

本科生毕业设计

|  |
| --- |
| 带电粒子人工降雨外场实验  的效果评估 |

|  |  |
| --- | --- |
| 院 系 | 计算机科学与技术 |
| 专业班级 | 计科1703 |
| 姓 名 | 马恒奔 |
| 学 号 | U201714601 |
| 指导教师 | 秦磊华 |

2021年05月30日

**学位论文原创性声明**

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包括任何其他个人或集体已经发表或撰写的成果作品。本人完全意识到本声明的法律后果由本人承担。

作者签名： 年 月 日

**学位论文版权使用授权书**

本学位论文作者完全了解学校有关保障、使用学位论文的规定，同意学校保留并向有关学位论文管理部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权省级优秀学士论文评选机构将本学位论文的全部或部分内容编入有关数据进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于 1、保密囗，在 年解密后适用本授权书

2、不保密囗 。

（请在以上相应方框内打“√”）

作者签名： 年 月 日

导师签名： 年 月 日

摘 要

“天水计划”是以带电粒子催化降雨为核心原理的人工降雨项目。为了对带电粒子催化降雨的效果进行评估，需要对实验期间的天气数据进行分析，近似求出实验期间目标区的假设未进行作业时的降雨量，以及采用多种评估手段对多方面的要素进行分析，从而使评估结果更加可靠。**评估意义**

传统的自然降雨的预测一般采用线性回归的方法，但是由于降雨的原理十分复杂，没有明显的线性关系，使用线性回归预测的结果一般不够理想。使用机器学习的方法可以更好地解决非线性的问题，通过搭建神经网络对自然降雨进行预测，使预测结果更加准确。对收集到的气象站的数据进行分析，做到多维度、多方面的分析检验。对“天水计划”评估设计中的随机试验进行分析，以及对采集到的离子数、全天空测量仪的图片数据进行分析，探究这些因素与降雨之间的关系。**传统的降雨预测多种方法，列举出来，把不理想的地方说出来**

**关键词**：人工增雨；机器学习；数据分析；统计检验

Abstract

Failure resilience is of pivotal importance in practical network function virtualization (NFV) systems, but has been mostly absent in the existing ones. The absence is mainly due to the challenge of patching source code of the existing NF software for extracting important NF states, a necessary step toward flow migration and replication to provide failure tolerance.

This paper proposes NFActor, a novel NFV system that uses the actor programming model to provide transparent resilience, high scalability and low overhead in network flow processing. In NFActor, a set of efficient APIs are provided for constructing NFs, with inherent support for scalability and resilience. A per-flow management principle is advocated, different from the existing practice, which provides dedicated micro service chain services for individual flows, enabling decentralized flow migration and scalable flow replication. We implement NFActor and show that it achieves good scalability, prompt flow migration and failure recovery with large numbers of concurrent flows. We also show that NFActor can enable applications such as live NF update and correct MPTCP subflow processing, which cannot be efficiently achieved in previous systems.

**Keywords:** Network Function Virtualization (NFV), Failure resilience, Actor model, Flow migration, Service Chain

撰写说明，阅后删除！！！

英文摘要在中文摘要定稿后再开始撰写，否则会多次修改，不要过于依赖百度翻译等工具。

Abstract是论文的英文摘要，一般对照中文摘要翻译，要求另起一页。Keywords对照中文关键词翻译，与英文摘要在同一页，以“Keywords：”另行顶左开始。Keywords之间用“,”隔开，最后一个Keyword后不用加任何标点符号。

目 录

[摘 要 I](#_Toc23945441)

[Abstract III](#_Toc23945442)

[1 绪 论 1](#_Toc23945443)

[1.1 课题背景 1](#_Toc23945444)

[1.2 国内外研究现状 2](#_Toc23945445)

[1.3 研究目的和主要内容 3](#_Toc23945446)

[1.4 论文结构 4](#_Toc23945447)

[1.5 课题来源 4](#_Toc23945448)

[2 方案论证（或具体背景技术概述） 5](#_Toc23945449)

[2.1 系统需求分析 5](#_Toc23945450)

[2.2 系统可行性分析 6](#_Toc23945451)

[2.3 开发工具分析及选择 6](#_Toc23945452)

[2.4 关键技术分析 6](#_Toc23945453)

[2.5 基本方案制定 7](#_Toc23945454)

[2.6 本章小结 7](#_Toc23945455)

[3 XXX系统设计 8](#_Toc23945456)

[3.1 功能需求 8](#_Toc23945457)

[3.2 系统总体设计 9](#_Toc23945458)

[3.3 功能模块设计 13](#_Toc23945459)

[3.4 本章小结 16](#_Toc23945460)

[4 XXX系统实现 17](#_Toc23945461)

[4.1 过滤器实现 17](#_Toc23945462)

[4.2 属性管理模块实现 18](#_Toc23945463)

[4.3 数据迁移模块实现 19](#_Toc23945464)

[4.4 本章小结 22](#_Toc23945465)

[5 性能测试与分析 23](#_Toc23945466)

[5.1 测试环境 23](#_Toc23945467)

[5.2 功能测试 23](#_Toc23945468)

[5.3 系统界面 23](#_Toc23945469)

[5.4 性能测试 23](#_Toc23945470)

[5.5 本章小结 25](#_Toc23945471)

[6 总结与展望 26](#_Toc23945472)

[致 谢 27](#_Toc23945473)

[7 毕业设计模板基本框架 28](#_Toc23945474)

[7.1 封面 28](#_Toc23945475)

[7.2 原创性声明页 29](#_Toc23945476)

[7.3 摘要 29](#_Toc23945477)

[7.4 目录 29](#_Toc23945478)

[7.5 参考文献 30](#_Toc23945479)

[7.6 附录 30](#_Toc23945480)

[7.7 毕业设计任务书 30](#_Toc23945481)

[7.8 成绩评定页 31](#_Toc23945482)

[8 毕业设计撰写要求 32](#_Toc23945483)

[8.1 图的格式 32](#_Toc23945484)

[8.2 表的格式要求 36](#_Toc23945485)

[8.3 公式 38](#_Toc23945486)

[8.4 流程图 39](#_Toc23945487)

[8.5 常见格式问题 39](#_Toc23945488)

[参考文献 41](#_Toc23945489)

[附录：大学期间发表或提交的论文 44](#_Toc23945490)

撰写说明，阅后删除！！！

目录不会自动更新，排版变动后必须更新目录，更新目录方式：鼠标点击目录，按F9键或者右键更新域，选择更新整个目录即可。

目录仅包含一级标题和二级标题和标题样式，目录是全论文的纲要。中文摘要、Abstract、论文正文的各级标题（不包括第三级）、致谢、参考文献、附录等都应编入目录，标注其页码对照关系，但目录本身不出现在其中。中文摘要、Abstract、目录等使用希腊数字“I、II、…”编连续页码；论文正文、致谢、参考文献、附录等使用“1，2，3，… ”编连续页码。

# 绪 论

本章介绍了“天水计划”项目的由来、主要内容和意义，然后指出了外场试验效果评估的重要性，也就是本文工作的意义。接着说明了国内外人工降雨效果评估方面的常见做法，提出了引入机器学习对传统评估进行改进的想法。

## 课题背景

### 人工降雨与“天水计划”

水资源短缺是一个全球性的难题。与世界主要发达国家相比，我国的水资源形势尤为严峻，人均水资源量仅为世界平均水平的四分之一[1]。气象统计表明，只有16%-18%的水汽能形成降水降落地面，大气水资源利用率严重不足。若能将大气水资源的利用率提高若干个百分点，也必将显著提升我国有效水资源总量，经济效益潜力巨大。

基于干冰、碘化银等催化剂的传统人工增雨技术，是目前大气水资源开发的一种主要技术手段。但是近年来，随着水资源供需矛盾日益突出，传统人工增雨技术的一些固有缺点逐渐凸显，如作业效率低，温湿度窗口过小，火箭、高炮、飞机等播撒方式管控严格等。这些缺点大大限制了大气水资源开发的成效和规模。

所以，本世纪以来，西方越来越多的国家开始进行新的人工降雨技术的研究，其中最具有代表性的技术之一就是带电粒子催化降雨。该技术至今已在墨西哥、阿联酋、阿曼等多个国家的干旱地区进行了成功的外场试验。[2-4]鉴于此技术对于充分缓解水资源短缺问题意义重大，国外相关机构都对此严加保密。

在此背景下，我们国家在2016年推出了“天水计划”。该项目以基于带电粒子催化降雨雪的大气水资源开发新技术为对象，围绕大气水资源开发与调控技术体系的构建，从理论、关键技术及装备到应用示范开展全链条研究。该项目重点开展带电粒子催化降雨雪科学规律的研究、突破关键技术与装备研发、人工降雨雪过程模拟与空陆一体化区域水资源适应性调控，以及带电粒子催化降雨雪系统集成与应用示范研究。“天水计划”的基本原理是在大气中播撒带电粒子，使空气中部分气溶胶带电。这些带电气溶胶粒子的静电场对其他中性水分子簇团存在极化效应，产生带电气溶胶粒子对被极化的水分子簇团的非接触的电场凝聚力，促使其凝结速率增加，促进降雨的形成[5]。

“天水计划”的外场试验区选择了六盘山地区和乌鞘岭地区，在六盘山地区建立了三个单电极带电粒子发生器装置和一个双电极带电粒子发生器装置，在乌鞘岭地区建立了三个单电极带电粒子发生器装置。

“天水计划”意在自行研发带电粒子催化降雨雪的核心技术，通过对外场试验中带电粒子发生装置的实际效果的分析，对带电粒子催化降雨雪试验的方案进行评估，进而更加充分利用大气水资源以缓解国内水资源匮乏的现状。

### 降雨效果评估

“天水计划”从2016年开始实施，先后完成了带电粒子催化降雨雪的科学规律和关键技术的研究，完成了应用示范具体选址，完成了气象监测网络的设计以及基站的集成控制。该项目在2020年确定了外场试验方案，并于2020年7月到2020年11月，在乌鞘岭和六盘山两个试验区成功开展了外场试验。

为了对外场试验的结果进行分析，以及对带电粒子人工降雨方案的效果进行评估，从而作为进一步开展实验的重要参考，需要进行带电粒子人工降雨外场实验效果评估。人工降雨效果检验是人工降雨试验中重要的一部分，它的基本要求为科学、客观和定量。[6]

为了进行科学有效的效果评估，一个重要工作是估计出实验区假设未播种情况下的自然降雨量，从而区分出实际降雨中的自然降雨部分和催化增加的部分。一般情况下，自然降雨量的求取采用线性回归的方法，通过线性回归的方法建立目标区域和对比区域的历史回归方程。但是预测精度不高，也不能充分利用采集到的气象数据。同时，“天水计划”所采取带电粒子催化降雨雪技术在国内并无先例可循，装置的运行特性、影响范围、作用时间都尚未可知，需要对多种气象要素进行分析。

而随着计算机技术的不断发展，机器学习在计算机视觉[7]、自然语言处理[8]、数据挖掘[9]、数据分析[10]等多个领域取得了很好的效果。机器学习是从大量数据中通过选取合适的算法，自动的归纳逻辑或规则，并根据这个归纳的结果(模型)与新数据来进行预测。相比于传统的基于统计学的算法，利用机器学习可以更好地解决非线性的降雨问题[11]。将机器学习地方法和传统统计方法相结合，可以进一步提高效果检验地精确度和可信度。本文使用了神经网络对自然降雨进行了预测，取得了不错的效果。

神经网络的预测模型也有自身的限制，它本身需要大量的数据进行训练来提高模型的精确度。但是“天水计划”外场试验周期较短，只有部分历史数据丰富的站点可以使用神经网络模型来预测自然降雨。“天水计划”项目的自建站点建成时间较晚，采集数据较少，这些数据无法构建神经网络模型，需要结合传统的统计检验方法和显著性检验方法。为此设计了多种分析手段，来从多方面的对多种气象数据进行分析评估。

### 面临的问题和挑战

降雨评估手段没有规范化的定量方法。目前还没有令人信服的科学依据表明各种形式的播云试验是有效或可行的增加降雨量的手段。而且，带电粒子人工降雨在国内也是首例，对于降雨效果的评估设计是困难的而没有依据可循，需要设计科学有效的评估方法。也要进行多方面的评估，从多角度来对结果进行评估。

数据清理、整合和分析。效果评估的数据多来源于自动气象站，该类数据的数据量较大，采样密度大，缺测数据也较多，且该类数据的观测项目较多，包含了很多干扰项目。不同地区，不同气象站的数据结构不同，没有统一标准。需要对这些数据进行辨别，甄别出有用数据与无用数据，以及对数据要进行清理和整合，最后这些数据才能进行分析，也要设计多种分析方法，来找到降水的变化以及其他气象要素和降雨之间的联系。

## 国内外研究现状

人工降水是目前大气水资源开发的主要技术手段。世界气象组织在近期发布的报告中指出，应把人工降水作为水资源综合管理战略的一部分，并建议在各国建立开展云、雾和降水气候学分析，开展跨国外场试验和独立专家评估等，为人工影响天气、云水资源研究和业务提供依据。

人工降雨效果检验主要采用物理效果检验和统计效果检验。其中物理效果检验是对降雨技术有效性的定性检验，而统计效果检验是在一定显著性水平下的定量的效果分析。统计效果检验从最初的简单对比，逐步发展出了随机试验的各种统计效果检验方法，科学性和可靠性都有了明显的提升。

以色列在1961-1967年进行的冬季人工降水试验中，使用了随机交叉试验的方案，分别利用秩和检验法和双样本检验法对增雨效果进行了评估，实验结果发现增雨效果可以达到21.6%[12]。在美国怀俄明州开展的WWMPP人工增雨试验中，也采用了区域回归随机试验的方法，并提出了更加精确的效果评估指标[13]。我国于1975至1978年，在福建古田水库开展的小火箭人工降雨试验中，也采用了区域回归随机试验的方法，通过分别拟合试验区和对比区之间的回归关系直线，采用了多个事件检验法与双样本分析，并根据催化云的类型以及天气的不同，进一步进行了分层检验，最终得出在显著性水平0.05的情况下，对锋面天气以及小雨天气的积状云催化效果最佳的结论[14]。北京地区在2001到2004年开展的火箭、高炮催化作业中，分别采用序列试验法、回归分析法、双比分析法与区域回归数值模拟对效果进行了评估，分别在一定显著性水平上给出了增雨效果[15]。随着自动气象站普及和互联网技术的进步，我国在2011年实现了基于.NET的华南区域人工增雨效果统计检验共享平台，该平台采用区域回归的方法进行分析，效果检验的精度得到了大幅度提高[16]。

在目前超过56个国家开展的人工影响天气项目中，不乏有取得可观成果的[17]，如CLIMAX I&II[18]，ISRAEL I &II[19]等。然而，绝大多数项目没有经过科学的定量评价与催化试验效果检验，从而不能在较高置信度水平上给出科学的人工影响天气试验的定量效果评估。此外，2003年美国国家研究委员会(NRC)[20]与2007年世界气象组织(WMO)的报告[21]中都指出，目前还没有令人信服的科学依据表明各种形式的播云试验是有效或可行的增加降雨量的手段。即便是目前有记录的结果积极的试验，在经过审查和质疑后，其科学性与准确性依然有待商榷。中国气象局编制的《全国人工影响天气发展规划（2014-2020年）》也指出，我国目前的人工影响天气工作存在“科技支撑不足，自主创新能力有待加强”等突出问题，“系统性的人工影响天气研究与关键技术研发滞后，作业效率明显低于先进国家”。

我国在开展试验的早期，由于缺乏必要的观测设备和观测手段，以及试验设计缺乏周密性和合理性，尽管开展了数次人工降雨作业，但几乎不能提供任何科学上有力的证据证明降雨作业的作用效果[22]。随着试验设计越来越科学完善，观测设备与观测手段进一步增强，我国对人工降雨试验的评估也逐渐走上正轨。就目前而言，国内外对人工降雨评估的手段和方法都比较单一，都是基于传统的统计方法对降雨效果进行评估[14]。

对于降雨预测问题，使用机器学习与统计手段相结合进行效果评估的方法，在学术界几乎是空白的。只有王伟健等人利用随机森林算法，对单次降雨作业进行了探索性的研究[23]。相比于传统的基于统计学的分析方法，使用机器学习可以更好地解决非线性的降雨问题[11]，提高降雨量的预测精度。使用机器学习还可以使用气象站测得的多种气象数据进行辅助分析预测，从而进一步的提高精度，这在传统效果检验中做不到的。通过引入温度、压强、湿度、能见度等多种气象要素，将降雨这一过程中的物理变化也纳入了考虑的范畴，这与传统的线性回归模型相比更加科学，结果也更加精确。

## 论文结构

本文的主要内容如下：

第一章首先介绍了天水计划的背景与研究方向，然后介绍了人工增雨的效果评估方法，对项目过程中的问题和挑战进行了分析，对国内外的研究情况进行了介绍，并对研究目的和主要内容进行了说明。

# 方案论证

本章对开展效果评估工作进行了整体分析和设计，介绍了评估工作使用到的技术手段和分析方法，指定了评估工作的整体方案。

## 需求分析

“天水计划”于2020年7月至11月，在乌鞘岭和六盘山地区分别开展了带电粒子催化人工降雨雪外场试验。该项目在两个试验区自建雨量站点和综合站点在试验期间收集到了大量的气象数据，此外试验区的自动气象站记录了包括试验期间在内的长期历史数据。

为了对外场试验结果进行评估，用来作为带电粒子发生装置外场有效性的外场试验证据，现需要对已经收集到的数据进行分析评估，针对实验特性和数据特点设计一套科学、有效、定量的外场试验评估方法。

## 可行性分析

“天水计划”于2019年分别在六盘山和乌鞘岭试验区建立了雨量站和综合站，这些站点记录了每小时的雨量、风向、风速、小时雨量、气温、相对湿度、空气中离子数浓度以及装置运行电压等数据。同时，六盘山和乌鞘岭国家气象站提供了十余年的多种气象数据作为分析的数据支持。通过数据处理和数据分析，为外场试验结果设计一套评估方法是可行的。

人工催化增雨试验的实验结果中最重要的部分就是该地区的降雨量，实验期间该地降雨量分为自然降雨量和催化增雨量，我们的目的在于求出催化增雨在降雨中所占的比例，所以重点在于得到自然降雨量。由于气象站有长达十余年的气象数据，比较充足。且考虑到自然降雨预测是一个非线性的问题，直接使用线性回归预测的方法效果会不够理想。基于上述分析，选择引入机器学习的方法来对自然降雨量进行预测。通过构建神经网络模型对六盘山和乌鞘岭试验区的试验期间自然降雨量进行预测，来实现精度较高的预测模型。这是对外场试验开展效果的宏观检验，通过将模型的预测结果与实际结果进行对比，清晰地说明在试验区降雨量增加与否，增加量多少。

使用可靠性高的神经网络模型对目标研究量进行预测是一个直观、可信度高的方法。但是，神经网络的训练需要时间维度较长的数据，对于“天水计划”自行建立的雨量点收集到的短期数据，以及自建的综合站点在实验期间的收集到的云量、离子数、风向风速等其他气象要素，由于数据的时间范围相对较小，导致实验数据比较少。使用机器学习的效果比较差，所以需要采用除机器学习之外的其他方法进行分析。对于这些数据的处理，需要灵活变通，根据现有的数据设计最科学的评估方法，来检验带电粒子人工降雨实验的效果。这些数据相对比较微观，讨论的是一个实验区内部的降雨变化。

## 开发工具分析及选择

开展效果评估首先要进行的就是数据处理，由于收集到的数据内容相对较多，且各种数据格式不同，数据均是来自自动测量仪记录的原始格式，所以需要进行大量的数据处理和数据清洗工作。另外，需要针对外场试验设计神经网络模型预测自然降雨量，所以选择广泛应用于数据处理和机器学习的python语言作为开发语言。

对于大量的数据处理和清洗工作，使用基于NumPy（提供高性能的矩阵运算）的Pandas分析工具对数据进行处理。Pandas是python用于数据处理和分析的软件库，它提供数据表格的操作方法和时间序列的数据结构。它是一个强大的分析结构化数据的工具集，用于数据挖掘和数据分析，同时也提供数据清洗功能。

搭建神经网络时，选择使用TensorFlow框架进行机器学习算法的实现。TensorFlow是一个开源软件库，是基于数据流编程的符号数字系统，被广泛应用于机器学习领域。TensorFlow功能强大且非常灵活，而且对于之前未接触过机器学习的初学者比较友好。

## 关键技术分析

### 全连接神经网络

全连接神经网络是一种结构简单的神经网络，包括输入层、输出层和隐藏层。其中输入层和输出层均只有一层，输入层的神经元个数是输入的特征数量决定的，输出层的神经元数量是网络预测的标签数量决定的。层与层之间神经元相互连接，层内神经元互不连接，而且下一层神经元连接上一层所有的神经元，这就是全连接的由来。一个仅有一个隐含层的神经网络就能拟合任何一个函数，所以他的非线性拟合能力很强，它的缺点是网络中需要较多的神经元。每个神经元的模型如下图所示。



神经元模型

激活函数f是一个非线性函数，它将线性的输出转换为非线性，来解决非线性的问题，如果没有激活函数，神经网络就只能解决线性问题。

神经元之间的连接是神经元中最重要的东西，每个连接都有一个权重w，对神经网络的训练就是让权重的值调整到一个最佳状态，使得预测结果最好。调整权重是通过反向传播(Back propagation，简称BP)算法实现的，通常与一种最优化方法（如梯度下降法）结合使用。它的基本原理是，将神经网络模型训练的过程中产生的模型输出与期望输出之间的损失loss，向前传播反馈到隐含层和输入层，各层根据反馈信号，以及输入输出权重和该神经元的阈值，使用梯度下降法计算出参数调整的方法，调整后进行新一轮的训练。一次来不断调整权重，改善模型性能。

### 长短记忆神经网络

在说明长短记忆神经网络之前首先对循环神经网络(RNN, Recurrent neural network)进行说明，RNN是专为处理序列数据的一种神经网络，是一种处理序列数据的网络结构，所有循环单元按照链式连接。RNN的网络结构如图所示。



图

与全连接神经网络相似，RNN也分为输入层(Xt)、输出层(Ht)和隐藏层(A)。但是不同的是，它的隐藏层A含有循环结构，循环结构有参数（主要为权值参数）共享的重要特性。由于序列数据前后之间有较强的关联性，所以RNN被设计为在处理当前信息时，会考虑前序信息，所以对于序列数据的预测效果很好。但是RNN也有一些缺点，比如它无法解决“长依赖问题”，为此，许多人对RNN进行了改进，一种广泛应用的网络就是长短记忆神经网络(LSTM，Long Short-Term Memory)。

循环神经网络都具有神经网络的重复模块链的形式。标准RNN中，该模块只有一个tanh层，LSTM对于此结构模块进行了改进，引入了四层结构。通过引入门(Gate)结构，LSTM能够增添信息到神经元中，门结构示意图如图所示。



图 门结构

模块代表Sigmoid神经网络层，该网络输出0到1之间的数字，代表了有多少信息可以通过。X代表点乘法运算结构，对通过Sigmoid的信息和其他信息进行点乘运算。在LSTM中有三个这样的门结构，用来保护和控制神经元的状态。

## 基本方案的制定

为了使降雨评估更加完善和全面，既要完成宏观层面的试验区的降雨预测，也要把握一个试验区内部的微观的降雨差异。所以基于得到的数据的实际情况，通过分析各站点数据的特征要素以及各种数据的时间维度，将主要的工作划分为两个部分，主要内容为：

(1) 建立基于神经网络的降雨量预测模型

对六盘山和乌鞘岭两个试验区分别求出实验期间的自然降雨量的预测值，利用气象学界普遍认为与降雨关联性强的气象要素，根据获取数据的情况及其特征数量选择合适的神经网络或某种特定的机器学习算法建立降雨量预测模型，通过对模型的训练和评估，获得相对准确的目标区域降雨量预测模型。将得到的模型用于预测实验期间的自然降水，通过比较自然降水和实际降水来评估人工增雨效果。

(2) 建立基于统计分析的人工增雨效果评估模型

对于试验区内部的，由于地理、时间、风向等因素带来的降雨差异，分别进行分析。由于这些数据来源于“天水计划”建立的测量装置，工作时间较短，所以历史数据不足，无法使用神经网络分析。就要建立基于统计分析的效果评估模型，对各天气要素分别采用相应的分析方法，力求从微观上说明降雨的差异，完成对人工降雨的效果检验。

## 本章小结

本章从需求分析入手，分析项目收集到的实验数据，对实现分析的编程语言、框架等进行了选择和确定。针对不同类型和时间维度的数据指定了不同的分析方法对实验结果进行评估，确定了基于统计分析和机器学习相结合的总体评估思路。

# 基于全连接神经网络的自然降雨预测模型

本章分析处理了六盘山气象站十余年的气象数据，搭建了全连接神经网络进行自然降雨量的预测，通过评估验证了模型的可靠性，并对实际问题进行了应用。

## 六盘山自然降雨预测模型说明

六盘山试验区内的六盘山自动气象站具有丰富的实验数据和历史数据，有利于使用全连接神经网络，进行实验期间降雨量的预测估计。全连接神经网络模型结构比较简单，但是非线性拟合能力比较强，用来进行降雨量预测是一个不错的选择。而且，气象站采集到的天气要素种类以及数据规模，相对神经网络主要应用中使用到的数据集来说比较小，所以可以忽略全连接模型要使用较多神经元的缺点。

对六盘山气象站的降雨数据进行分析，来设计全连接神经网络模型。使用历史数据作为神经网络模型的输入，预测实验期间也就是2020年7月到11月之间，如果不加人工干扰的情况下的自然降水量，通过与实验期间的实际降雨量进行对比，分析带电粒子装置开启后带来的实际效果。整体上看，我们的模型是一个回归模型，目的在于预测出实验期间的自然降雨量。

## 数据处理与特征选取

在建立神经网络预测模型之前，首先要完成历史数据的筛选和清洗，以及进行特征的选取。由于气象站数据为自动站自动记录的，所以存在缺测、异常等多种干扰，需要对拿到的原始数据进行分析和异常检测，对数据清洗并提取特征。

### 数据整合

六盘山气象站拥有的气象数据时间维度为2008年1月1日至2020年11月30日，它的采集频率为一小时一统计，采集要素包括气压、温度、最高气温、最低气温、露点温度、相对湿度、最小相对湿度、最小相对湿度出现的时间、水汽压、过去1小时降雨量、两分钟平均风向、两分钟平均风速、最大风速的风向等多种气象数据。

由于降雨过程一般情况下是持续数个小时甚至几天的，所以每小时的降雨数据对我们来说意义不大。同时，对每小时的降雨预测难度非常大，所以需要对数据进行整合。按照一天为单位，对数据进行重采样，降低数据的时间分辨率。整合中要对不同的数据类型采用不同的方式。对于气温、气压、湿度、可见度等连续型标量，为了保持其物理意义，取它们24小时统计量的均值作为日均值。对于降雨量这种离散型标量，由于每小时统计的是过去1小时收集到的数值，所以将这些数据进行相加处理。对于风向等矢量数据，直接对角度求代数平均或者进行矢量运算都是不符合物理意义的，所以对于每天的24个风向信息，要找出其中的主导风向。

### 数据标准化处理

由于各个数据的量纲和数值范围不同，如果将原始特征直接放入网络中训练，数值较小的特征对网络的影响也会较小，实际上各个特征对降雨的影响程度是不清楚的。而且，神经网络中的激活函数是有非线性范围的，数值较大的特征进行训练会影响网络的初始化，甚至会影响网络的收敛性。所以需要对数据进行标准化处理。

其中*Dn*是标准化处理后的数据矩阵，*D*是原始数据矩阵，*Hmean*是训练集数据矩阵各特征的均值，*Hstd*是训练集数据矩阵各特征的标准差。对于神经网络划分的训练集、验证集和测试集，进行数据标准化处理时，采用的均是训练集数据的均值和标准差，因为我们假设认为训练集的均值和标准差代表了这项气象特征在该地的分布规律。

### 特征选取

选取特征时，要选择与降雨关系密切的气象要素，如气温、湿度等。选取特征的数量不宜过多也不宜过少，特征过少的话，神经网络难以从已有的特征中获取降雨规律的信息；特征选择过量的话，会人为地给神经网络引入噪音，使非相关因素影响了网络的效果。

## 预测模型的建立

### 模型数据说明

为了得到一个有效的模型，可靠地预测出2020年7月到11月的降雨数据。我们需要对设计的神经网络模型可靠性进行检验。因此，我们先不预测2020年实验期间的降雨量，而是使用历史数据来对未开展试验的一个年份进行自然降雨预测，通过比较预测数据与实际降雨数据之间的差异，来评价该模型是否有效，再使用该模型预测2020年7月到11月的自然降雨量。

因此我们选择2008年到2016年期间的数据作为训练集， 2017年到2018年的数据作为验证集，2019年的数据作为测试集。选择全连接神经网络作为网络结构，选取的特征为气压、温度、每小时降雨量、风速、风向。

### 模型的评估

因为2019年没有进行人工降雨，所以理论上来说，好的模型的预测结果应该与实际降雨量基本吻合，为了定量地说明模型的预测结果的可靠程度，引入决定系数R2对预测结果进行评估，它的定义为：

决定系数R2是用来度量因变量的变化中，可由自变量解释所占的比例，由公式可以看出，预测值和实际值的差异越小，R2的值越接近1。按照这一思想对模型预测结果进行检验。

### 模型初步结果与改进

神经网络训练过程中可能会出现过拟合现象，也就是模型在训练集上的预测准确度很高，但是在验证集上的准确度很低，这是由于模型过度迎合了验证集的数据特征。为了避免过拟合现象的出现，对网络采用了提前终止策略，通过监控验证集的均方误差在预定训练轮数中没有下降则提前终止训练。

搭建一个全连接神经网络，根据划分好的数据集和选取的特征数据进行训练。，图给出了神经网络迭代过程的损失曲线。

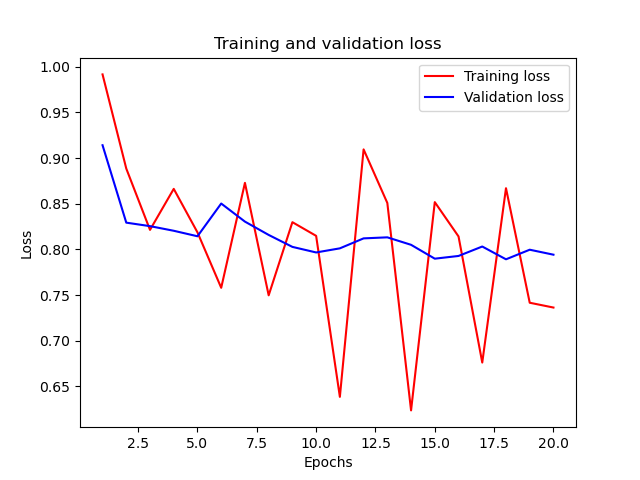


图5 预测2019年过程中的loss曲线

可以看到训练集和验证集的loss曲线没有随着训练轮数的增加而呈现不断下降趋势，训练集的loss值在上下振荡，而验证集的loss值虽有下降，但是下降幅度较小，一直维持在较高水平，这两者说明模型不仅在验证集中表现不好，甚至在训练集中都没有达到较好的效果，我们的模型准确度不够。我们使用R2对测试结果进行定量分析，将2019年的预测降雨量和实际降雨量代入公式进行计算，得到计算结果为0.21。这说明我们的预测量与实际量相差很大，相关性大致为21%，当前模型效果很差。我们分析预测结果来探寻原因，如图所示。

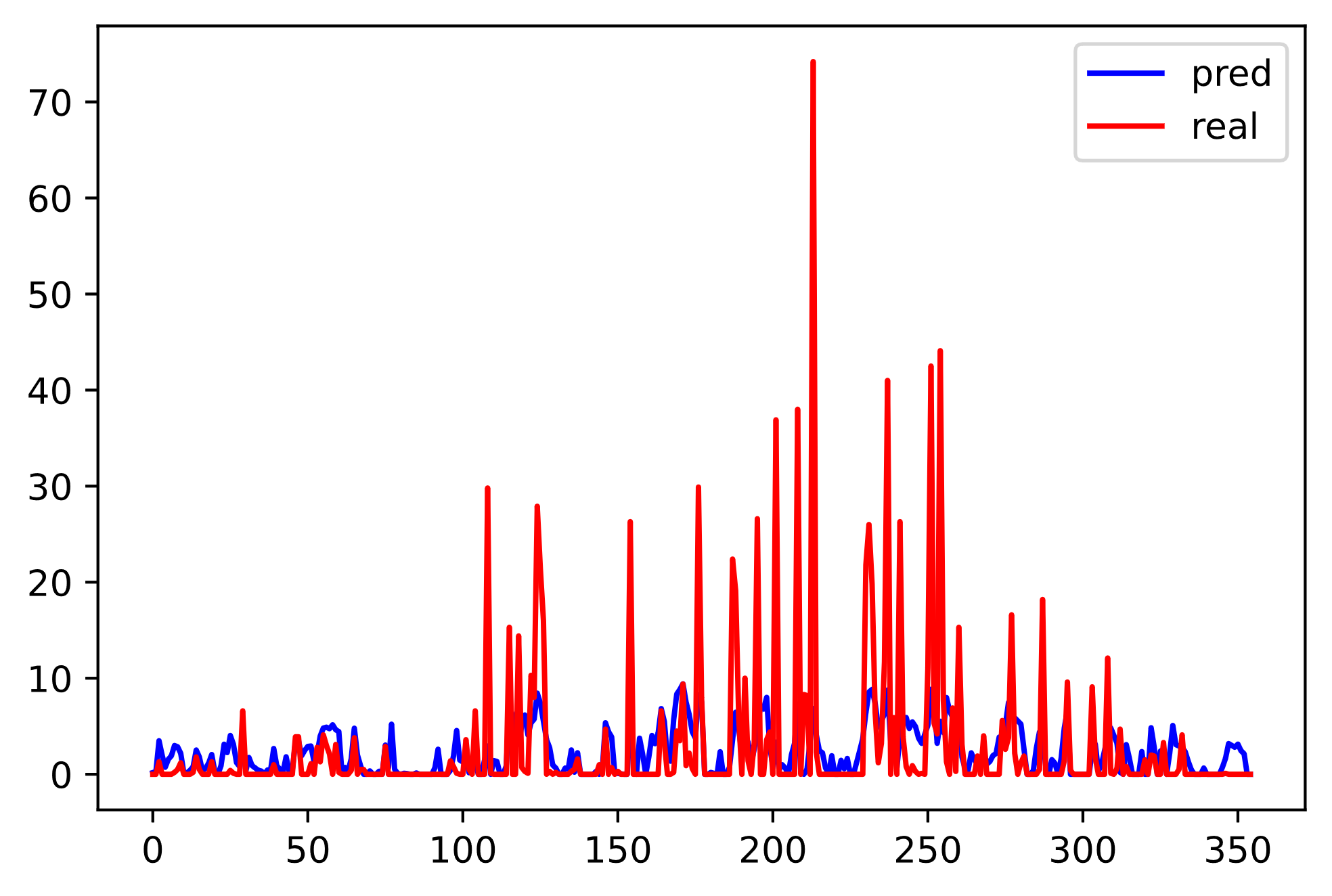


图 2019年预测降雨

红色曲线代表2019年的实际降雨量随日期的变化曲线，蓝色曲线代表模型预测的降雨量随日期的变化曲线。x轴表示2019年的第几天，y轴表示该日的降雨量。

由于我们在六盘山地区开展试验的时间为2020年的8月至11月，所以我们要着重观察下2019年同期的降雨预测效果。发现2019年8月到11月的真实降水量为508.20mm，使用模型预测的降水量为383.06mm，预测值与真实值相比，误差达到了32.67%。结合上面的降水量图像，我们可以发现，每年的降水量的分布是不均匀的，春冬季节降水量较小，大概在每年的第一百天左右开始，降水量突然增大，在200天左右达到峰值，第100天到第250天之间一直维持在较高水平。但是模型预测值随时间的变化并不明显，预测值一直在较低水平。对于夏季雨水的增加，模型不能即使地做出响应，导致不能预测到大的降雨周期。

因为降水存在有季节差异，在雨量少的季节中，日雨量中有较多零值，这样网络可能会偏向于使用零值，来作为网络的输出。为了凸显出降雨的季节差异，以及减少大量零值对神经网络的影响，尝试将日雨量作为预测值的标准进一步修改，改为预测月雨量。与之前每小时雨量整合为日雨量类似，将日雨量按照月份累加，得到月雨量。将气压、湿度等气象特征求平均值作为每月的特征值，重新进行模型训练，预测出来的降雨量如下表所示。

表 2019年预测结果

|  |  |  |  |
| --- | --- | --- | --- |
|  | 实际月平均降雨量 | 预测月平均降雨量 | 误差 |
| 2019年1-7月 | 71.31mm | 54.77mm | -30.20% |
| 2019年8-11月 | 124.97mm | 82.67mm | -51.17% |

从表中可以看到，预测结果与实际值相差较大。同时，预测值和真实值之间的R2值为0.55，这也验证了两者之间的差异。

分析月降雨量的预测结果，发现除了在2月左右有稍许偏高外，其他时间都低于实际值。而且七八月份实际降雨量会出现徒增，但是模型并没有发现这一规律。考虑到模型用到的特征值，我们猜想8月左右的强降雨天气不能被该模型识别，也就是说当前的特征的不足以让模型学会何时会发生强降雨。因此我们决定加入其他特征来进行训练，希望这些特征的加入，可以让模型“嗅”到强降雨到来前的味道。

因为我们对降雨的机制不了解，我们已经将与降雨有关联的天气数据都加入进行了训练，所以没有哪个天气的特征值加入会让模型大幅度提升。实验之前进行方案设计时，设计了区域回归实验，也就是使用其他没做实验的站点作为参考，来预测实验地区的自然降水量。而我们当前的工作是用实验站点的历史数据来进行预测的，我们可以将区域回归实验融入进来，将其他站点的降雨量作为特征值。要求这些地方不但没有受到人工降雨的影响，而且与实验地区相距不远，且气候条件非常相似。将这些加入作为特征值进行训练，可以让模型知道8月份左右会有强降雨发生，从而改善模型的效果。经过筛选，我们选择加入隆德气象站和泾源气象站同时期的降雨数据作为模型的特征值，重新训练模型。得到预测的降雨数据如图所示，训练过程中的loss曲线如图所示。

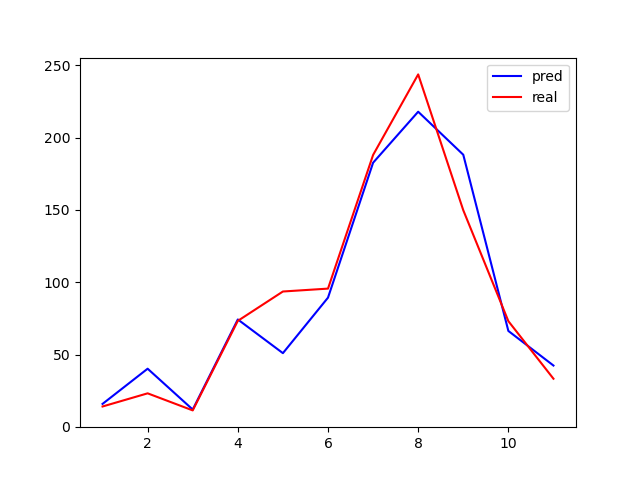


图 降雨曲线

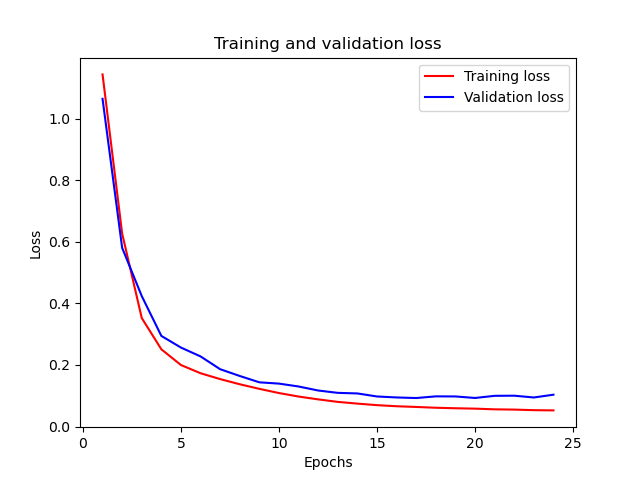


图 loss曲线

从雨量图中可以看到降雨量吻合的效果比较好，而图像的loss曲线平滑下降。具体结果如表所示。

表 2019年预测结果

|  |  |  |  |
| --- | --- | --- | --- |
|  | 实际月平均降雨量 | 预测月平均降雨量 | 误差 |
| 2019年1-7月 | 71.31mm | 66.47mm | -6.79% |
| 2019年8-11月 | 124.97mm | 128.70mm | +2.98% |
| 2019年1-11月 | 90.83mm | 89.11mm | -1.89% |

可以看到误差值都非常小，模型效果比较理想。注意到2019年8月到11月的预测值是大于实际值的，这说明模型不会得到较低的预测值来使实际值看起来虚高。如果结果如我们所愿，实际值更大，这是比偏高的预测值还要大，更有力地说明我们实验的有效性。

在改善模型的过程中，我们发现，固原、泾源气象站降雨量的加入，会让模型得到大幅度的提升。现在，我们重新着眼于日雨量，希望通过加入两个气象站降水量数据的模型可以成功地预测出日雨量。对模型的训练数据进行相应的调整，将月雨量修改为日雨量，重新训练，预测2019年的自然降雨量。结果如下：

表 2019年预测结果

|  |  |  |  |
| --- | --- | --- | --- |
|  | 实际日平均降雨量 | 预测日平均降雨量 | 误差 |
| 2019年1-7月 | 2.35mm | 2.29mm | -2.55% |
| 2019年8-11月 | 3.50mm | 3.71mm | +6% |
| 2019年1-11月 | 2.81mm | 2.86mm | +1.78% |

过程中loss曲线如图所示，训练过程中的loss曲线下降明显。且2019年的日降雨量的预测值和实际值的R2值为0.91。说明该模型对日雨量的预测比较精确。2019年的全年降水量的预测值和实际值相差1.78%，理想情况下该值应该接近0，但是这个数值不大，说明该模型比较合适。上半年的预测值比实际值低2.55%，数值较低，上半年的预测比较准确。下半年的预测值比实际值大了6%，这个值相对来说比较大，但是考虑到预测结果比实际结果是要高的，这说明模型不会得到较低的预测值来使实际值看起来虚高。如果实际值更大，这是比偏高的预测值还要大，更有力地说明我们实验的有效性。

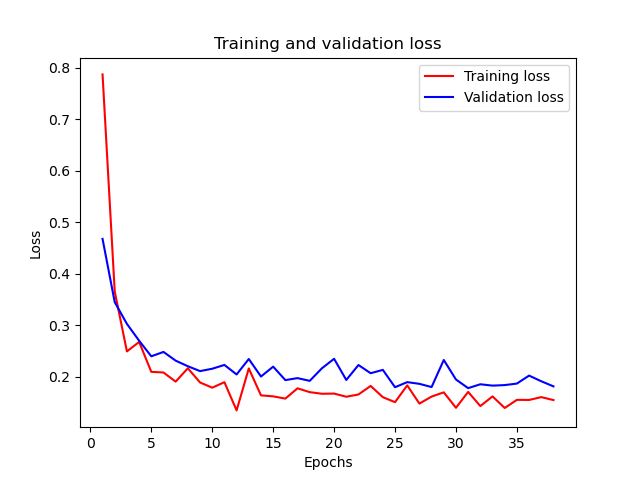


图 loss曲线

## 模型应用

使用改进后的模型，对六盘山气象站2020年的月降雨量和日降雨量进行预测。训练集为2008年到2017年的数据，包括六盘山气象站、固原气象站和隆德气象站的降雨数据以及六盘山气象站的气压、温度、湿度、风速、风向，选取这些作为神经网络的特征输入，使用2018年到2019年的降雨量进行验证，来得到最终需要的2020年的降水量的预测值。

### 预测月雨量

模型输入的雨量格式为月雨量，对模型进行训练，训练过程中的loss曲线如图所示。

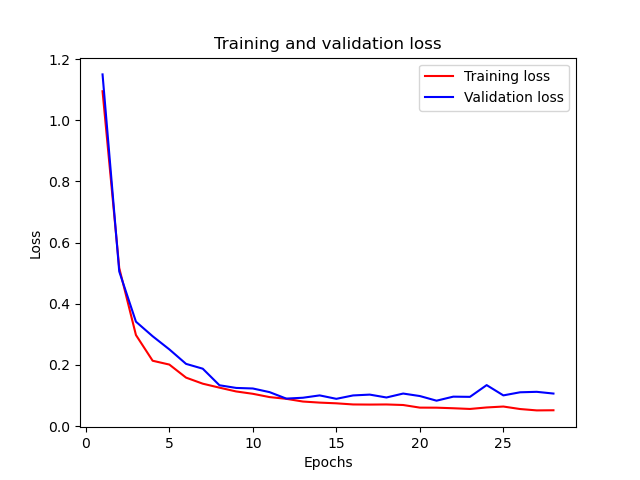


图 训练过程中的loss曲线

可以看到训练集的loss曲线和验证集的loss曲线都不断下降，说明模型对2018和2019年的预测结果是比较满意的。模型计算结果如下表所示。

表 2020年预测结果

|  |  |  |  |
| --- | --- | --- | --- |
|  | 实际月平均降雨量 | 预测月平均降雨量 | 误差 |
| 2020年1-7月 | 64.96mm | 64.82mm | -0.23% |
| 2020年8-11月 | 122.30mm | 103.99mm | -15.79% |
| 2020年1-11月 | 85.81mm | 79.03mm | -7.9% |

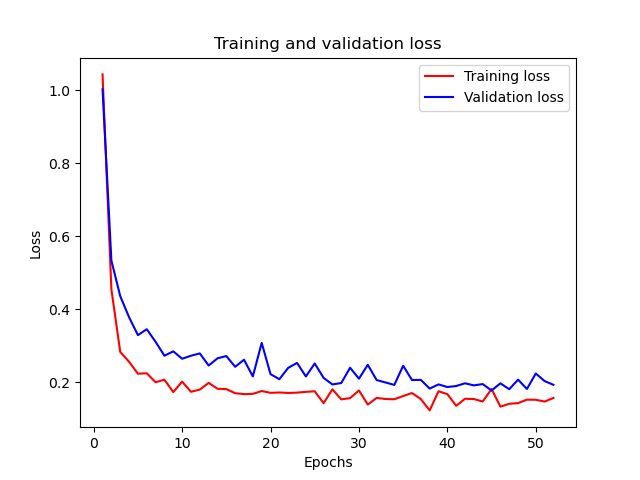
分析结果，可以看到全年的降水量实际值比预测值是大的，在保证预测值是可靠的前提下，这个结果是我们希望看到的。

观察降雨增量在上下半年的分布，看到2020年1月到7月预测值和真实值只相差0.23%，因为这期间没有进行实验，所以预测值和真实值理论上来说应该是相同的，这一数据也证明了这一点，再次证明了模型是可靠的。

下半年预测值比实际值小了15.79%，这说明下半年的降雨受到了除自然降雨以外的影响，也就是我们的实验起了效果。再结合之前预测2019年下半年降雨的结果来看，该模型会对下半年降雨量的预测值偏高，预测值比实际值高6%左右。但是在2020年，预测值与实际值低了15.79%，所以再考虑上模型的误差，这个实际值增加15.79%的这个值会更大，为21%左右。

### 预测日雨量

模型输入的雨量格式为日雨量，对模型进行训练，训练过程中的loss曲线如图所示。



2020日雨量loss

可以看到loss曲线整体下降，保持在了较低水平。模型计算结果如下表所示。

表 2020年预测结果

|  |  |  |  |
| --- | --- | --- | --- |
|  | 实际日平均降雨量 | 预测日平均降雨量 | 误差 |
| 2020年1-7月 | 2.14mm | 2.25mm | +6.54% |
| 2020年8-11月 | 4.01mm | 3.51mm | -12.47% |
| 2020年1-11月 | 2.82mm | 2.71mm | -3.9% |

综合2019年的结果，可以看到2019年的上半年预测值小了2.55%；在2020年，预测值大了6.54%；同时2019上半年的实际日降水量平均值为2.35mm，2020年为2.13mm。这两项数值表明，2020年上半年的雨量比2019年是有所降低的，(2.13-2.35)/2.35=-9.36%，即大概降低9.36%，也就是说自然降雨的影响下，2020年的上半年比2019年的雨水偏少，这样导致了2020年的预测值会有些偏大。

2020年下半年的预测值低了12.47%，我们注意到2019年的模型中下半年预测值大了6%，也就是说，该模型得到的下半年的预测值会偏高6%左右。假设模型在2020年下半年也有该误差，则扣除误差，增量大概为12.47%+6%=18.47%左右。按照实际值来看，下半年的实际日降雨量为4.01mm，而2019下半年的实际日降雨量为3.50mm，增加了12.47%。而根据上半年显示2020雨水少，上半年减少了9.36%，假设下半年也会减少9.36%，实际却增加12.47%，这表示降雨效果达到了9.36%+12.47%=21.47%。两个方面的证据都表示降雨的增加会达到20%。

我们再来分析一下2020年的降水量的残差图，该图显示的数据是用真实值减去预测值得来的。如下所示。可以看到，上半年的残差变化比较小，维持在0值附近，下半年的起伏比较大，而多为正值，这是实验带来的结果，这一图非常直观。

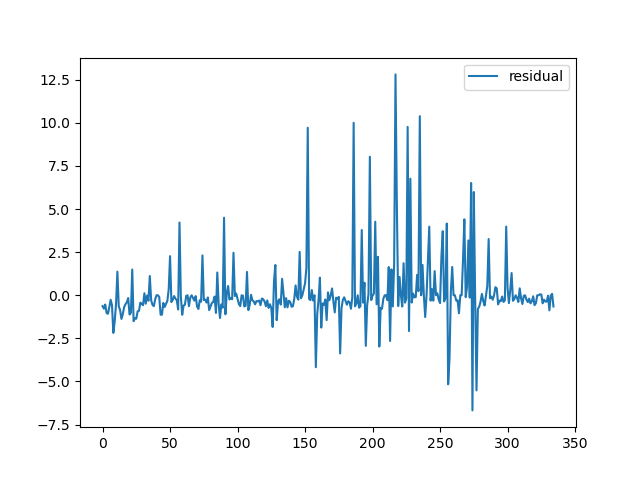


图 2020年日降雨量残差图

## 本章小结

1. 如果要预测出一个气象站的降雨量，其他气象站的降雨数据非常重要。仅仅从气压、温度、湿度、风等当地的自然气象条件很难判断出降雨的强弱。

2. 实验模型对日雨量和月雨量都进行了研究，在两个方面都得到了实验期间降水量增加了20%左右的结论。

3. 实验模型有2019年作为检测，具有比较高的可信度。且在2020年预测效果也符合预期。

# 基于LSTM的

# 基于统计分析的人工增雨效果评估模型

本章是针对“天水计划”自建站点收集的数据所作的评估分析，这些数据由于时间周期较短，不符合神经网络对数据量的要求，所以采用基于统计分析的方法，对随机对比试验的结果进行分析，并对实验区内部各雨量站的差异与风向建立联系。

## 效果的统计检验

统计检验的主要内容是显著性检验。如果实测雨量与估计雨量之间存在差异，则需要对这个差值进行统计检验，并指出由于降水的自然差异造成这个差值的可能性是多少。如果这个可能性很大，我们就不能认为人工的影响显著改变了雨量，即效果不显著；如果可能性很小，例如小于5%，则我们就有较大的把握认为人工影响是有效的，即效果显著。上述可能性大小即显著性水平，用α表示。

显著性检验在数理统计中又称为假设检验，在人工降水效果评估中常用的是t-检验法。t检验，也称为Student t检验，他适用于样本较小，样本标准差未知的正态分布。t检验法要求人工影响的措施只改变总体平均值，而不改变总体方差。如果这两个条件不满足，t-检验的依据就不成立。

## 随机对比试验

### 随机对比试验说明

“天水计划”在2020年确立了在乌鞘岭地区和六盘山地区开展随即对比实验。将两个试验区的带电粒子发生装置按照随机规则进行开关，每次开关周期为一天，装置开关次数比例为1:1，乌鞘岭地区从2020年7月开始实验，六盘山地区从2020年8月开始实验，两个试验区都在2020年11月结束实验。装置开启的日期作为随机对照试验的实验组，关闭的日期作为随机对照试验的对照组。“天水计划”在两个试验区分别安装了8个雨量筒，用来记录实验期间该地的雨量情况，通过清洗雨量筒的数据，得到该地实验期间的降雨。通过统计实验组和对照组的降雨量，来计算出实验组相比于对照组的增加量，同时还要对两组数据进行显著性检验，来推论差异发生的概率。

### 乌鞘岭随机对比实验结果分析

完全按照随机对照试验的原则，乌鞘岭试验区实际实验组数据为14天，对照组数据为13天。将这些数据清洗并统计之后，得到实验组合计14天的日降雨总和为176.74mm，日均降水量12.62mm，而对照组合计13天的日降雨总和为129.12mm，日均降水量为9.93mm，则由

可计算出实验期间的带电粒子催化降雨的效率为27.1%，达到了项目要求的20%的增雨效率。式中，A为增雨效率，R0为平均自然降雨（即对照组平均雨量），R1为平均催化降雨（即实验组平均雨量）。

针对这两组数据进行t检验，因为实验组的均值偏大，所以选择单尾检验，而选择定义t检验的类型时，选择了双样本等方差假设，得到的t检验结果为0.3743，也就是大概有37.43%的概率说明实验组和对照组的均值相同，不属于同一分布的概率为62.57%。

进一步地，由于随机实验采取以天为周期随机开展实验的方法，尽管其置信度较高，但其实验数据偏少且不连续，无法体现出装置长时间开启的增雨效果，故而继续对长期实验进行统计。选取研究时间为，将2020年8月24日到2020年8月30日，其间9月15日至9月22日期间由于装置闪络损坏未能正常开启，将这个时间段作为对照组，其余作为实验组进行相同的统计与分析。

结果显示实验组合计30天的日降雨总和为707.23mm，日均降水量23.57mm，而对照期间合计8天的日降雨总和为83.19mm，日均降水量为10.40mm。计算出实验期间的带电粒子催化降雨的效率为126.63%，远远超出了项目要求的20%的增雨效率。对这一数据同样进行t检验，得到结果为0.2293，也就是大概有22.93%的概率说明实验组和对照组的均值相同，不属于同一分布的概率为77.07%。

将上述所有的实验组数据合并在一起，所有对照组数据合并在一起，得到实验组共44天，对照组共21天。实验组的总降雨量为883.97mm，日平均降雨量为20.09mm。对照组的总降雨量为212.3mm，日平均降雨量为10.11mm。相对于对照组，实验组的日均降雨量高了99%，远高于要求的20%。t检验结果为0.1521，也就是大概有15.21%的概率说明实验组和对照组的均值相同，也就是说均值不同，不属于同一分布的概率为84.79%，这一值相对还是很高的。

上述分析了带电粒子发生装置开启与否对整个乌鞘岭实验区带来的影响，接下来分析实验区内部的降雨差异。将各个雨量站的数据分开，按照各个站点的实际经纬度位置标注起来，如图所示。

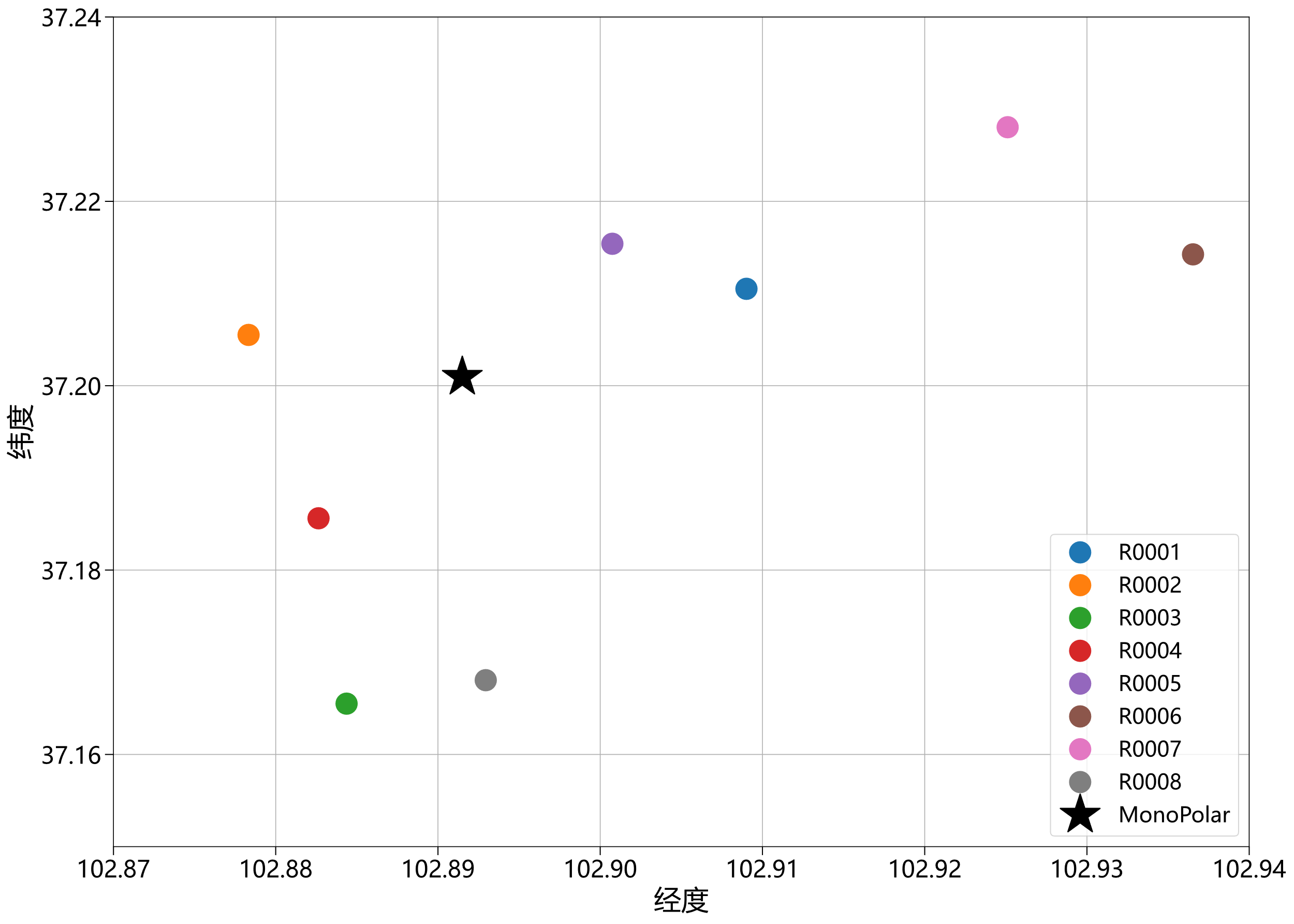


图 乌鞘岭各个雨量站位置示意图

我们将东经102.87°~ 102.94°，北纬37.15° ~ 37.24°的范围作为乌鞘岭地区的实验区，按照纬度和经度将试验区划分为网格，将雨量站的数据按照各自实际的纬度和经度写入到网格内。对于没有数据也就是该地没有雨量站的网格，使用Scipy中的griddata方法，对未知点进行二维插值。根据已知的雨量站数据估算出该地的降雨数据。分别对实验组和对照组进行插值操作，将插值后的数据绘制成heatmap图像，如下图所示。

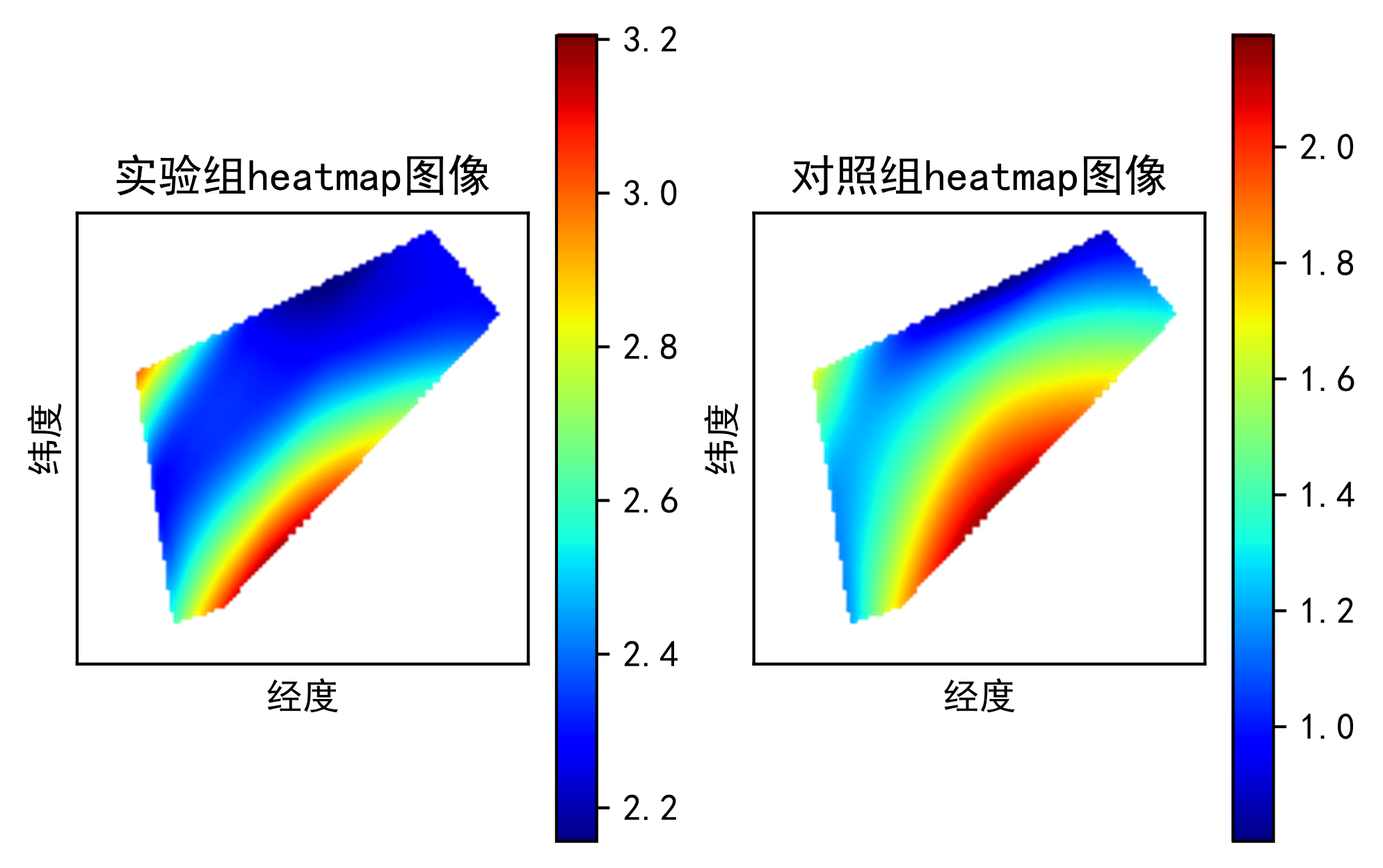


图 乌鞘岭试验区的降雨插值结果

对于地形网格来说，有数据的网格很少，二维插值方法需要用到离所求点距离较近的有效数据，所以对于远离雨量点的区域，插值结果为0，整体插值结果表现为雨量站所在地附近范围生成有效的插值数据。从图上看，有效数据范围的轮廓与雨量站集合轮廓基本吻合，也就是对实验区进行了插值处理，这一部分是实验的核心区域，也是我们重点探讨的范围。我们近似地将该地区的降雨插值结果看作该地区的降雨分布，通过观察上图，我们可以容易地发现，实验组的插值结果相比对照组来说图像颜色较深。为了更加直观地观察实验组相对对照组的增加量，我们将实验组的插值结果减去对对照组的插值结果，得到降雨增长分布，将他画为heatmap图像，如下图所示。

图 试验区降雨增长的heatmap图像

从图中我们可以直观的看出，在实验期间，乌鞘岭实验组各地降雨量都有不同程度的增加。

### 六盘山随机对比实验结果分析

与乌鞘岭单电极随机对比试验相似，需要选取实验组和对照组，通过比较雨量的增量，以及进行显著性检验来分析降雨效果。但是与乌鞘岭单电极不同的是，六盘山地区考虑到试验周期，对试验方案进行了调整，在实验期间，所有装置功率全开，这样导致了随机对比试验没有了对照组，所以选择了没开装置的7月份的降雨数据作为对照组。

由于装置始终开启，能获取到的没有装置开启的时间只有几天。再考虑12小时的缓冲时间，则8、9月没有可以用来做同等时间尺度下对比的数据。故而选用7月的雨量数据与8、9月实验期间的进行对比。

经过数据的清洗和整理，发现7月雨量缺测比较严重，1号与8号雨量筒基本没有采集到数据，其余雨量筒只有前半月有数据。所以不再考虑1号和8号雨量筒，将2号到7号雨量筒的数据进行加和，作为六盘山地区的降雨数据。除去缺测的情况，在8月和9月中清洗出了54天作为催化降雨的实验组，选择7月1日到7月13日共13天的数据作为对比区。

根据统计结果并进行对比可知，实验期间合计54天的日降雨总和为1100.89mm，日均降水量20.39mm，而对照期间合计8天的日降雨总和为234.24mm，日均降水量为18.02mm。可计算出实验期间的带电粒子催化降雨的效率为13.15%，相比项目要求的20%的增雨效率尚有一定距离。对两组数据进行t检验，结果为0.4184，也就是大概有41.84%的概率说明实验组和对照组的均值相同，也就说两个均值不同，不属于同一分布的概率为58.16%。

上述分析了带电粒子发生装置开启与否对整个六盘山实验区带来的影响，接下来通过将各个雨量站的数据分开，按照各个站点的实际经纬度位置标注起来，如图所示。

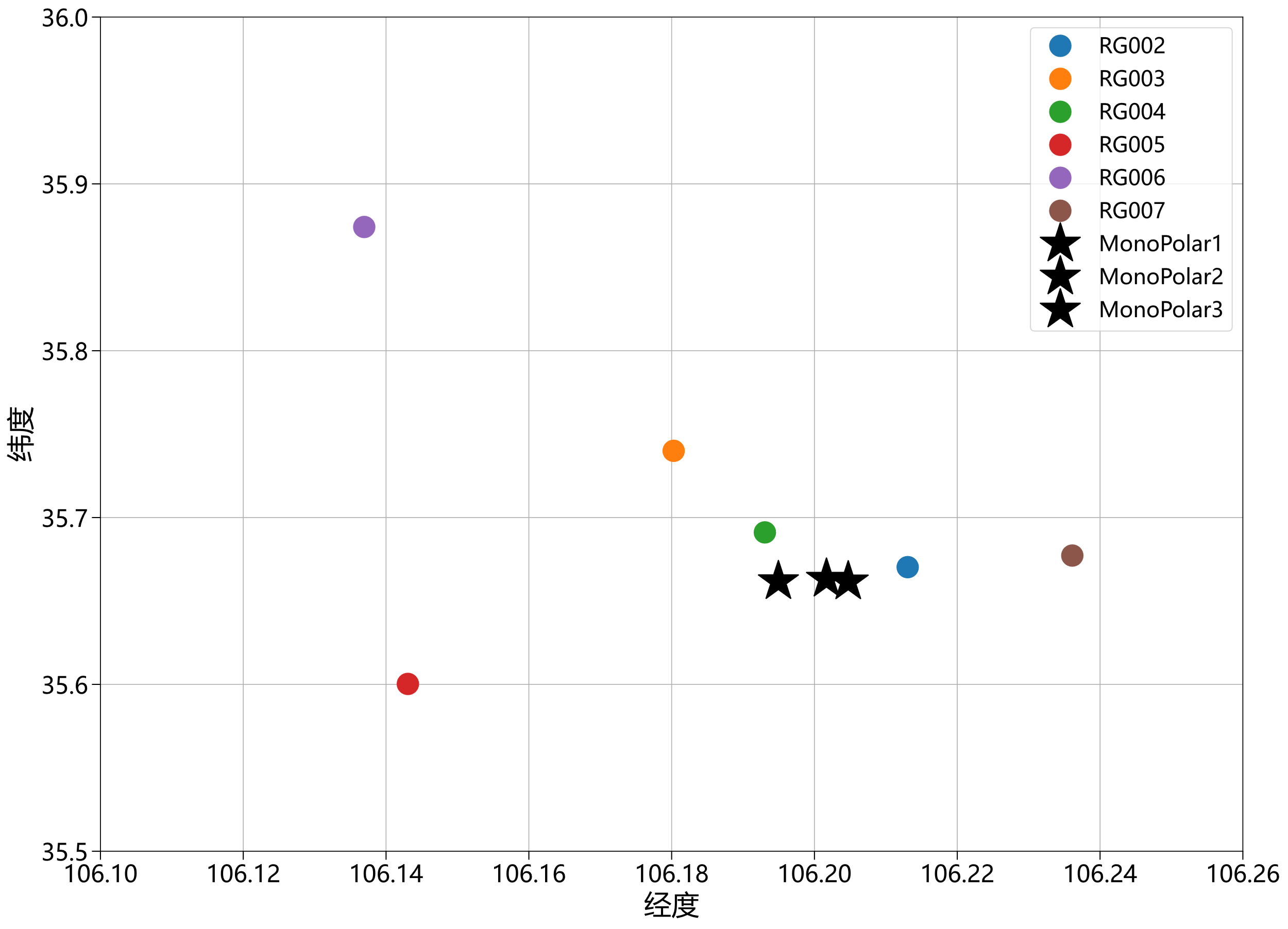


图 六盘山各个雨量站位置示意图

我们将东经106.10°~ 106.26°，北纬35.5° ~ 36.0°的范围作为六盘山地区的实验区，按照纬度和经度将试验区划分为网格，将雨量站的数据按照各自实际的纬度和经度写入到网格内。对于没有数据也就是该地没有雨量站的网格，使用Scipy中的griddata方法，对未知点进行二维插值。根据已知的雨量站数据估算出该地的降雨数据。分别对实验组和对照组进行插值操作，将插值后的数据绘制成heatmap图像，如下图所示。

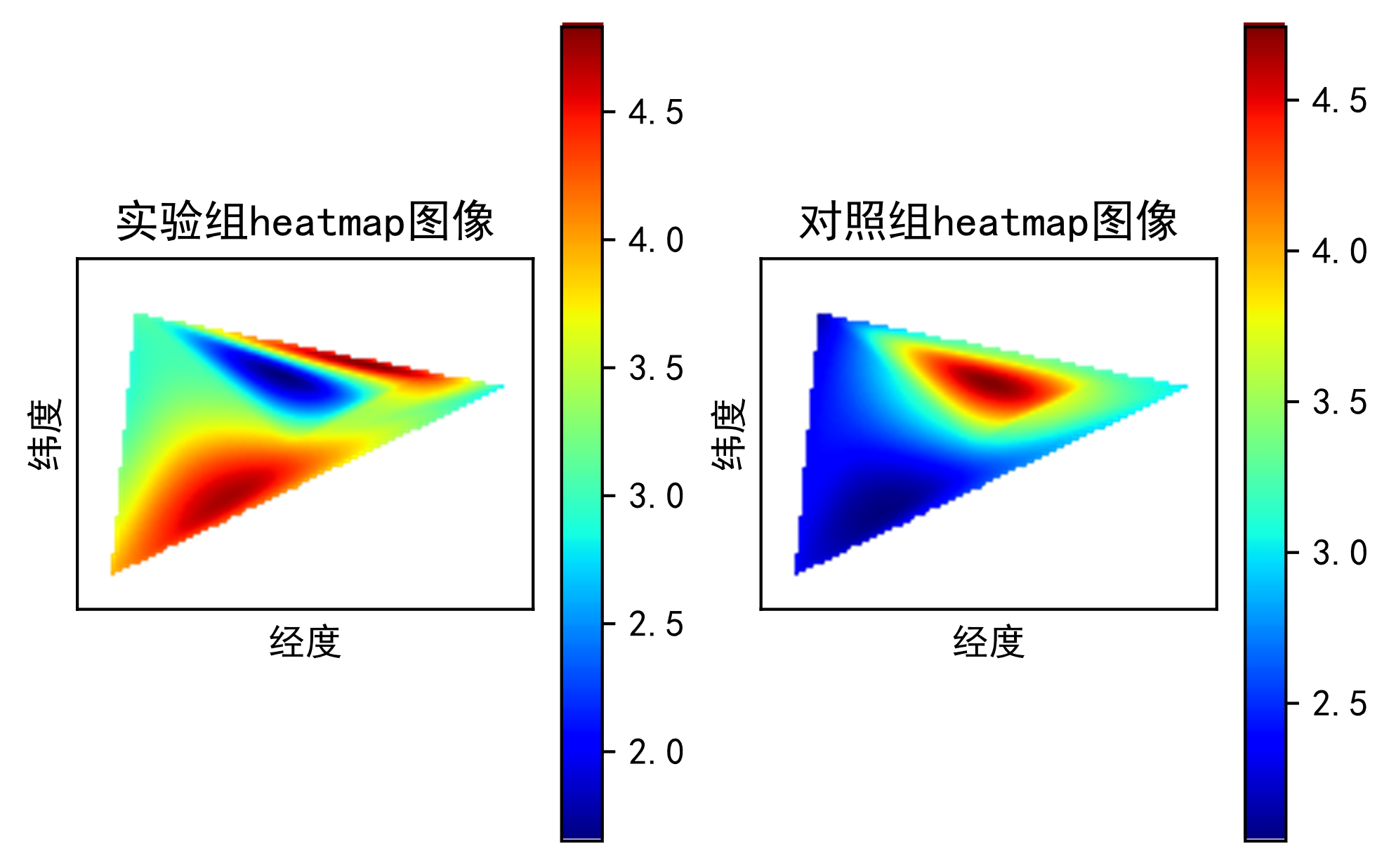


图 六盘山试验区的降雨插值结果

与乌鞘岭地区相似，图中有效数据范围的轮廓与雨量站集合轮廓基本吻合，这一部分试验的核心区域，是我们重点探讨的范围。通过观察上图，可以发现，除三角区域的中心未知以外，实验组的图像颜色比对照组的更深。为了分析原因，我们列举出六盘山的6个有效雨量站的实验组和对照组数据，如下表所示。

表 六盘山有效站点雨量说明

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | R0002 | R0003 | R0004 | R0005 | R0006 | R0007 |
| 实验组 | 4.386284 | 4.045162 | 2.352403 | 3.108337 | 3.955586 | 2.541928 |
| 对照组 | 3.335749 | 2.835177 | 4.491484 | 2.042235 | 2.330283 | 2.935451 |

可以看到除了R0004和R0007站点以外，其他站点的实验组降雨量均比对照组降雨量多，而这两个站点在试验核心区域的位置正是在右上角，与图像表现出来的信息相符。这是实验实际存在的现象，但是考虑到随机试验持续时间不长，具有偶然性，对于两个站点的异常属于可以接受的范围。为了更加直观地观察实验组相对对照组的增加量，我们将实验组的插值结果减去对对照组的插值结果，得到降雨增长分布，将他画为heatmap图像，如下图所示。

图 试验区降雨增长的heatmap图像

从图中我们可以直观的看出，除右上角区域外，核心区域内降雨其他地点的降雨量都有所增加，此结果与上述分析相符合。

## 风向影响的降雨差异分析

### 风向影响降雨的总体说明

带电粒子生成装置在工作过程中，风的影响是很大的。生成的带电粒子需要在风的作用下扩散出去，然后来催化该地区的降雨。如果没有风，带电粒子的扩散速度很慢，就会导致装置的影响范围很小，只能影响距离装置很近的地方。如果有风，但是风向是固定的，那么理论上，下风处的降雨催化效果要比上风处的催化效果要好，因为下风处获得了更多的带电粒子。以上的推断都建立在带电粒子生成器正常工作且有效的前提下。

为了验证实际上带单粒子生成器的有效性，我们分析了乌鞘岭地区和六盘山地区的自建雨量站点的降雨数据，以及综合站的记录下的风向数据，来探究在装置生成器附近的风向与降雨催化效果之间的关系。

乌鞘岭地区和六盘山地区各有8个雨量站和若干个综合测试站，综合测试站记录了该地每小时的风向风速信息。雨量站的位置不同，收集到的降雨数据不同，以降雨数据为核心，通过分析风向风速的变化与降雨变化之间的关系，找到上风处和下风处的雨量站降雨差异，根据各个雨量站与带电粒子发生器之间的位置关系，找出风向影响降雨是否与各自相对位置之间的联系。

### 数据处理

风向影响降雨实验的数据处理主要分为降雨数据和风向风速数据。雨量站和综合站点采集的数据是按照每小时采样的，但是降雨过程往往会持续数个小时，且负离子传播催化是需要时间的，研究风向在一个小时内对不同点之间的降雨影响是不合适的。所以我们选择以天为单位进行研究。

降雨量为标量，按照天为单位进行研究时，可以将24小时的降雨直接加和，作为该站一天的降雨量。但是为了尽可能地消除不同站点由于自身原因带来的各站之间降雨的差异，凸显出风向对不同站点降雨数据的影响，我们需要找出8个雨量站之间自然存在的降雨差异。最简单的方法就是根据未进行实验时的降雨数据，分别求出各个站的降雨均值，然后用实验期间的降雨数据减去这一均值，就得到了催化的效果，公式如下：

其中是实际降雨量，是根据未进行实验时的降雨数据计算出的平均将欲练，是催化的雨量。

综合站点收集的风向数据是按照角度为单位记录的，是一个矢量，24小时的风向角度不同。我们认为风向代表了催化的方向，如果我们将24小时的风向看成矢量进行简单的矢量运算显然是不合适的，因为合成的矢量并不是风吹的方向。考虑到一天之中，风向会有一个大致方向，也就是主导风向。因此我们将360度的各个方向分为16份，将每小时的风向确定为其中的一份，按照每天为单位，统计出每天的主导方向，以此作为主导风向，也就是催化的主要方向。

### 乌鞘岭试验区的详细分析

乌鞘岭试验区的带电粒子发生装置附近除了有8个自建的雨量计站外，还有4个综合站点，这些站点记录了该地每小时的降雨数据和风向风速数据。经过数据处理、分析，选择了数据较为完整的W0003综合站点的风向数据作为该地区的风向数据，但是该站只有2020年的7、8、9三个月的数据。为了消除8个雨量站点本身的降雨差异，先求出各雨量站的自然降雨量。

因为雨量站在2019年5月已经建好并开始记录数据，所以统计计算出2019年5月到2020年6月的降雨信息。同时，因为乌鞘岭地区进行了随机对照试验，所以在2020年7月之后仍然有几天未开装置。而且2020年9月15日至9月22日装置出现故障，也未进行试验。将这些日期的降雨数据统计在一起，得到每个站点的平均降雨数据如下表所示。

表 雨量站的自然降雨数据日均值

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 雨量站 | R0001 | R0002 | R0003 | R0004 | R0005 | R0006 | R0007 | R0008 |
| 自然降雨量 | 3.213 | 3.825 | 3.378 | 2.976 | 2.832 | 3.390 | 3.039 | 3.657 |

将试验期间的各个站点的实际降雨数据，减去各自的日均自然降雨量，就得到了催化效果降雨数据。观察实验期间的催化增雨效果时将各个站点均无降雨的日期删除，因为实际降雨都是0时，再减去自然降雨量之后，各站的数据均为负值且都不相等，容易做出错误的判断。

求风向时，因为乌鞘岭地区只有W0003气象站的风向信息能用，所以将W0003的风向信息作为单电极的风向信息，找出主导方向后，需要以单电极装置为界限确定上风处和下风处，计算出上风处的站点的平均增雨量和下风处的站点的平均增雨量，再计算出下风处增量相比上风处增量的增长值。发现在很多日期中，下风处的雨量增加量都比上风处的雨量增加量要大，如下表所示。

|  |  |  |  |
| --- | --- | --- | --- |
| 日期 | 上风处站点雨量平均增量 | 下风处站点雨量平均增量 | 下风处增量相比上风处的增幅 |
| 2020/7/1 | 2.888151 | 5.495293 | 90% |
| 2020/7/11 | 0.406718 | 1.272235 | 213% |
| 2020/7/16 | -6.21792 | -4.40449 | 29% |
| 2020/7/17 | 7.790935 | 12.77013 | 64% |
| 2020/7/18 | 11.34078 | 23.34237 | 106% |
| 2020/7/23 | 28.07284 | 35.03279 | 25% |
| 2020/7/24 | -11.0821 | -7.57941 | 32% |
| 2020/8/3 | -1.88452 | -0.07358 | 96% |
| 2020/8/7 | -8.20118 | -4.61615 | 44% |
| 2020/8/13 | -12.6059 | -10.9839 | 13% |
| 2020/8/14 | 48.93238 | 51.39373 | 5% |
| 2020/8/17 | -0.75605 | 3.574104 | 573% |
| 2020/8/19 | -12.4133 | -11.381 | 8% |
| 2020/8/22 | 8.685797 | 13.9958 | 61% |
| 2020/8/23 | 0.334511 | 5.389129 | 1511% |
| 2020/8/29 | 99.64446 | 105.7174 | 6% |
| 2020/8/30 | -6.61201 | -3.06868 | 54% |
| 2020/9/5 | -5.51891 | -3.45382 | 37% |
| 2020/9/8 | 17.3028 | 18.71562 | 8% |
| 2020/9/13 | 21.13172 | 29.57258 | 40% |
| 2020/9/14 | 16.93538 | 30.81197 | 82% |
| 2020/9/17 | -13.1992 | -11.8777 | 10% |
| 2020/9/21 | 14.32866 | 33.30321 | 132% |
| 2020/7/9 | -0.7750 | -2.8075 | -262% |
| 2020/7/10 | -3.0225 | -3.4250 | -13% |
| 2020/7/13 | 1.1600 | 0.5920 | -49% |
| 2020/7/22 | -1.3300 | -2.8750 | -116% |
| 2020/7/25 | -3.0675 | -3.4050 | -11% |
| 2020/7/27 | -2.9675 | -3.1025 | -5% |
| 2020/8/1 | -3.0425 | -3.4050 | -12% |
| 2020/8/4 | 2.7675 | 2.6325 | -5% |
| 2020/8/5 | 9.1600 | 7.3650 | -20% |
| 2020/8/6 | -3.0175 | -3.3525 | -11% |
| 2020/8/8 | -3.1175 | -3.4025 | -9% |
| 2020/8/9 | -2.4100 | -2.4900 | -3% |
| 2020/8/11 | -3.1175 | -3.3050 | -6% |
| 2020/8/12 | 1.7500 | 1.4850 | -15% |
| 2020/8/16 | -2.8200 | -2.9300 | -4% |
| 2020/8/21 | 6.3700 | 6.3500 | 0% |
| 2020/8/31 | 0.0575 | -1.3275 | -2409% |
| 2020/9/1 | -2.5700 | -1.1200 | 56% |
| 2020/9/6 | -3.1500 | -3.1460 | 0% |
| 2020/9/9 | 1.1775 | 1.1750 | 0% |
| 2020/9/11 | -3.0675 | -3.4525 | -13% |
| 2020/9/15 | -3.1175 | -3.4275 | -10% |
| 平均值 | 4.0256 | 6.3461 | 58% |

从表格中我们可以看到，上半部分的是下风处雨量增长量更多的日期，下半部分是上风处雨量增长量更多的日期。所有日期加起来的平均雨量为：上风处雨量站的平均增加雨量为4.0256mm，下风处雨量站的平均增加雨量为6.3461mm，下风处的增长量比上风处的增长量大58%，也就是说下风处的催化效果比上风处强58%，这说明了催化效果与风向关系很大，也证明了我们一开始的假设是正确的。同时，不管上风处还是下风处，雨量增加量均是正值，证明了装置的有效性。

下面的几张图是从选取下风处催化效果更好的几个日期的数据绘制的。图中的散点代表了8个雨量站的具体地理位置，数值代表了该日降雨量减去该站的平均降雨量，黑色的五角星是乌鞘岭单电极装置的位置，红色的五角星是综合气象站W0003的位置，由它的数据来计算风向。右下角是风向玫瑰图。图中的带箭头的射线代表了风，箭头的方向是风向，线段长度代表风速的大小。

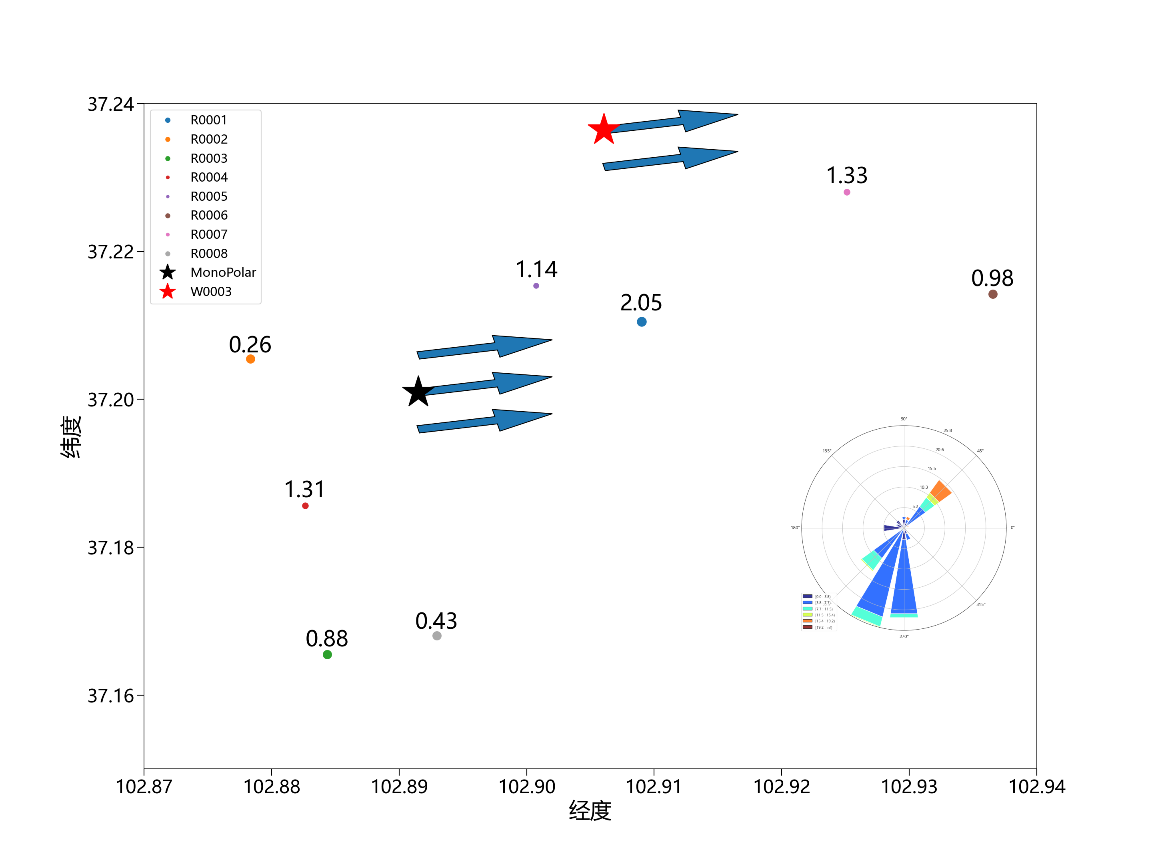
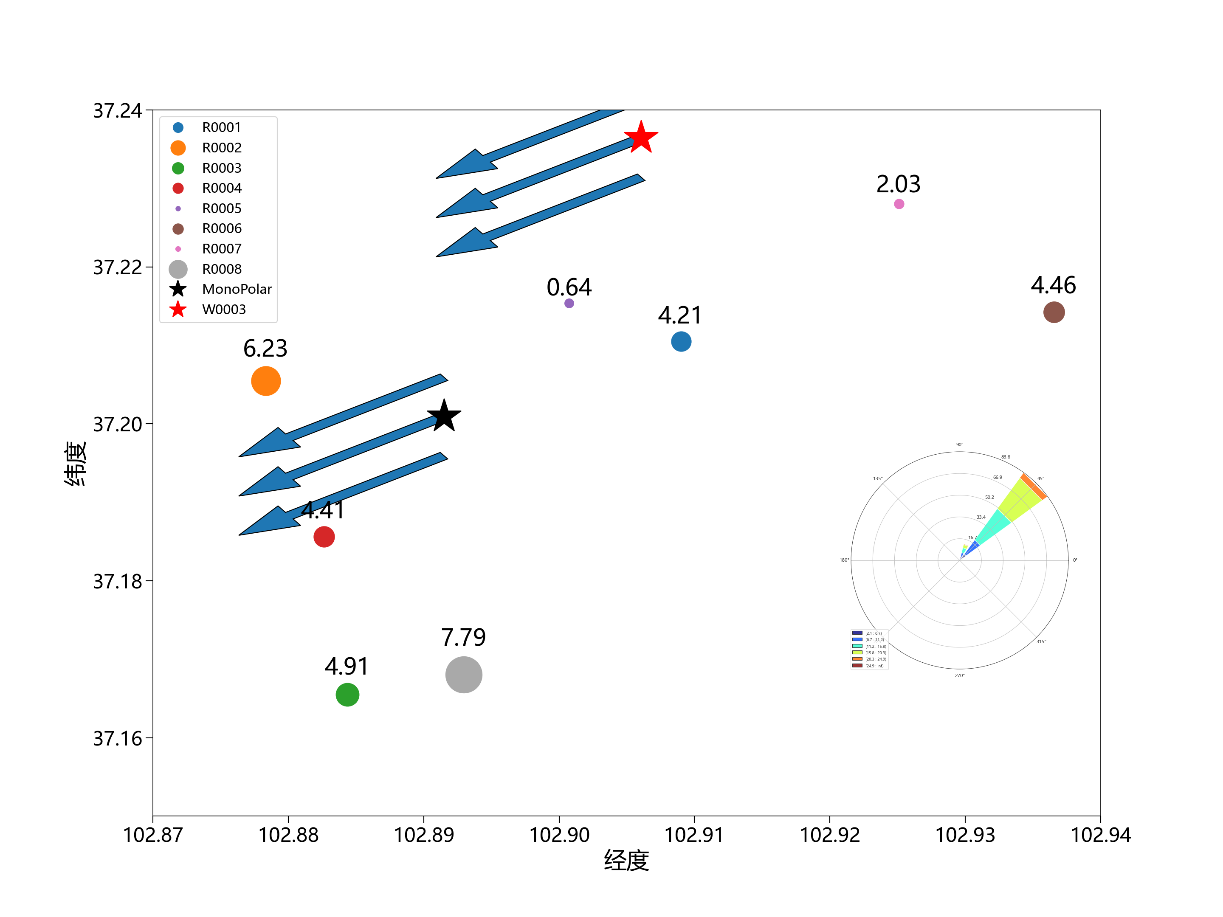


图 7月1日风向雨量图



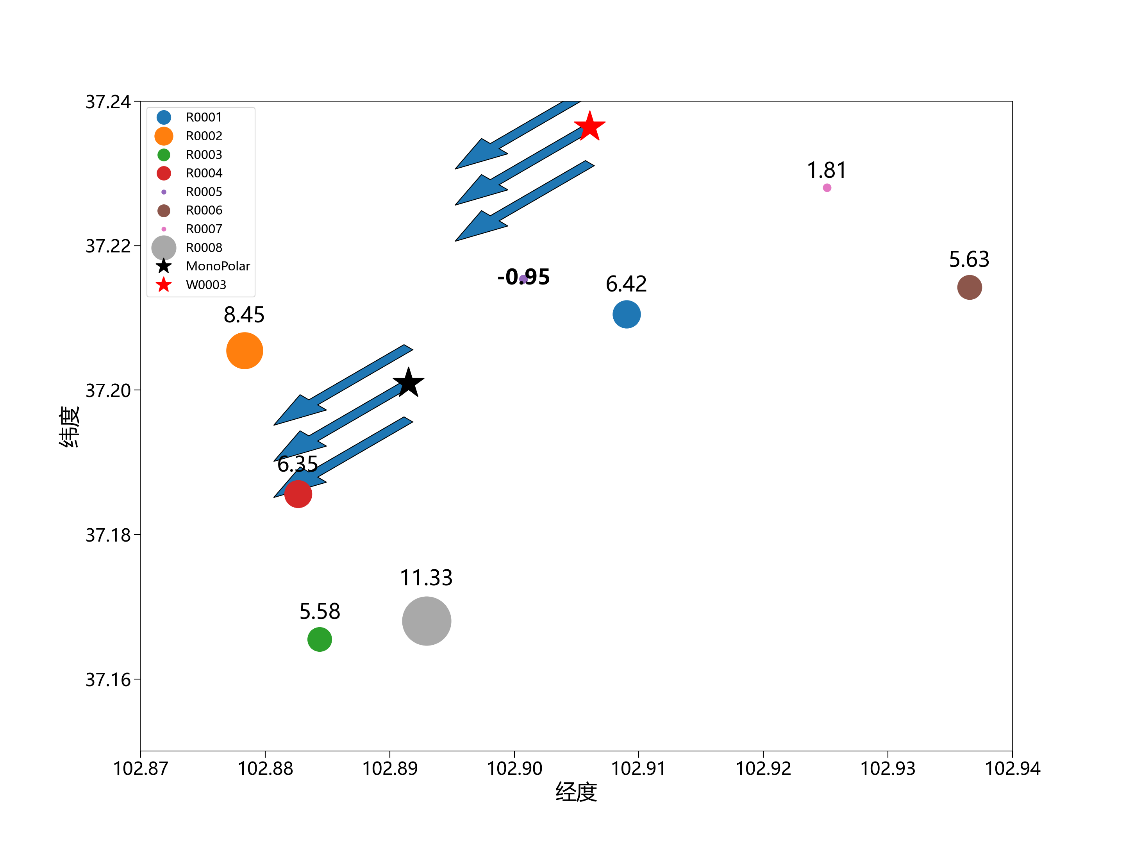
7月18日

为了使我们的结果更具有说服力，我们将研究周期从一天改为了一个降雨过程，我们把8个雨量站的降雨全为0定义为一个降雨过程的结束，把一个雨量站测得降雨量定义为一个降雨过程的开始。将三个月的数据进行分析划分，总结出了以下几个降雨过程，如下表所示。

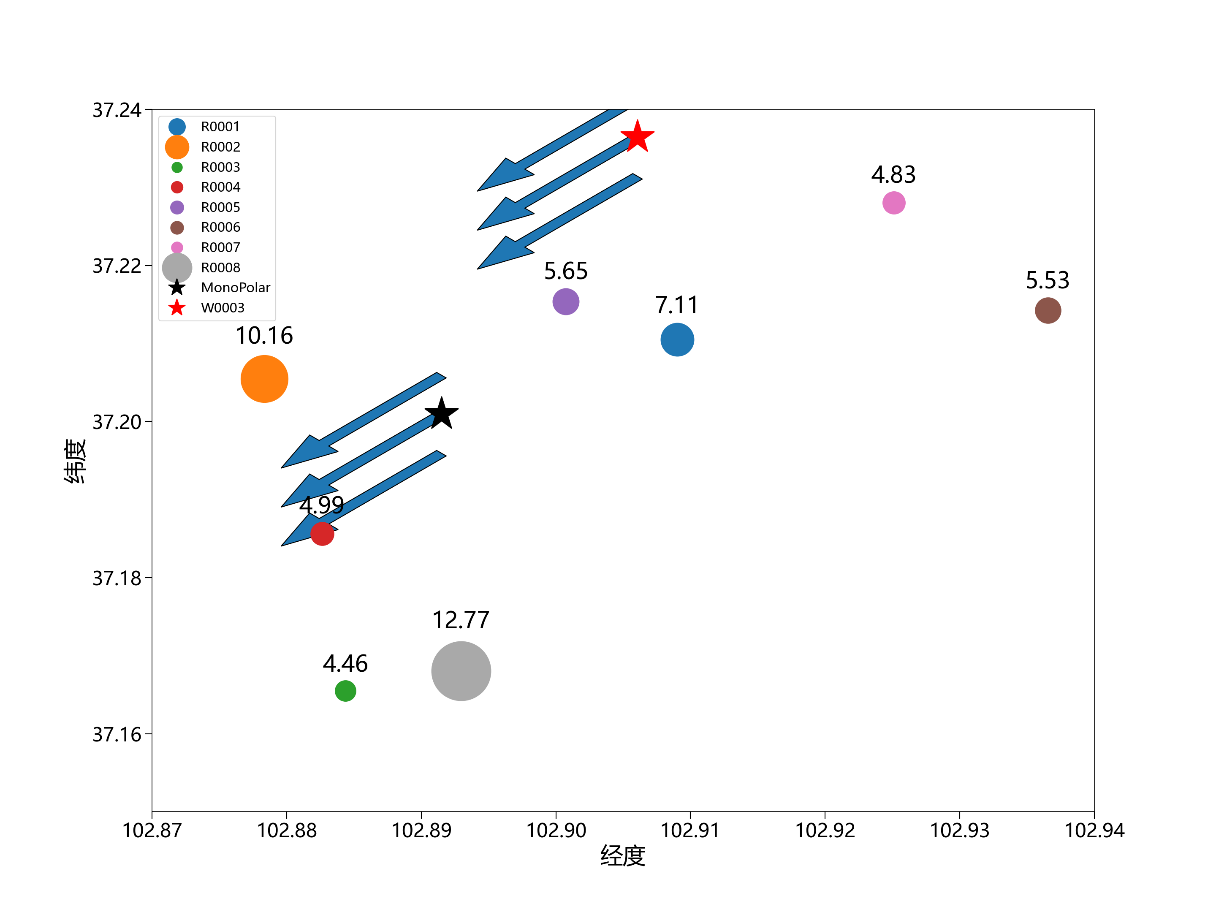
|  |  |  |  |
| --- | --- | --- | --- |
| 降雨过程 | 上风处站点雨量平均增量 | 下风处站点雨量平均增量 | 下风处增量相比上风处的增幅 |
| 0709-0713 | -1.89 | -4.01 | -112% |
| 0716-0718 | 3.228449 | 7.927001 | 146% |
| 0722-0727 | -3.34571 | -2.51216 | 25% |
| 0803-0809 | 1.1975 | -0.7725 | -165% |
| 0811-0817 | 5.7275 | 5.23 | -9% |
| 0819-0823 | 5.778832 | 8.092871 | 40% |
| 0829-0901 | 19.35542 | 22.88791 | 18% |
| 0905-0909 | 1.383653 | 1.437085 | 4% |
| 0911-0917 | 0.360255 | 4.923628 | 1267% |
| 平均值 | 3.5329 | 4.8004 | 36% |

可以看到在9个降雨过程中，有6个降雨过程中，下风处的降雨增加量要大于上风处。从平均值来看，在9个降雨过程中，上风处的降雨平均增加量为

3.5329mm，下风处的降雨平均增加量为4.8004mm，下风处增量相比上风处增量增加了36%，也就是说下风处比上风处的催化效果强了36%左右。同时，下风处和上风处的降雨增加量均为正值，证明了装置的有效性。以下是降雨过程的几张图片。



7月16日到7月18日



8月19日到8月23日

### 六盘山试验区的详细分析

六盘山试验区的单电极装置附近有8个自建的雨量计站点，以它们在实验期间收集的降雨数据为核心进行比较，主要到试验开始后RG001雨量计没有降雨数据记录下来，所以实际上选择RG002~RG008共7个雨量计站点的数据。除了雨量站点之外还有3个综合站点，为 WG005、WG006、WG007。其中WG006建在单电极装置处，所以以该站的风向数据为主要风向参考数据，将WG005和WG007的数据作为辅助风向参考。为了消除8个雨量站点本身的降雨差异，先求出各雨量站的自然降雨量。

观察到雨量站在2020年5月开始记录数据，六盘山从8月开始做实验。所以统计计算出2020年5月到2020年7月的降雨信息将这些日期的降雨数据统计在一起，得到每个站点的平均降雨数据如下表所示。

表 雨量站的自然降雨数据日均值

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 雨量站 | RG002 | RG003 | RG004 | RG005 | RG006 | RG007 | RG008 |
| 自然降雨量 | 3.718 | 7.584 | 4.929 | 3.544 | 2.760 | 3.701 | 2.321 |

将试验期间的各个站点的实际降雨数据，减去各自的日均自然降雨量，就得到了催化效果降雨数据。观察实验期间的催化增雨效果时将各个站点均无降雨的日期删除，因为实际降雨都是0时，再减去自然降雨量之后，各站的数据均为负值且都不相等，容易做出错误的判断。此处与乌鞘岭地区还有个不同点，由于乌鞘岭地区的雨量站在未做试验前有13个月的降雨数据，所以求出的平均值比较可靠。但是六盘山地区未做试验的数据只有3个月的，且在5~7月份雨水较多，所以求得的各个的平均值较高。我们的实际降雨量数据减去平均值后会有较多的负值。为了减小带来的影响，探究催化效果时，不仅删去了没有降雨的日期，而且删去了只有一个站有少量降水的日期，减少了减去大的平均值带来的对实验结果的误判。

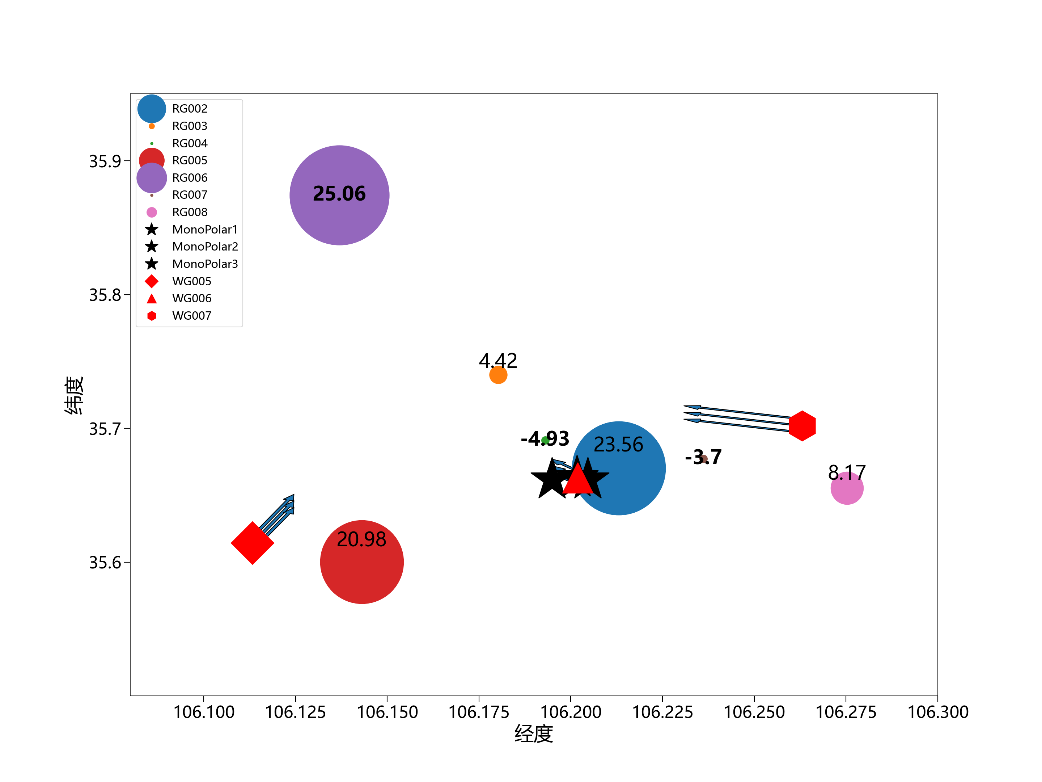
求风向时，因为WG006建在单电极装置处，所以以该站的风向数据为主要风向参考数据，将WG005和WG007的数据作为辅助风向参考。找出主导方向后，需要以单电极装置为界限确定上风处和下风处，计算出上风处的站点的平均增雨量和下风处的站点的平均增雨量，再计算出下风处增量相比上风处增量的增长值。如下表所示。

表

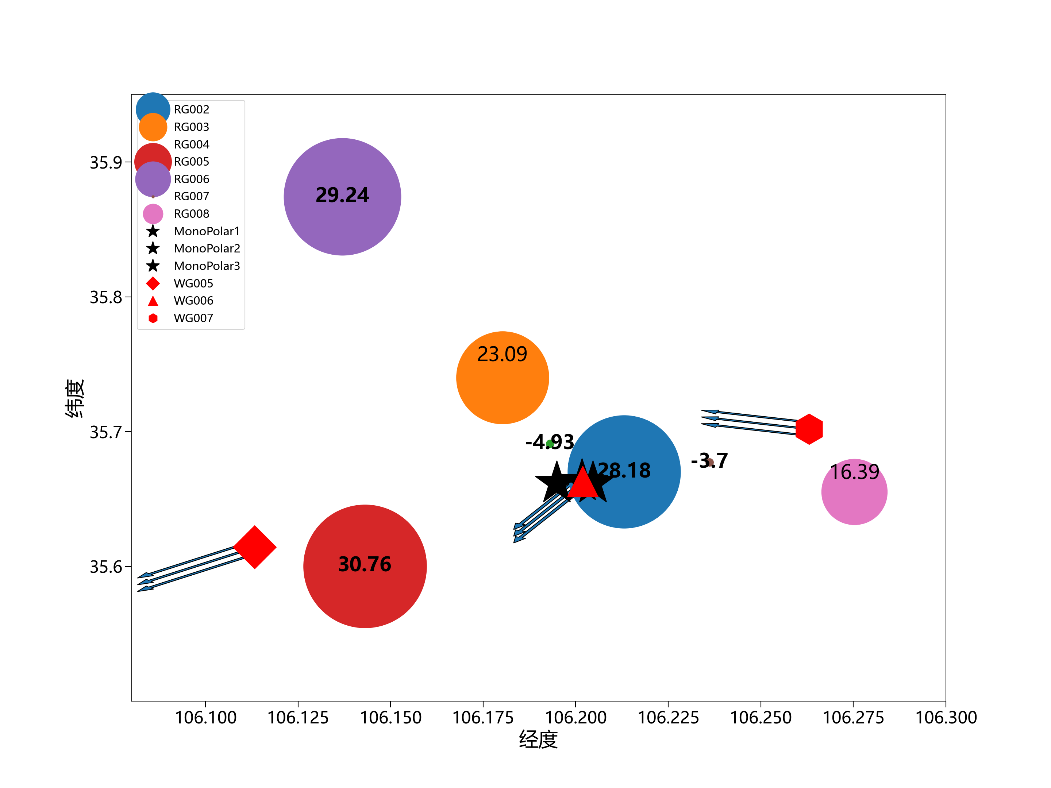
|  |  |  |  |
| --- | --- | --- | --- |
| 日期 | 下风处站点雨量平均增量 | 上风处站点雨量平均增量 | 下风处增量相比上风处的增幅 |
| 2020/8/4 | -3.614 | -1.765 | -105% |
| 2020/8/5 | 6.66 | 1.875 | 255% |
| 2020/8/6 | 12.0275 | 8.483333 | 42% |
| 2020/8/8 | -3.575 | -3.08667 | -16% |
| 2020/8/10 | 2.285 | 0.913333 | 150% |
| 2020/8/11 | -1.12 | -2.15167 | 48% |
| 2020/8/12 | 16.28667 | 17.5875 | -7% |
| 2020/8/13 | -3.53 | -4.258 | 17% |
| 2020/8/14 | 12.644 | 14.31 | -12% |
| 2020/8/15 | -1.86 | -1.562 | -19% |
| 2020/8/16 | 19.275 | 13.97667 | 38% |
| 2020/8/17 | 28.8125 | 25.07333 | 15% |
| 2020/8/21 | -4.2 | -2.205 | -90% |
| 2020/8/22 | -3.225 | -1.80667 | -79% |
| 2020/8/23 | 24.39 | 23.94 | 2% |
| 2020/8/25 | -2.7975 | -2.34333 | -19% |
| 2020/8/26 | -4.452 | -1.43 | -211% |
| 2020/8/27 | -4.378 | -2.28 | -92% |
| 2020/8/29 | 6.28 | 6.93 | -9% |
| 2020/8/30 | 17.15 | 12.1875 | 41% |
| 2020/8/31 | -3.76333 | -4.0225 | 6% |
| 2020/9/1 | -4.2775 | -2.55333 | -68% |
| 2020/9/3 | -4.318 | -2.93 | -47% |
| 2020/9/5 | -4.765 | -2.70667 | -76% |
| 2020/9/9 | 4.91 | 8.696667 | -44% |
| 2020/9/10 | -4.364 | -2.88 | -52% |
| 2020/9/13 | 5.72 | 4.423333 | 29% |
| 2020/9/14 | -4.84 | -2.80667 | -72% |
| 2020/9/15 | -1.055 | -2.96 | 64% |
| 2020/9/16 | -0.42 | -1.4625 | 71% |
| 2020/9/17 | -4.6525 | -3.05333 | -52% |
| 2020/9/20 | -2.325 | -1.72 | -35% |
| 2020/9/21 | 5.805 | 2.843333 | 104% |
| 2020/9/22 | -3.975 | -3.708 | -7% |
| 2020/9/25 | 6.5325 | 3.616667 | 81% |
| 2020/9/26 | -2.42 | -3.34167 | 28% |
| 2020/9/27 | -4.025 | -3.524 | -14% |
| 2020/9/28 | -1.81 | -1.7325 | -4% |
| 2020/9/29 | -5.41 | -3.055 | -77% |
| 2020/9/30 | 5.892 | 4.2 | 40% |
| 2020/10/1 | -1.92 | -3.406 | 44% |
| 2020/10/12 | -2.94 | -3.258 | 10% |
| 2020/10/13 | -3.26667 | -2.988 | -9% |
| 2020/10/14 | -2.64667 | -1.395 | -90% |
| 2020/10/15 | -3.33 | -3.134 | -6% |
| 2020/10/28 | -0.89667 | -0.7525 | -19% |
| 2020/10/29 | 5.64 | -1.878 | 400% |
| 2020/10/30 | -0.86 | -2.774 | 69% |
| 2020/10/31 | -5.37667 | -3.055 | -76% |
| 2020/11/4 | -3.58 | -4.198 | 15% |
| 2020/11/5 | -4.625 | -2.70667 | -71% |
| 2020/11/17 | -5.41 | -2.4825 | -118% |
| 2020/11/18 | -4.78667 | -2.03 | -136% |
| 2020/11/19 | -3.12 | -3.35167 | 7% |
| 2020/11/23 | -3.265 | -5.02 | 35% |
| 平均值 | 0.893 | 0.750597 | 19% |

从表格中我们可以看到，上风处雨量站的平均增加雨量为0.750597mm，下风处雨量站的平均增加雨量为0.893mm，下风处的增长量比上风处的增长量大19%，也就是说下风处的催化效果比上风处强19%，这说明了催化效果与风向关系很大，也证明了我们一开始的假设是正确的。同时，不管上风处还是下风处，雨量增加量均是正值，证明了装置的有效性。

下面的几张图是从选取下风处催化效果更好的几个日期的数据绘制的。图中的散点代表了7个雨量站的具体地理位置，数值代表了该日降雨量减去该站的平均降雨量，三个黑色的五角星是六盘山三个单电极装置的位置，红色的三角形是综合气象站WG006的位置，由它的风向作为该试验区的主要参考风向，将WG005和WG007的风向作为辅助参考风向。因为有三个独立的气象站风向数据，所以在图中没有绘制风向玫瑰图。图中的带箭头的射线代表了风，箭头的方向是风向，线段长度代表风速的大小。



8月6日



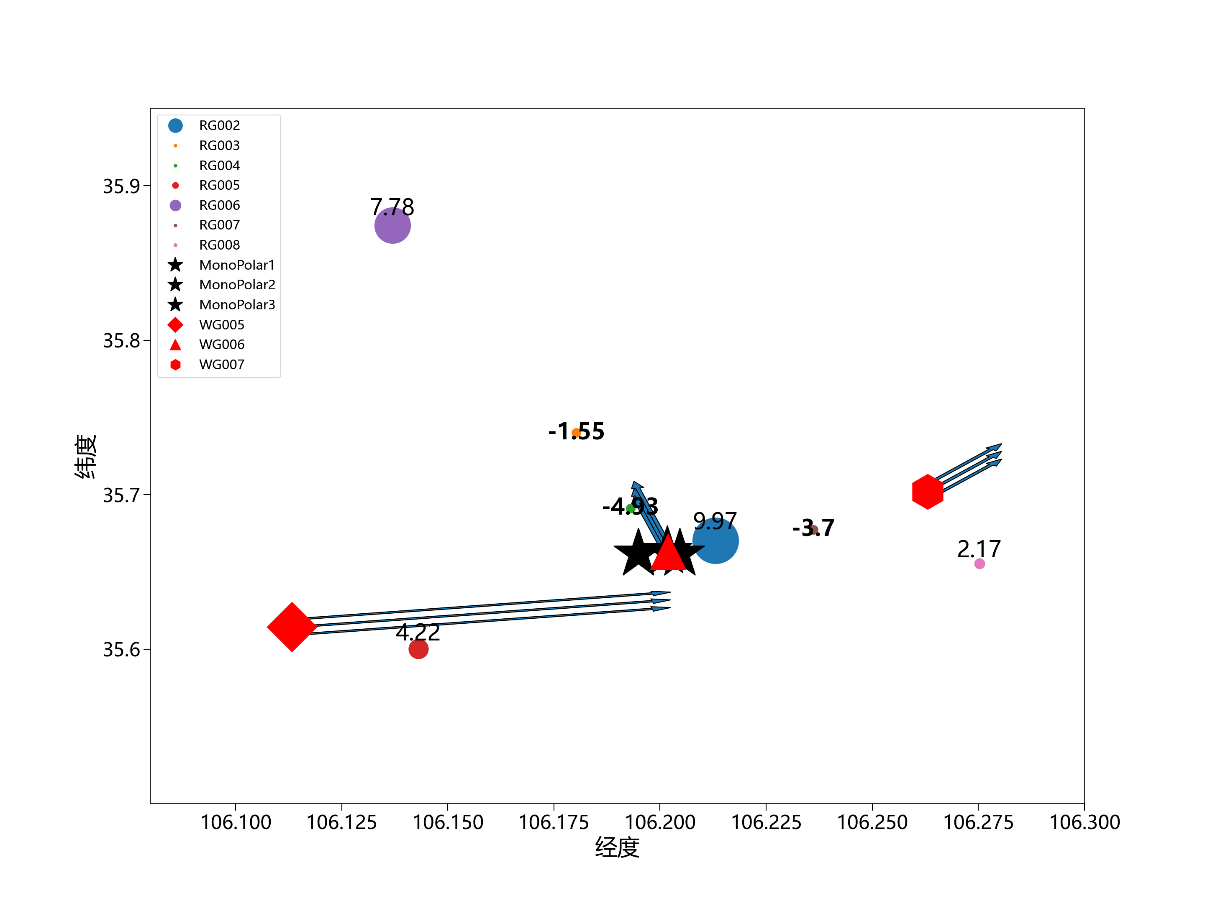
8月16日

为了使我们的结果更具有说服力，我们将研究周期从一天改为了一个降雨过程，我们把8个雨量站的降雨全为0定义为一个降雨过程的结束，把一个雨量站测得降雨量定义为一个降雨过程的开始。将降雨数据进行分析划分，总结出了以下几个降雨过程，如下表所示。

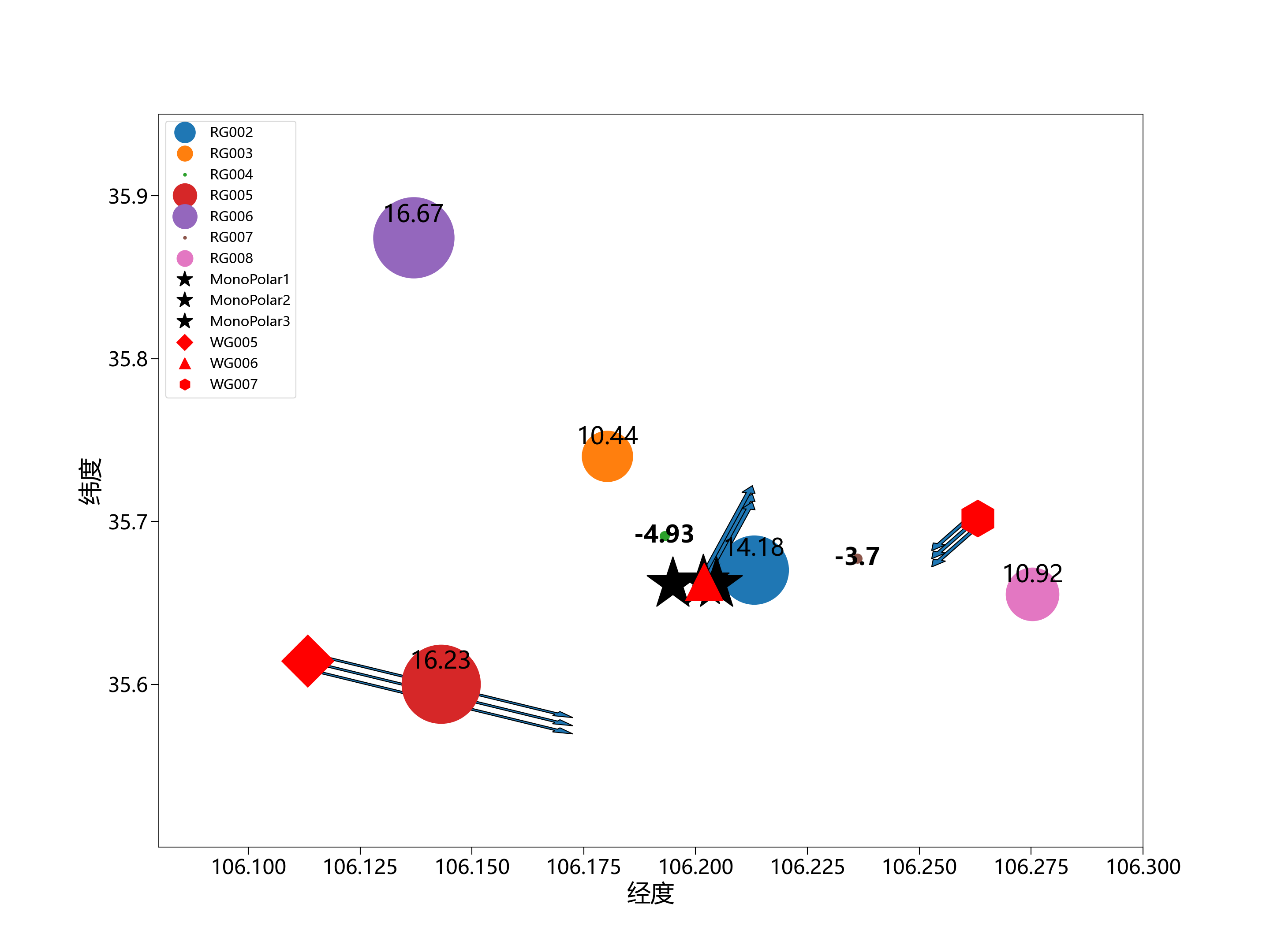
表

|  |  |  |  |
| --- | --- | --- | --- |
| 降雨过程 | 下风处站点雨量平均增量 | 上风处站点雨量平均增量 | 下风处增量相比上风处的增幅 |
| 0804-0808 | 2.8175 | 0.896667 | 214% |
| 0810-0817 | 10.518 | 3.66 | 187% |
| 0819-0905 | -0.15167 | 0.08 | -290% |
| 0909-0917 | -1.156 | -0.53 | -118% |
| 0920-1002 | -0.57333 | -1.445 | 60% |
| 1012-1015 | -2.7 | -3.1875 | 15% |
| 1027-1031 | -1.816 | -1.38 | -32% |
| 1104-1106 | -4.52 | -0.52 | -769% |
| 1117-1123 | -5.92 | -2.98667 | -98% |
| 平均值 | -0.38906 | -0.60139 | 35% |

从平均值来看，上风处的降雨平均增加量为-0.60139mm，下风处的降雨平均增加量为-0.38906mm，下风处增量相比上风处增量增加了35%，也就是说下风处比上风处的催化效果强了35%左右。这一数值和乌鞘岭地区的36%非常相近。与乌鞘岭地区不同的时，乌鞘岭地区的下风处和上风处的降雨增加量均为正值，但时六盘山地区都是负值，探究原因应该是前文提到的，由于求得的平均值相比自然水平引起的。以下是降雨过程的几张图片。



8月4日到8月8日



8月10日到8月17日

## 本章小结

本章按照统计统计分析的方法，采用了显著性检验，对“天水计划”设计的随机对比试验的实验结果进行了分析，探究了风向对六盘山和乌鞘岭实验区内部的雨量分布带来的影响，通过定量的分析说明了带电粒子发生装置的有效性。

# 总结与展望

基于对象的存储是为了克服当前基于块的存储存在的诸多难题，在存储接口

致 谢

论文完成之际，首先要感谢我的导师秦磊华教授。他站在学科发展的前沿，从论文的选题，研究工作逐步深入，到论文的撰写，都给我以细致的指导和建设性的意见，使我得以圆满而顺利地完成……。XXX严谨的治学态度、诲人不倦的师德和一丝不苟的工作作风将会给我留下不可磨灭的记忆……。

在…… 四年里，得到了许多老师和同学的大力帮助和支持，在此表示深深的谢意。感谢XXX等老师对我的关心和帮助。在课题研究和项目开发过程中，我与很多同学一起度过了一段段难忘的时光，值得怀念。我们大家共同创造的良好的学术氛围，将给我以永远而美好的回忆。

……

最后，我要深深地感谢我的父母，他们给予我无尽的关怀和无止境的爱，….

致谢属于论文的辅文部分。使用第一人称，采用散文体，对指导教师以及协助完成设计的有关人员表示谢意，并可简述自己通过本次毕业设计的体会，注意只写是查重最容易出问题的地方，请千万不要看别人写的，照搬。

参考文献

[1] 陈灿平. 我国水资源问题的成因分析和对策研究[J]. 西南民族大学学报(人文社会科学版),2014, 35(06): 141-144.

[2] Khain A, Arkhipov V, Pinsky M,et al. Rain enhancement and fog elimination by seeding with charged droplets. Part I: Theory and numerical simulations[J]. Journal of applied meteorology, 2004, 43(10):1513-1529.

[3] Rohwetter P, Kasparian J, Stelmaszczyk K,et al. Laser-induced water condensation in air[J]. Nature Photonics, 2010, 4(7):451-456.

[4] Ju J, Liu J, Wang C,et al. Laser-filamentation-induced condensation and snow formation in a cloud chamber[J]. OPTICS LETTERS, 2012, 37(7):1214-1216.

[5] 潘垣, 于克训. 带电粒子催化人工降雨雪新原理新技术及应用示范(天水计划)[J]. 中国环境管理,2017, 9(03): 115-116.

[6] 房彬, 肖辉, 王振会, 等. 聚类分析在人工增雨效果检验中的应用[J]. 南京气象学院学报,2005,(06): 739-745.

[7] Liang M, Hu X. Recurrent convolutional neural network for object recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015:3367-3375.

[8] Sundermeyer M, Schlüter R, Ney H. LSTM neural networks for language modeling[C]. Thirteenth annual conference of the international speech communication association, 2012.

[9] Lu H, Setiono R, Liu H. Effective data mining using neural networks[J]. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, 1996, 8(6):957-961.

[10]Yu L, Wang S, Lai K K. An integrated data preparation scheme for neural network data analysis[J]. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, 2005, 18(2):217-230.

[11]Alpaydin E. Introduction to machine learning[M]. MIT press, 2014.

[12] Gabriel R K, Baras M. The Israeli rainmaking experiment 1961-67: Final statistical tables and evaluation[M]. Hebrew University, 1970.

[13]Breed D, Rasmussen R, Weeks C,et al. Evaluating winter orographic cloud seeding: design of the Wyoming Weather Modification Pilot Project (WWMPP)[J]. Journal of Applied Meteorology and Climatology, 2014, 53(2):282-299.

[14]叶家东, 范蓓芬. 人工影响天气的统计数学方法[M]. 科学出版社, 1982.

[15]李书严, 李伟, 赵习方. 北京市人工增雨效果评估方法分析[J]. 气象科技,2006,(03): 296-300.

[16]黄彦彬, 姚展予, 游积平, 等. 基于“.NET”的华南区域人工增雨效果统计检验共享平台的设计与实现[J]. 气象,2011, 37(06): 727-734.

[17] Silverman B A. A Critical Assessment of Glaciogenic Seeding of Convective Clouds for Rainfall Enhancement[J]. BULLETIN OF THE AMERICAN METEOROLOGICAL SOCIETY, 2001, 82(5):903-924.

[18] Mielke P W, Berry K J, Medina J G. Climax I and II: Distortion Resistant Residual Analyses[J]. Journal of Applied Meteorology, 1982, 21(6):788-792.

[19] Levi Y, Rosenfeld D. Ice Nuclei, Rainwater Chemical Composition, and Static Cloud Seeding Effects in Israel[J]. Journal of Applied Meteorology, 1996, 35(9):1494-1501.

[20] Council N R. Critical issues in weather modification research[M]. National Academies Press, 2004.

[21]WMO Statement on Weather Modification. The Commission for Atmospheric Sciences Management Group, Second Session, Oslo, Norway, 24-26 September 2007[EB/OL].2007[https://www.wmo.int.

[22] 叶家东. 人工降水的试验设计和效果检验[J]. 气象,1979,(02): 26-29.

J Ye.1979,02): 26-29(in Chinese).

[23]王伟健, 姚展予, 贾烁, 等. 随机森林算法在人工增雨效果统计检验中的应用研究[J]. 气象与环境科学,2018, 41(02): 111-117.

附录：大学期间发表或提交的论文

1. A. Gember, R. Grandl, A. Anand, T. Benson, and A. Akella. Stratos: Virtual Middleboxes as First-class Entities[R]. Technical report, UW-Madison 2012.
2. Palkar S, Lan C, Han S, et al. E2: a framework for NFV applications[C]. Symposium on Operating Systems Principles. ACM, 2015:121-136.

此页以后的内容为装订材料！先仔细阅读！

撰写论文时此页以后的内容全部删掉！

装订毕业论文时应将任务书原件装订到最后



**本科生毕业设计任务书**

|  |  |
| --- | --- |
| 题 目 | 带电粒子人工降雨外场实验 |
|  |  |

（任务起止日期：2021年1月1日～2021年6月10日）

|  |  |
| --- | --- |
| 院 系 | 计算机科学与技术 |
| 专业班级 | 计科1703 |
| 姓 名 | 马恒奔 |
| 学 号 | U201714601 |
| 指导教师 | 秦磊华 |

教研室（系、所）负责人 2021年1月6日 审查

院（系）负责人 2021年1月9日 批准

|  |
| --- |
| 课题内容 |
| 实现一个简单的分布式网络系统，实现文件的分布式网络化存储。，实现文件的分布式网络化存储。，实现文件的分布式网络化存储。，实现文件的分布式网络化存储。，实现文件的分布式网络化存储。，实现文件的分布式网络化存储。，实现文件的分布式网络化存储。，实现文件的分布式网络化存储。 |
| 课题任务要求 |
| 研究学习文件系统的相关原理及实现研究分布式系统概念及应用研究网络文件系统的原理、架构及实现。研究学习文件系统的相关原理及实现研究分布式系统概念及应用研究网络文件系统的原理、架构及实现。研究学习文件系统的相关原理及实现研究分布式系统概念及应用研究网络文件系统的原理、架构及实现。 |
| 主要参考文献 |
| 1. 董晓明，谢长生．基于对象的进化存储系统研究．计算机科学，2005，32(11): 223~226 2. 庞丽萍编．操作系统原理（第二版）．武汉：华中理工大学出版社，1994.9．225~270 3. （美）Nils J. Nilsson著；郑扣根等译．人工智能（Artificial Intelligence）．北京：机械工业出版社，2000.9．177~194 4. （美）Tom Mitchell著；曾华军等译．机器学习（Machine Learning）．北京：机械工业出版社，2003.1．38~56 5. 谢长生，董晓明，万继光，谭志虎，刘瑞芳．磁盘阵列控制器的设计与原型实现．小型微型计算机系统，2006, 27(1): 173~176 6. （美）Nils J. Nilsson著；郑扣根等译．人工智能（Artificial Intelligence）．北京：机械工业出版社，2000.9．177~194 |
| 同组设计者 |
| 无 |

**成 绩 评 定**

**指导教师评定意见**

装订论文时应将成绩评定页放最后，导师评语放前面！

一、对毕业设计（论文）的学术评语（应具体、确切、实事求是）

|  |
| --- |
|  |

二、对毕业设计评分

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 评分项目  (分值) | 调研论证  (10分) | 外文翻译  (5分) | 设计(论文)撰写质量  (10分) | 学习态度  (10分) | 基本理论和基本技能  (50分) | 创 新  (15分) | 合 计  (100分) |
| 得分 | **8** | **4** | **8** | **8** | **40** | **12** | **80** |

指导教师签字：**秦磊华** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 年 月 日

**答辩小组评定意见**

答辩小组意见和综合评定页放最后，双面！

一、评语（根据学生答辩情况及其设计（论文）质量综合评价）

|  |
| --- |
|  |

二、评分

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 评分项目  (分值) | 答 辩 情 况 | | 论 文 质 量 | | 合 计  (100分) |
| 答辩情况  (15分) | 回答问题情况  (25分) | 规范要求与文字表达  (20分) | 学术水平  (40分) |
| 得分 |  |  |  |  |  |

答辩小组长签字：\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 年 月 日

**毕业答辩及成绩评定说明**

1. 毕业答辩
2. 答辩前，答辩小组应详细审阅每个学生的毕业设计（论文），为答辩做好准备。
3. 严肃认真组织答辩，开好答辩会。
4. 指导教师应参加所指导学生的答辩会，但评定其成绩时宜回避。
5. 答辩中要做好记录以供成绩评定时参考。
6. 成绩评定
7. 答辩前每个学生都要将自己的毕业设计（论文）在指定时间内交给指导教师，由指导教师审阅，写出评语并预评分。
8. 答辩工作结束后，答辩小组应举行专门会议进行讨论，在参考指导教师预评结果的基础上，结合学生毕业设计（论文）质量和学生答辩情况，综合评定每个学生的成绩。
9. 院（系）对专业答辩小组提出的优秀和不及格的毕业设计（论文），要组织院（系）级答辩，最终确定成绩，并向学生公布。
10. 各专业学生的最后成绩应符合正态分布规律。
11. 请用蓝、黑钢笔手写或五号宋体字编辑，签名须手写，A4纸双面打印。

**毕业设计（论文）成绩评定**

|  |
| --- |
| 班号：**计科1703** 学生姓名：**马恒奔**  综合成绩：\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_分（折合等级\_\_\_\_\_\_\_\_\_\_\_\_\_\_）  答辩小组长（签名）：\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 年 月 日 |