—2013年期间的TRMM PR和VIRS的融合轨道资料。挑选出中国东部地区典型的梅雨锋面降水个例和中尺度热对流降水个例，与TRMM 轨道资料进行比对和匹配。  **2）轨道资料的格点化处理**。对10年夏季所有上述轨道资料进行格点化处理，得到0.5度水平空间分辨率气候态的降水及其相应光谱及云参数的格点数据产品，要素包括：地表降水率，可见光（0.63和1.6um）反射率，红外亮温（10.8um）以及云滴有效半径和光学厚度。  **3）基于轨道数据，进行典型个例分析**：主要分析锋面降水和中尺度对流系统的降水中的对流和层云降水的光谱信号、云参数特征及其差异。  **4）基于格点数据，进行空间分布特征分析**：分析层状云和对流云降水的地表降水率，可见光（0.63和1.6um）反射率，红外亮温（10.8um）以及云滴有效半径和光学厚度的空间分布特征及其差异，重点给出层状云和对流云的云参数和降水的相关性分析。  **5）基于轨道和格点数据，建立TRMM卫星光谱和降水信号关系的机器学习模型**：建立卫星光谱可见光（0.63和1.6um）反射率，红外亮温（10.8um）、云参数（云滴有效半径和云光学厚度）与地表降水的随机森林模型。给出重点给出层云降水和对流降水的上述经验统计关系的差异特征。  **6）TRMM卫星光谱和降水信号关系的机器学习模型在静止卫星估算降水的应用。**进一步将机器学习模型应用于静止卫星（风四和葵花八）光谱和云参数的降水反演和预测，探讨对流降水和层云降水的光谱信号和云参数的差异特征及其对降水的指示意义。  **7）基于高密度自动站和静止卫星光谱的机器学习模型研究。**该方案将风云四号（FY-4）和向日葵8号(HIMAWARI-8)卫星的降水资料，光谱资料同高密度自动站的地面降水资料进行时空匹配，通过机器学习，研究云参数和降水强度的关系。整个方案的技术路线如下3个步骤：  **a云参数和地面降水数据的时空匹配处理。**风云卫星反演得到的云参数产品，是空间分辨率为5km的格点数据。为了进行云参数与地面降水的时空匹配统计，在空间上，以观测点经纬度为中心，取其周边最近的9个格点上相应的云参数的算术平均值，作为该点对应的云参数值；在时间上，取同一时次的探空和卫星反演的云参数值，与其后一个时次的小时雨量观测值进行对比分析。  **b云参数和降水数据的统计分类。**根据降水强度（r，每小时降水量；单位：mm/h）大小，将降水分为：无降水（r = 0mm/h）、弱降水（r < 1mm/h）、一般降水（1mm/h ≤r <10mm/h）和强降水（r ≥10mm/h）四类。云顶高度、云顶温度、云光学厚度、云粒子有效半径等云参数按数值大小范围进行分档。为统计方便，规定某时次、某站点对应的数据集为一个统计样本，包括卫星和探空反演的各类云参数数值和地面雨强观测值。  **c搭建随机森林模型，预测降水强度。**从各类云参数数值和地面雨强观测值的统计样本中有放回的抽样，构造子数据集，子数据集的数据量是和统计样本相同的。利用子数据集构建子决策树，将数据放至每个子决策树中，将分好档的各类云参数作为决策树的特征，每个子决策树输出一个降水强度结果。最后，如果有了新的云参数数据需要通过随机森林得到分类结果，就可以通过对子决策树的对降水强度的判断结果投票，得到随机森林的输出结果，从而实现了对降水的预测。  TMRR PR和VIRS的融合资料  静止卫星  （风四和葵花八）  高密度自动站  地面降水资料  随机森林模型  降水的反演和预测  卫星光谱及其反演的云参数因子重要性排序  应用至  降水的反演和预测  搭建  对模型进行验证和改进  高密度自动站  地面降水资料  随机森林模型  搭建  降水的反演和预测  对比两种机器学习模型的结果，选取预测精度检验较高的一组  时空匹配  时空匹配  **2、拟解决的问题**  传统的静止轨道气象卫星检测降水的指标不确定性很大，本项目将对PR和VIRS的融合轨道资料进行机器学习回归模型分析，建立云参数预测降水的随机森林模型并应用到静止卫星上；同时，为降低统计回归预测降水的误差，我们将采用风云四号（FY-4）和葵花8号(HIMAWARI-8)卫星的降水资料，光谱资料同高密度自动站的地面降水资料进行时空匹配，通过机器学习建模，找到可靠的云顶信息与降水的关系，以提高降水预测的准确性。  **3、预期成果**  本项目将利用同步观测的揭示云参数和降水强度的关系，并利用卫星和探空反演的各类云参数数值和地面雨强观测值，构建基于云参数的反演和预测降水的机器学习模型，提高静止卫星光谱反演和预测降水的能力。研究成果预期可在国内外期刊上发表论文1篇。   1. **项目研究进度安排**   研究计划实施时间为2020年4月－2021年4月，研究周期12个月，季度研究计划如下：  **第一季度：**整编 TRMM PR和VIRS的融合轨道资料并进行格点化处理，分析典型个例和空间分布特征，研究对流降水和层云降水的云顶可见光、红外信号、对流降水和层云降水云的云参数的时空分布特征。  **第二季度：**通过机器学习（随机森林）进行降水光谱信号以及云参数和地表降水强度的关系建模，并给出的卫星光谱及其反演的云参数因子重要性排序。将机器学习模型应用于静止卫星（风四和葵花八）光谱和云参数的降水反演和预测，探讨对流降水和层云降水的光谱信号和云参数的差异特征。  **第三季度：**将地面高密度自动站的雨强观测值与机器学习模型对降水的反演和预测进行时空匹配，对已建立的机器学习模型进行验证和改进；基于静止卫星光谱和高密度的雨量计的机器学习建模。总结撰写相关论文。  **第四季度**：进行课题成果梳理和总结，申请软件著作权等，完成论文发表。   1. **已有基础**    1. **与本项目有关的研究积累和已取得的成绩**   项目组成人员结构合理，专业方向涉及大气探测、大气物理、卫星遥感等多方面学科交叉，在云和降水观测、云微物理分析方面有专业背景知识，为本项目的顺利开展提供了有利的技术累积和人力保证。  指导老师杨元建，目前主要从事大气遥感、城市化-边界层-气溶胶相互作用及其气候环境效应研究。已在PNAS, Earth-Sci. Rev.、GRL、ACP、JGR等国内外期刊发表或接收学术论文100余篇 (含SCI 50余篇，EI 3篇)，其中第一作者或通讯作者论文50余篇（含SCI 20余篇）。曾获谢义炳青年气象科技奖(2016),中国气象局青年英才(2016), Science of the Total Environment杰出审稿人(2018), Advances in Climate Change Research期刊最佳论文贡献奖(2019),安徽省直机关五一劳动奖章(2017), 安徽省重大气象服务先进个人(2016)等奖励或荣誉称号，为本项目的顺利开展、管理与实施提供了丰富经验。在卫星遥感与降水方面，在博士期间，建立了一套基于TRMM的降水云参数和降水的数据集，在区分降水云和非降水云前提下，从气溶胶、降水类型和地形角度揭示了夏季降水云宏观和微观特性特征及成因，提供了气溶胶-降水云参数相互作用的卫星观测研究新视角。发展了基于卫星遥感的云参数和单柱辐射传输模式相结合的大范围网格化的云辐射计算方案，定量解决了高云（卷云和云砧）辐射强迫不确定性问题。随机森林机器学习方法已成熟熟练作为主体方法在夜间灯光人口空间反演与空气污染暴露，以及地面太阳辐射估算等研究中，这些研究成果都为本项目的开展提供了坚实的工作基础。此外，通过参与研发专项，已获得近5年高密度自动观测站的地面降水小时资料充足，风云四号卫星和葵花8号可提供准确度较高的降水和光谱资料。   * 1. **已具备的条件，尚缺少的条件及解决方法**   目前关于项目的软硬件计算条件都已具备，还缺少数据的存储介质，拟通过本项目预算解决。 |

* 1. 经费预算

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 开支科目 | 预算经费  （元） | 主要用途 | 阶段下达经费计划（元） | |
| 前半阶段 | 后半阶段 |
| 预算经费总额 | **3000** |  | **1600** | **1400** |
| 1. 业务费 |  |  |  |  |
| （1）计算、分析、测试费 |  |  |  |  |
| （2）能源动力费 |  |  |  |  |
| （3）会议、差旅费 | **1000** | 高密度雨量计站址调研和学术交流差旅费 | **500** | **500** |
| （4）文献检索费 | **200** | 检索论文、期刊文献 | **100** | **100** |
| （5）论文出版费 |  |  |  |  |
| 2. 仪器设备购置费 |  |  |  |  |
| 3. 实验装置试制费 |  |  |  |  |
| 4. 材料费 | **1800** | 购买资料存储介质（3T移动硬盘），其他电脑打印等设备维修（包括电脑耗材、墨盒、硒鼓、网卡等） | **1000** | **800** |
| 学校批准经费 |  |  |  |  |

* 1. 指导教师意见

|  |
| --- |
| 随着卫星遥感技术不断发展，可见光/红外光谱扫描仪已成为目前最为广泛的星载探测仪器，尤其是新一代静止卫星（风云四号和葵花八号）上的可见光/红外仪器可以提供大范围更高时空分辨率的探测资料。利用卫星观测的可见光/红外信号进行估算地表降水率，一直都是卫星遥感降水研究中的重要课题，对降水监测与预报研究、防灾与减灾具有重要的科学意义和社会价值。  本项目申请书提出的研究方案和技术路线合理可行，作为申请人的导师对申请人及其研究小组有很好的了解，学风严谨，成绩优良，思维敏捷，对待问题有很强的洞察力，同时也有很好的工作基础和工作条件。本人也就全心全力实施跟踪指导，相信他们能在预定时间内完成任务，达成申请书所提出的研究目标。  特此推荐。    **导师（签章）：**  **年 月 日** |

* 1. 院系大学生创新创业训练计划专家组意见

|  |
| --- |
| **专家组组长（签章）：**  **年 月 日** |

* 1. 学校大学生创新创业训练计划专家组意见

|  |
| --- |
| **负责人（签章）：**  **年 月 日** |

* 1. 大学生创新创业训练计划领导小组审批意见

|  |
| --- |
| **负责人（签章）：**  **年 月 日** |