

# 短时强降水智能临近预报的一种新型众创机制

■ 魏晓琳 陈训来 江崧 李辉 毛夏

为促进人工智能(AI)技术在天气预报上的应用,深圳市气象局连续两年(2017—2018年)通过与阿里巴巴公司、香港天文台合作,共同组织了短时强降水智能临近预报为主题的“全球AI气象挑战赛”,征集到卷积神经网络、深度神经网络、决策树模型等基于AI技术的短时强降水临近预报方法,形成了一种新型的社会化众创机制。

DOI: 10.3969/j.issn.2095-1973.2019.03.030

短时强降水是指1 h雨量大于等于20 mm的降水事件,其导致的主要灾害是暴洪,是所有气象相关灾害中发生频次最高且导致伤亡最多的灾害,临近预报则是指短时间内(0~6 h,其中0~2 h为重点)发生明显变化的天气现象的预报,包括雷暴、强对流、降水和沙尘暴等。由于致灾程度高,雷暴和强对流是研究最多的、最具挑战性的临近预报研究对象,从实际业务的角度看,目前对大多数雷暴和强对流天气的高时空分辨率的可用预报和警报时效仍然不超过2 h。研究表明,基于高分辨率的天气预报模式WRF(Weather Research and Forecasting)以及逐时同化雷达、GPS水汽、风廓线等观测资料后,利用深圳稠密的自动气象站监测数据针对短时强降水的定时定量预报的严格TS评分显示,第一小时的降水临近预报的准确率在10%以下,第二小时起准确率均在3%以下。因此对于短时强降水的短临预报,特别是第二小时开始的临近预报是天气预报中的重点和难点问题。

传统的临近预报技术主要是基于多普勒天气雷达观测数据并结合其他资料对雷暴生成、发展和衰减过程的临近预报。新一代天气雷达由于可探测大气中的多种天气系统及其结构,由于其大功率、高灵敏度和全相参性能,能可靠地定量测量降水以及降水的生消、演变、范围、强弱以及降水分布,在临近天气预报和灾害性天气警报方面有着很高的应用价值。利用新一代天气雷达进行临近预报的方法有外推法、相似法、经验预报、降水的定量估测及预报、通过同化雷达资料利用中尺度数值模式开展数值临近预报等,虽然我国强对流天气短时临近预报业务已经取得了巨大进展,但在强对流天气的监测技术、快速同化更新的

数值预报和集合数值预报支撑技术、强对流天气预报检验技术、强对流天气机理研究和气候时空分布特征分析等方面仍然面临很多挑战。

近年来,人工智能技术发展较快,在天气预报领域的应用,无论在国内均持续受到关注,并被认为是较为适合处理天气预报这类复杂问题的有效方法。面对复杂的大气系统的演变,与传统的临近预报方法从探寻大气运动的本质规律出发不同,人工智能具有针对模糊、不确定性问题进行分析、联想、记忆学习和推断的能力,在气象领域的应用已具有较宽的覆盖面,从探测、数据处理、预报预测、到服务产品均有涉及,可通过大数据分析、信息识别、机器学习等技术解决各类难以通过数理模型直接给出确定性方案的复杂问题。AI技术除了通过机器学习解决预报模式中的不确定因素外,另一个明显优势是对各类复杂信息的识别与处理,可以改进模式初始信息的质量,也可以通过对信息的识别与学习开展短时临近预报。Shi Xingjian等通过机器深度学习,模拟雷达回波未来2 h的移动路径,较传统基于光流矢量计算方法预测雷达回波移动具有了新的优势。Yao and Li等采用卷积神经网络技术,结合了尺度不变的特征变换和泰勒冻结假设相结合的云团轨迹追踪技术,基于雷达资料对第二小时的单点强降水预报取得良好的效果。

为促进传统的天气预报方法与AI技术的融合应用,深圳市气象局自2017年起建立了一种新型的众创机制,通过在阿里巴巴的天池平台上发布标准雷达数据集,面向全球举办AI气象挑战赛,吸引从事AI、机器学习的人才到平台上来针对短时强降水智能临近预报问题发展原创的算法。在2017—2018年的实践经验

收稿日期:2018年11月30日;修回日期:2019年4月10日

中，取得了一定的成效。

## 1 全球AI气象挑战赛介绍

### 1.1 2017年赛事组织概况

#### 1.1.1 主办方、赛题和数据集

2017年，深圳市气象局以短时强降水智能临近预报技术为突破点，联合阿里云计算有限公司和CIKM2017（第26届信息与知识管理国际会议）作为主办方，首次组织了以短时强降水智能临近预报为主题的挑战赛，引导社会力量参与气象智能预报技术研究。在赛题的选择上，针对目前短时临近预报方法对短时强降水的定时、定点、定量预报能力有限，尤其是第二小时以后的预报准确率显著下降的特点，以提高单点短时强降水的临近预报精准度为目标，提出了“单点短时定量降水预测”的赛题。

竞赛使用的数据集共包含14000组数据，其中1万组为训练数据集，另外各2000组作为测试数据集，分别用于比赛第一阶段和第二阶段的竞赛。其中训练数

据集包含雷达数据和目标站点的第一至第二小时之间的降水量，而测试数据集仅包含雷达数据。选手需根据主办方提供的1万组训练数据集通过机器学习等方法进行训练，建立算法模型，然后对第一和第二阶段的测试数据给出未来第一至第二小时之间的降水量的预测。

数据集的编制基于广东省2014—2016年3年的历史多普勒雷达拼图数据（数据已做脱敏处理），以自动站小时雨量 $\geq 20$  mm为主要指标，通过时空抽稀和随机选取等方式确定了14000个目标自动站点及目标时刻，以目标站点为中心提取水平方向上 $100\text{ km} \times 100\text{ km}$ 的正方形、垂直方向上共4层（从 $0.5 \sim 3.5\text{ km}$ 每隔 $1\text{ km}$ ）、目标时刻过去 $1.5\text{ h}$ 内逐 $6\text{ min}$ 的雷达回波图，在训练数据集中每组数据还提供了目标站点在目标时刻未来 $1$ 和 $2\text{ h}$ 的降水量数据（图1），评测的方法为均方根误差，均方根误差越小排名越靠前。

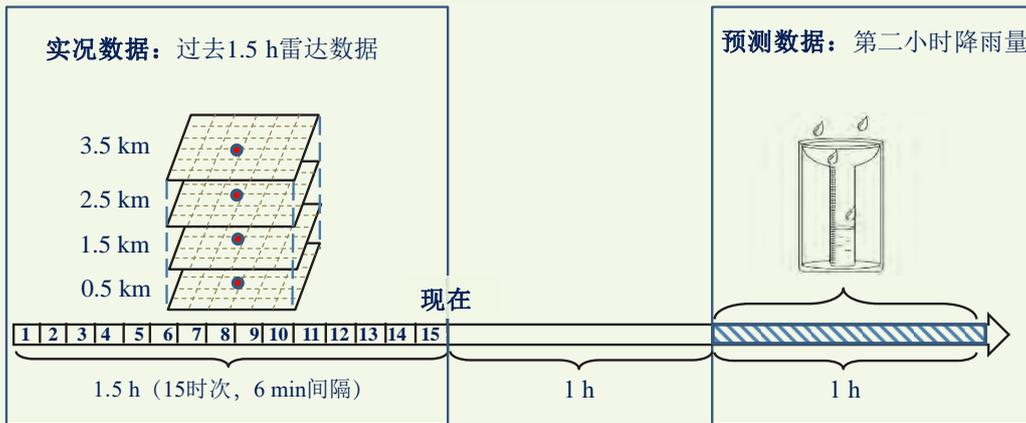


图1 2017年全球AI气象挑战赛赛题示意图

#### 1.1.2 比赛规则及结果

除主办方内部工作人员外，任何人均可参赛。竞赛分为发布训练数据与第一阶段测试数据集、第一阶段比赛、第二阶段比赛、入围选手提交报告和源代码及公布成绩4个阶段。在第一阶段、第二阶段比赛期间，参赛选手可以提交尽可能多的结果，以第二阶段的历史提交最优的一次结果作为最终的比赛成绩。

经CIKM2017大会组委会评审，深圳市气象局联合阿里云共同组织的挑战赛被确定为由国际计算机学会（ACM）于2017年11月在新加坡主办的第26届CIKM国际会议的四大赛题之一，并通过CIKM国际会议官网和阿里天池平台向国内外同时公布赛题和数据，比赛时间持续近半年，最后共有来自中国、新加坡、美国、英国以及港澳台等10余个国家和地区的

1395个团队、1650人参加，参赛团队包括清华大学、新加坡国立大学，香港科技大学、中国科学院等国内外著名高校院所以及高科技企业的专家和研究人员。参赛团队充分应用卷积神经网络、深度神经网络、决策树模型等深度学习算法，不断优化结果，总共提交了多达7917次的评估结果，前三名队伍及成绩见表1。

第一名清华大学团队采用卷积神经网络技术，结合了尺度不变的特征变换和泰勒冻结假设相结合的云团轨迹追踪技术，预报效果明显高于其他团队，比赛

表1 2017年全球AI气象挑战赛获奖队伍及成绩

排名	参赛队	所在组织	RMSE
1	Marmot	清华大学	10.99720
2	大时代啊	西安交通大学	12.33323
3	AI Rookies VN	越南VNG技术公司	12.94975

最终成绩较大赛提供的RMSE基准线14.69减小了3.7，这对气象部门开展强降水的智能预报技术研究提供了借鉴和参考。在CIKM国际会议期间，大会组委会安排气象智能预报挑战赛专题分会场报告，得到了大数据、人工智能等研究人员的广泛关注。

## 1.2 2018年赛事组织概况

### 1.2.1 主办方、赛题和数据集

2018年香港天文台加入赛事，由深圳市气象局、香港天文台、阿里云计算有限公司和IEEE ICDM2018国际会议4家联合主办，IEEE ICDM的全称是IEEE International Conference on Data Mining，为数据挖掘领域的三大顶级国际会议。IEEE ICDM提供了一个展示原创研究成果的国际论坛，研究范围涉及数据挖掘与人工智能领域的许多内容，包括算法、软件和系统以及应用程序等，吸引了大量数据挖掘与人工智能相关领域的研究人员和应用程序开发人员，涵盖统计学、机器学习、模式识别、数据库和数据仓库、数据可视化、高性能计算等各个学科和领域。此次ICDM2018主办的大数据竞赛有两项，其中之一为全球AI气象挑战赛，另一赛题为美国麻省理工学院主办

的多用途语音识别大数据竞赛。

经精心设计，2018年的赛题由2017年的“由面到点”（即由雷达回波图预测单点降水）难度加大为“由面到面”（即由雷达回波图预测雷达回波图）。数据集的量显著加大，融入香港的雷达数据，选取2010—2017年前汛期广东及香港的雷达数据，数据集采取灰度的国际通用标准数据格式，构成标准雷达数据集SRAD2018（Standardized Radar Dataset 2018），SRAD2018共提供32万组数据，其中30万组为训练数据集，另外各1万组为初赛和决赛的测试数据集，训练数据集的个例数量为2017年的30倍。训练数据集的每一组数据包括有61张雷达图像，空间分辨率为1 km，时间分辨率为6 min。而初赛和决赛的测试数据集的每组数据仅提供前3 h、间隔6 min的雷达数据，选手需要预测每组数据在后3 h内以30 min为间隔的雷达图像结果（图2）。在评测办法上，为了提高征集到的算法在实际业务中的应用价值，基于Heidke技巧评分（Heidke skill score, HSS）（其说明可参考<http://www.cawcr.gov.au/projects/verification/>），深圳市气象局和香港天文台联合对HSS方法进行改进，当预报时效越长和强度越大，则所占的评分比重越高。

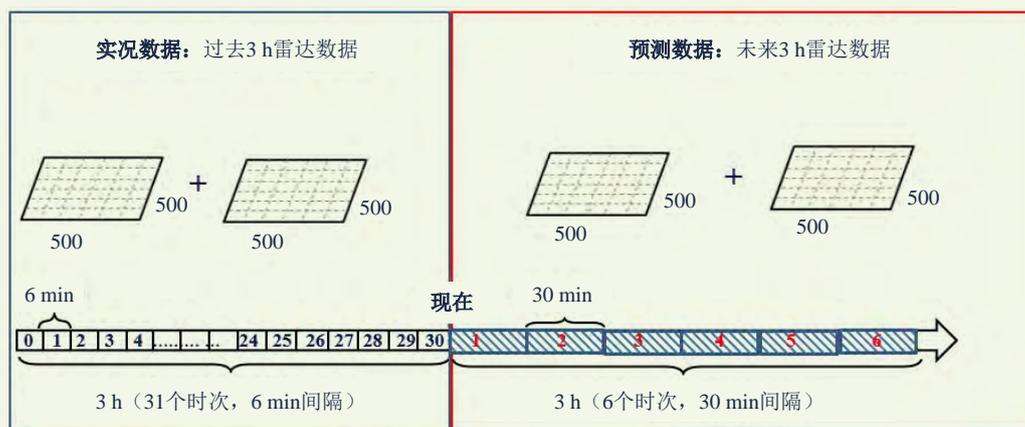


图2 2018年全球AI气象挑战赛赛题示意图

### 1.2.2 比赛规则及结果

与2017年相同，大赛面向全社会开放报名，除主办方内部工作人员外任何人士均可参赛。选手可1~4人组队参赛，每位选手只能加入一支队伍，否则会被取消参赛资格；除官方发布的SRAD2018数据集外，如参赛者还使用了其他数据信息将被视为作弊并取消参赛资格。与2017年相同，竞赛同样分为4个阶段，取第二阶段的历史提交最优的结果为代表队最终成绩。

2018年5月23日深港气象部门通过阿里云天池平

台再度发起全球AI气象挑战赛，前后历时6个月，影响力持续提升，共有来自全球45个国家和地区的1739支队伍、2012名选手参加，其中海外选手243人，覆盖45个国家和地区（含港、澳、台），共有114支队伍产出有效成绩，选手分布42%为学生，5%为教职工、科研人员，29%为公司职员，24%为其他。53%的选手具有硕士及以上学历，参赛团队来自北京大学、中国科学院、香港中文大学、加州大学洛杉矶分校、康奈尔大学、新加坡国立大学等国内外著名高校院所以及高科技企业。根据参赛团队提交的技术报告

和源程序，2018全球AI气象挑战赛组委会组织专家进行了审核，从程序复现运行结果一致性、技术原创性、文档完整性、技术合理和业务可行性等方面综合评审，来自香港中文大学（深圳）、中山大学等单位和个人获本年度全球AI气象挑战赛的前3名（表2），此次赛事征集到基于trajectory GRU、Conv2D LSTM以及传统光流法等针对赛题进行原创改进的智能临近预报算法，以本次竞赛的数据评分效果来看，基于AI方法的评分高于光流法的评分，显示了人工智能技术在解决临近预报难题中的巨大潜力。

表2 2018年全球AI气象挑战赛获奖队伍及成绩

排名	参赛队	所在组织	得分
1	CUHK(SZ)_group	香港中文大学（深圳）	50.85480036
2	两个芒果	中山大学	50.21705342
3	RandomScore	个人参赛	48.96835493

2018与2017年最大的不同点在于题目的难度增加、数据量显著加大，2017年是由多层雷达回波图预测单点的降水，而2018年是由雷达回波图预测雷达回波图，在实际的短临预报业务中，显然后者更为实用，征集的算法可以迅速投入业务使用，实践证明在竞赛中，人工智能的算法获得了冠军，表现优于传统的光流法，当然这仍需在实际业务中进行更多的检验。

### 1.3 对标国际一流，建立了一种短时强降水临近预报的新型众创机制

连续两年的全球AI气象挑战赛参考和借鉴了ImageNet的发展轨迹，ImageNet项目是世界上图像识别最大的数据库，提供了超过1400万个图像，自2010年以来，该项目每年举办一次软件比赛，即ImageNet大规模视觉识别挑战赛（ILSVRC），竞赛在给定的数据集上评估其算法，并在几项视觉识别任务中争夺更高的准确性，2010—2017年短短几年时间里，物体分类冠军的精确度从71.8%上升到97.3%，超越了人类物体分类水平，也有力证明了更大的数据会带来更好的决策（[https://www.sohu.com/a/160316515\\_680198](https://www.sohu.com/a/160316515_680198)）。

2017—2018年深圳市气象局联合香港天文台、阿里云计算有限公司通过全球AI气象挑战赛的成功举办已经初步构建了从发布数据集、提供算法运行环境和计算存储空间，吸引位于全世界各地的从事AI技术的专业人员到平台上针对智能临近预报的赛题提交原创算法，在给定的数据集上评估其算法的新的社会化众创机制。从众创的结果来看，征集到卷积神经网络、trajectory GRU、Conv2D LSTM等人工智能算法

用于解决气象上降水临近预报难题，并且在竞赛给定的个例中，总体表现上智能临近预报方法优于传统的光流法。

## 2 众创机制未来的发展方向展望

### 2.1 持续完善数据集及算法评测方法

ImageNet创始人李飞飞教授指出：构建优良数据集的工作是AI研究的核心，数据和算法一样至关重要（[https://www.sohu.com/a/160316515\\_680198](https://www.sohu.com/a/160316515_680198)）。因此数据集是人工智能天气预报中至为重要的一环，未来将致力于从以下几个方面进一步完善：一是将现有的30万组标准雷达数据集，扩展为百万量级以上的数据集，这需要将当前粤港澳的雷达数据拓展至更大的空间范围；二是采取更严格、更适用于AI算法研发的雷达数据控制技术，对雷达基数据带来的杂波、缺值等做更精心的处理，最大限度降低对机器学习带来的干扰；三是对智能临近预报的客观评价算法继续优化，使评分的结果与预报员的主观感受更匹配；四是继续完善标准雷达数据集下载、使用以及人工智能算法共享的使用条款。

### 2.2 创建AIMNet平台

针对众创机制未来的发展方向，我们的构想是在现有的众创机制的基础上，创建致力于解决气象业务难题的社会化众包平台AIMNet，英文AIMNet的含义即基于AI技术针对性的解决气象难题（Meteorology）的网络平台。该平台将主要包含如下几种功能：一是持续更新发布标准雷达数据集，并向全球征集经过质量控制的雷达数据，按照开源、脱密、共享、共研、共用、格式统一的原则汇集更多国家和地区的雷达基础数据，形成人工智能训练资源库；二是与超算中心合作，为AIMNet平台提供算法运行环境和计算存储资源，吸引从事人工智能研究的人员到平台上提交原创或改进的临近预报算法，形成短时强降水智能临近预报算法库；三是提供客观的、专业的评测方法，对进入平台上的短时强降水智能临近预报算法进行客观评分，形成产品排行榜及产品超市；四是将征集优选的算法在全国气象部门推广落地应用，并由气象业务部门提供应用反馈的意见。最终，通过AIMNet平台的搭建，形成短时强降水智能临近预报的大数据资源库、算法征集、客观优选、应用反馈的社会化众创机制的良性闭环，促进AI与天气预报的高度融合和螺旋式上升发展。

## 3 结论

本文总结了2017—2018年深圳市气象局与阿里巴

巴公司、香港天文台（2018年参与）联合主办的全球AI气象挑战赛，总结了挑战赛的经验和成果，并展望了未来发展的思路 and 理念，主要结论如下。

1) 在人工智能与气象这种交叉融合的新兴技术上，气象专业人员缺乏人工智能领域的专业知识，迫切需要通过技术的融合，促进智能临近预报的快速发展。全球AI气象挑战赛的成功举办显示了一种新型的短时强降水智能临近预报的众创机制已初见雏形和成效，有效激发了从事AI的专业人士研究气象问题，并征集到一些优秀的人工智能算法为气象所用，尤其是2018年征集到的算法可直接在业务中投入使用。

2) 未来智能临近预报的发展，数据和算法同样重要，深圳市气象局将通过AIMNet平台的创建和发展，着力推动标准雷达数据集的共享共用以倍数级增长，吸引更多AI技术人员参与短时强降水智能临近预报算法的发展，在统一的平台上，依托相同的大数据和评测方法，形成客观、权威的评分和排序，优选的算法将在全国气象部门内进行落地应用，形成问题的提出、解决、反馈的良性发展闭环。

3) 未来AIMNet平台的创建除了可以促进气象部门行业内的应用，同时也将搭建权威的、公开的平台，成效突出的算法可以迅速在平台上获得商业价值的认可和应用，促进商业气象经济的发展。

致谢：本项目由国家重点研发计划(2016YFC0203602)资助。

## 深入阅读

- Shi X, Chen Z, Wang H, et al, 2015. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: MIT Press.
- Shi X, Gao Z, Lausen L, et al, 2017. Deep learning for precipitation nowcasting: a benchmark and a new model. Proceedings of the 31th Conference on Neural Information Processing Systems. Long Beach, CA, USA.
- Wilson J W, Feng Y R, Chen M, et al, 2010. Nowcasting challenges during the Beijing Olympics: Success, failures, and implications for future nowcasting systems. Weather and Forecasting, 25: 1691-1714.
- Yao Y, Li Z, 2017. Short-term precipitation forecasting based on radar reflectivity images. Proceedings of International Conference on Information and Knowledge Management. Shenzhen Meteorological Bureau.
- 段鹤, 夏文梅, 苏晓力, 等, 2014. 短时强降水特征统计及临近预警. 气象, 40(10): 1194-1206.
- 许小峰, 2018. 从物理模型到智能分析——降低天气预报不确定性的新探索. 气象, 44(3): 341-350.
- 俞小鼎, 2013. 短时强降水临近预报的思路与方法. 暴雨灾害, 32(3): 202-208.
- 俞小鼎, 周小刚, 王秀明, 2012. 雷暴与强对流临近天气预报技术进展. 气象学报, 70(3): 311-337.
- 张蕾, 王明洁, 李辉, 2015. 短时强降水临近预报相对准确率的探讨. 广东气象, 37(2): 1-6.
- 张沛源, 杨洪平, 胡绍萍, 2008. 新一代天气雷达在临近预报和灾害性天气警报中的应用. 气象, 34(1): 3-11.
- 郑永光, 张小玲, 周庆亮, 等, 2010. 强对流天气短时临近预报业务技术进展与挑战. 气象, 36(7): 33-42.

(作者单位：魏晓琳、陈训来、李辉、毛夏，深圳市气象局；江崧，深圳市国家气候观象台)

(上接198页)

的雷暴过程，因此对于海上作业应采取必要的防雷措施，尤其是大型船舶的卸油（气）更应避免在雷暴活动时开展。

## 参考文献

- [1] 吕海勇, 李文飞, 吴坚. 1995更多数据和年广东省雷电灾害事故的统计分析. 广东气象, 2015, 37(6): 52-55.
- [2] 黄崇福, 张俊香, 陈志芬, 等. 自然灾害风险区划图的一个潜在发展方向. 自然灾害学报, 2004, 13(2): 9-15.
- [3] 程向阳, 谢五三, 王凯, 等. 雷电灾害风险区划方法研究及其在安徽省的应用. 气象科学, 2012, 32(1): 80-85.
- [4] 高隼, 蒙小亮, 劳小青. 基于聚类分析的海南岛雷电灾害易损度风险区划. 自然灾害学报, 2013, 22(1): 175-182.
- [5] 赵伟, 杨续超, 张斌. 浙江省雷电灾害风险分析及区划. 热带气象学报, 2014, 30(5): 996-1000.
- [6] 崔逊, 庄燕洵, 王洪生. 基于组合评价法的江苏省雷电灾害风险区划. 自然灾害学报, 2015, 24(6): 187-194.
- [7] 杨天琦, 林彬彬, 韩照全, 等. 基于GIS的区域雷电灾害风险管理方法研究. 西南大学学报, 2016, 38(2): 155-161.
- [8] 陈广昌, 崔逊, 田心如. 江苏省雷电灾害危险性量化评估及区划. 灾害学, 2017, 32(1): 32-35.
- [9] 吕海勇, 庄燕洵, 尹娜. 广东省雷电灾害易损性分析与风险区划. 广东气象, 2016, 38(2): 50-53.
- [10] 陈思蓉, 朱伟军, 周兵. 中国雷暴气候分布特征及变化趋势. 大气科学学报, 2009, 32(5): 703-710.
- [11] 郭虎, 熊亚军. 北京市雷电灾害易损性分析、评估及易损度区划. 应用气象学报, 2008, 19(1): 35-40.
- [12] 李彩莲, 赵西社, 赵东, 等. 陕西省雷电灾害易损性分析、评估及易损度区划. 灾害学, 2008, 23(4): 49-53.
- [13] 王惠, 邓勇, 尹丽云, 等. 云南省雷电灾害易损性分析及区划. 气象, 2007, 33(12): 83-87.
- [14] 陈晨辉, 张明生, 陈军, 等. 雷电与地形. 北京: 中国气象学会第七届中国国际防雷论坛论文集摘要, 2008.
- [15] 钟连宏, 常美生. 不同土壤电阻率对雷电放电过程的影响. 高压技术, 1997, 23(1): 64-66.