



文献 CSTR:
32001.14.11-6035.csd.2024.0070.zh



文献 DOI:
10.11922/11-6035.csd.2024.0070.zh
数据 DOI:
10.57760/sciencedb.j00001.01124

文献分类: 地球科学

收稿日期: 2024-04-07
开放同评: 2024-05-28
录用日期: 2024-11-25
发表日期: 2024-12-22

VLF 至 MF 频段雷电电场变化波形分类数据集

王宇^{1,2*}, 肖力郎³, 贺恒鑫³, 傅中⁴, 程晨⁴, 谷山强^{1,2}, 陈维江⁵

1. 国网电力科学研究院武汉南瑞有限责任公司, 武汉 430206
2. 国网电力科学研究院有限公司, 南京 211199
3. 华中科技大学电气与电子工程学院, 武汉 430032
4. 国网安徽省电力有限公司电力科学研究院, 合肥 230022
5. 国家电网有限公司, 北京 100031

摘要: 雷电作为一种复杂的大气放电现象, 其放电过程中会产生多种类型的电场变化波形。对这些波形进行准确地识别不仅是提升雷电监测系统性能的关键, 也直接关联到对雷电物理过程的深入理解。本研究基于 2022 至 2023 年间于安徽合肥市部署的 13 套雷电电场变化测量设备, 综合收集了合肥及周边地区雷暴活动期间的电场变化波形数据。数据时间跨度覆盖 3 月至 10 月合肥主要雷电活动季节, 空间跨度涉及整个城市及其周边地区, 确保了数据的代表性与全面性。在数据质量控制方面, 采用了严格的数据筛选与预处理流程, 确保了数据集的准确性和可靠性。采用先进的卷积神经网络 (CNN) 模型, 对波形数据进行了系统分类, 成功辨识出了四种雷电放电过程中的电场变化波形, 具体包括地闪回击过程 (RS, 15290 例)、初始击穿过程 (PB, 3946 例)、双极性窄脉冲事件 (NBE, 3919 例) 以及一般云闪过程 (IC, 4111 例)。分类模型的准确率超过 99%, 展现了高效的识别能力。本研究创建了一份大样本量、高质量分类的雷电电场变化波形分类数据集, 本数据集不仅能够支持开发更为先进的雷电波形分类算法, 促进雷电监测、核爆电磁脉冲监测等技术的进步, 也可为深入理解雷电物理过程提供重要的数据基础。

关键词: 雷电电场变化波形; 地闪; 云闪; 初始击穿过程; 双极性窄脉冲事件

数据库 (集) 基本信息简介

数据库 (集) 名称	VLF 至 MF 频段雷电电场变化波形分类数据集
数据通信作者	王宇 (wangyu16@sgepri.sgcc.com.cn)
数据作者	王宇, 肖力郎, 贺恒鑫, 傅中, 程晨, 谷山强, 陈维江
数据时间范围	2022–2023 年
地理区域	30°51'N–32°51'N, 116°16'E–118°16'E
数据量	2.02 GB
数据格式	*.csv 和 *.png
数据服务系统网址	https://doi.org/10.57760/sciencedb.j00001.01124
基金项目	国家电网有限公司总部科技项目 (5500-202120583A-0-5-SF)
数据库 (集) 组成	数据集共包括 4 个文件夹, 其中: 名为 RS 的文件夹是地闪回击波形数据, 共有 15290 例, 分为 *.csv 和 *.png, 一个 *.csv 对应一个 *.png;

* 论文通信作者
王宇:wangyu16@sgepri.sgcc.com.cn

数据库（集）组成	名为PB的文件夹是初始击穿过程波形数据，共有3946例；名为NB的文件夹是双极性窄脉冲事件波形数据，共有3919例；名为IC的文件夹是一般云闪波形数据，共有4111例。
----------	--

引言

雷电是一种发生在大气对流层的强烈气体放电现象，既具有极强破坏力，如引起森林火灾、威胁人身安全和电力设施安全等，同时也蕴含很高的科研价值。雷电放电过程产生的宽频电磁场信号为揭示雷电生成与发展机制，优化雷电监测、预测与减灾策略，研究大气科学、临近空间物理学等提供了珍贵数据。

雷电电场变化波形由雷电电场变化测量仪（又称快慢天线）传感测量而获得^[1-2]，此类波形可记录雷电放电全过程的电场变化信号，是雷电研究和雷电监测的重要数据。自 20 世纪 50 年代以来，随着测量采集技术不断进步，雷电电场变化波形数据的时间分辨和数据质量越来越高，在雷电物理研究和雷电监测领域发挥出日益重要作用^[3-4]。通过雷电电场变化波形数据分析，可以判断雷电是否击中地面物体，即区分地闪和云闪^[1-3]。受放电瞬时性和雷暴云遮挡等多重因素影响，对云闪放电过程的辨识和理解较为困难。借助高质量电场变化波形信号和 VHF 定位等观测技术，一些云闪放电过程逐渐被发现，如初始击穿过程^[5-8]、双极性窄脉冲事件（也称袖珍云闪）^[9-11]，这些进展加深了对云闪放电机制的理解。

传统的基于雷电信号时域波形特征判别雷电放电类型的方法存在一定局限性。近年来，机器学习技术被引入到雷电信号的识别中，证明了其在提高识别准确率和泛化能力方面的显著优势^[12-13]，有望推进雷电放电过程的物理解向更深层次发展。同时，雷电电磁信号与远区核爆电磁信号具有较强的相似性，是核爆电磁脉冲探测的主要干扰源^[14]。雷电电磁信号的准确辨识可促进核爆电磁信号的识别和探测。本研究收集和分类的雷电电场变化波形数据集是目前本领域规模最大、类别较全的数据集，不仅能够支持雷电波形分类算法的迭代优化，促进雷电监测、核爆电磁脉冲监测等技术的进步，也可为进一步理解雷电物理过程提供重要的数据基础。

1 数据采集和处理方法

1.1 数据采集方法

雷电电场变化测量仪的基本原理是通过金属板天线（平板天线）感应由云层或云与地面间的电荷分布引起的电场强度变化。传感器与电子电路相连，包括运算放大器等组件，用于信号的放大和滤波。当雷电形成时，电场强度发生变化，传感器上感应到的电荷也会变化，进而产生电压变化，通过电子电路进一步处理这些电压信号，可以获得关于电场变化的详细信息^[1,2,15]。雷电电场变化测量仪现场安装图如图 1 所示。

本研究研制的雷电电场变化测量仪技术指标如表 1 所示。由于 3 dB 带宽为 2 kHz-3.4 MHz，本数据包含 VLF 至 MF 频段信号。采样率为 10 Msps，时间分辨率为 0.1 μ s。测量仪采集的原始数据为 200 ms^[16]，本研究在分类处理时，提取了其中的有用信号，每段波形信号时长为 2 ms，采样点数为 20000，采样时间间隔为 0.1 μ s。本数据采集期为 2022 年 3 月至 2023 年 10 月，共部署测量仪 13 套，站点分布图如图 2 所示。



图 1 雷电电场变化测量仪实物图

Figure 1 On-site installation of lightning electric field change meter

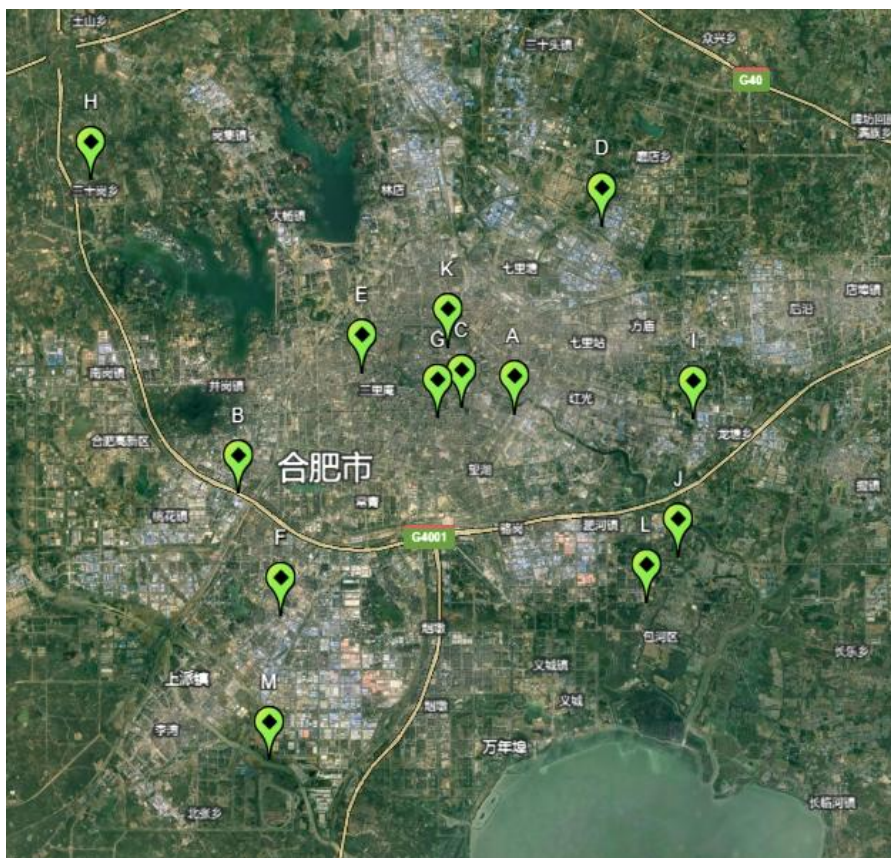


图 2 站点分布图

Figure 2 Site distribution map

表 1 雷电电场变化测量技术指标

Table 1 Technical specifications for the measurement of lightning electric field changes

指标	数值
带宽	2 kHz-3.4 MHz
时间常数	1 μ s
灵敏度	15 V/m
时间分辨率	0.1 μ s
垂直分辨率	16 bit
授时精度	50 ns
记录时长	200 ms

1.2 数据分类方法

本研究设计了适用于雷电电场变化波形的分类模型结构（见图 3），提出了一种基于多通道卷积神经网络的雷电电场变化信号分类方法。由于雷电波形特征时间尺度丰富，本研究在卷积层中以并行结构引入多尺寸卷积核，以有效提取雷电波形多尺度特征；通过 shortcut 连接实现特征重用，加速模型训练收敛。利用在合肥部署的 13 套雷电电场变化测量仪测量数据构建了 4 分类数据集，包括回击（RS）、初始击穿（PB）、双极性窄脉冲事件（NBE）以及云闪（IC）。模型分类准确率达到 99.4%，优于支持向量机、随机森林等机器学习算法^[16]。基于此分类模型，本研究分类筛选出地闪回击过程波形 15290 例、初始击穿过程波形 3946 例、双极性窄脉冲事件波形 3919 例，一般云闪过程波形 4111 例。需要说明的是，本研究目标在于多类型雷电放电波形信号准确识别，尽可能在保证模型识别准确率的情况下提升模型泛化性能，因此筛选出的 4 种类型雷电放电波形未区分距离、极性，且数量比例不代表各类型的真实占比。

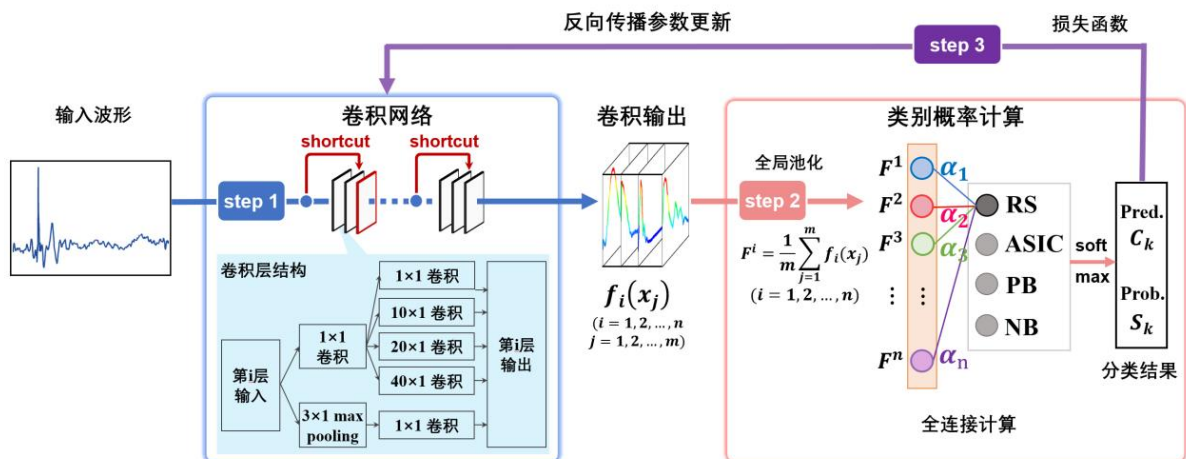


图 3 本研究设计的雷电电场变化波形分类模型结构图

Figure 3 The structure diagram of the lightning electric field change waveform classification model designed in this study

2 数据样本描述

本数据集共包括 4 类雷电波形数据。图 4 展示了回击（RS）、初始击穿（PB）、双极性窄脉冲事件（NBE）和云闪（IC）典型电场变化波形。图 4（a）为典型回击波形，呈快速波形前沿、相对慢速波形后沿和小幅反极性过冲。反极性过冲的有无及大小受传播距离影响较大。图 4（b）为典型初始击穿过程电场波形，表现为双极性脉冲簇，幅值先增大后减小。图 4（c）为典型双极性窄脉冲事件波形，从命名可以看出其电场波形是孤立的双极性窄脉冲，一般在主脉冲后跟随 2 个电离层反射脉冲。图 4（d）是一般云闪波形，表现为离散脉冲群，一般具有不规则性。本研究采用物理学符号定义，即雷暴云中负电荷向靠近测站的方向移动表现为电场负向增大^[2]。

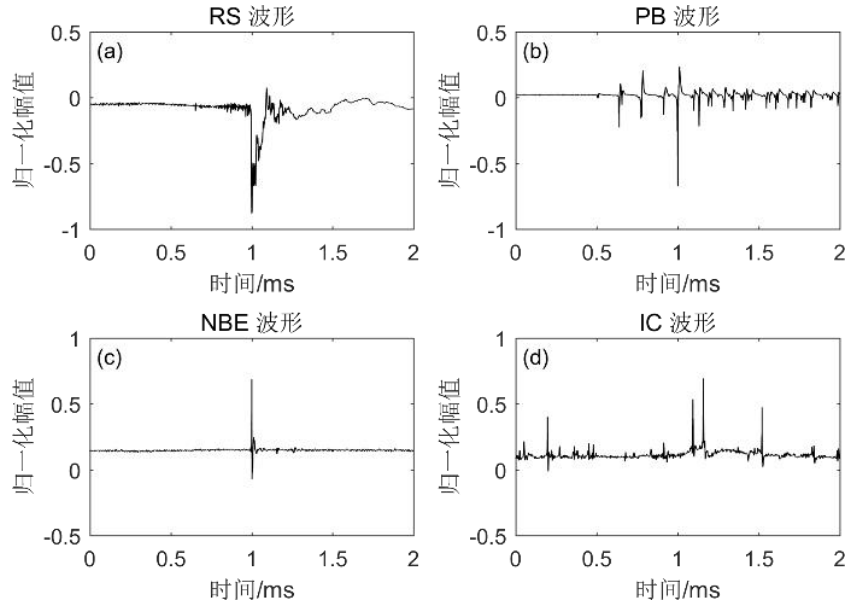


图 4 回击（RS）、初始击穿（PB）、双极性窄脉冲事件（NBE）和云闪（IC）典型电场变化波形

Figure 4 Typical electric change waveform of RS, PB, NBE, and IC

3 数据质量控制和评估

所用波形预处理包含分帧筛选标注以及滤波降噪两部分。由于原始波形持续 200 ms，存在大量背景噪声，因此需要提取其中雷电辐射脉冲部分并标注其类型。本研究采用滑动窗口法分割原始波形，窗口大小为 2 ms，滑动步长为 0.5 ms，以分割后每一片段为一帧。将每一帧波形数据归一化至 ± 1 区间后计算帧能量，当某帧的帧能量小于设定阈值时将该帧被标记为噪声帧并舍弃，帧能量可表示为：

$$\begin{cases} e_{\min} < e \\ e = \sum_{i=1}^h w_i^2 \end{cases} \quad (1)$$

式中： $e_{\min}$ 为帧能量设定阈值； e 为帧能量； h 为帧长度； w_i 为帧内第 i 个数据点。采用经验模态分解（Empirical Mode Decomposition, EMD）去除主频小于 200 Hz 的低频分量，采用小波去噪方法使用 sym5 小波基分解 5 层，通过软阈值方法处理分解所得小波系数，滤除帧内高频噪声^[16]。

4 数据价值

雷电电场变化波形是雷电监测和科研所需重要数据集。本数据采集所用硬件装备具有频带宽、垂直分辨率和时间分辨率高的特点；数据分类采用的自行开发的卷积神经网络分类模型分类准确度高。本数据集在雷电信号分类算法开发、雷电监测和核爆电磁脉冲监测技术提升以及雷电物理、空间物理研究方面具有重要价值。

5 数据使用方法和建议

VLF 至 MF 频段雷电电场变化波形分类数据集共包括 4 类，其中标记为 RS 的是地闪回击波形；标记为 PB 的是初始击穿过程波形数据；标记为 NBE 的是双极性窄脉冲事件波形数据；标记为 IC 的是一般云闪波形数据。每类数据有 CSV 格式和 PNG 格式两种，PNG 格式数据是 CSV 格式数据的图片形式。可使用 Matlab、Python、R 语言等编程语言读取、查看、分析，也可以使用 Pytorch、Tensorflow 等平台建立雷电信号分类模型。

致 谢

感谢合肥工业大学潘平、魏常勇、吕增威等老师和学生在探测站测试、安装和数据打标过程中的贡献。

数据作者分工职责

王宇（1987—），男，湖北省宜昌市人，博士研究生，高级工程师，研究方向为雷电物理及电力系统防雷技术。主要承担工作：研究方案设计、硬件设计、算法设计、波形数据标注。

肖力郎（2000—），男，四川省内江市人，在读博士研究生，研究方向为电力系统防雷技术。主要承担工作：算法设计，波形数据标注，硬件调试。

贺恒鑫（1982—），男，重庆市人，博士研究生，副教授，研究方向为电力系统过电压与绝缘配合。主要承担工作：算法设计，波形数据标注。

傅中（1977—），男，安徽省无为市人，博士研究生，正高级工程师，研究方向为输电运维技术。主要承担工作：硬件安装部署。

程晨（1994—），男，湖北省黄冈市人，博士研究生，研究方向为电力系统过电压与绝缘配合。主要承担工作：硬件安装部署。

谷山强（1980—），男，河南省周口市人，博士研究生，教授级高工，研究方向为高电压与电力系统防雷技术。主要承担工作：研究方案指导。

陈维江（1958—），男，山东省淄博市人，硕士研究生，教授级高工，研究方向为电力系统电磁瞬态分析方法与防护技术。主要承担工作：项目负责人、研究方案指导。

参考文献

[1] KITAGAWA N, BROOK M. A comparison of intracloud and cloud-to-ground lightning discharges[J].

- Journal of Geophysical Research, 1960, 65(4): 1189–1201.
- [2] 郄秀书, 张其林, 袁铁, 等. 雷电物理学[M]. 北京: 科学出版社, 2013. [QIE X S, ZHANG Q L, YUAN T, et al. Lightning Physics [M]. Beijing: Science Press, 2013.]
- [3] QIE X, ZHANG T, CHEN C, et al. The lower positive charge center and its effect on lightning discharges on the Tibetan Plateau[J/OL]. Geophysical Research Letters, 2005, 32(5). <https://onlinelibrary.wiley.com/doi/abs/10.1029/2004GL022162>. DOI:10.1029/2004GL022162.
- [4] 周璧华, 郭建明, 邱实, 等. 闪电电场变化测量仪研究[J]. 地球物理学报, 2012, 55(4): 1114–1120. DOI: 10.6038/j.issn.0001-5733.2012.04.006. [ZHOU B H, GUO J M, QIU S, et al. Research of the lightning electric field change meter[J]. Chinese Journal of Geophysics, 2012, 55(4): 1114–1120. DOI: 10.6038/j.issn.0001-5733.2012.04.006.]
- [5] WANG Y, QIE X S, WANG D F, et al. Beijing Lightning Network (BLNET) and the observation on preliminary breakdown processes[J]. Atmospheric Research, 2016, 171: 121–132. DOI: 10.1016/j.atmosres.2015.12.012.
- [6] CLARENCE N D, MALAN D J. Preliminary discharge processes in lightning flashes to ground[J]. Quarterly Journal of the Royal Meteorological Society, 1957, 83(356): 161–172. DOI: 10.1002/qj.49708335603.
- [7] NAG A, RAKOV V A. Pulse trains that are characteristic of preliminary breakdown in cloud-to-ground lightning but are not followed by return stroke pulses[J]. Journal of Geophysical Research: Atmospheres, 2008, 113(D1): e2007jd008489. DOI: 10.1029/2007jd008489.
- [8] 王宇, 郄秀书, 王东方, 等. 正地闪和负地闪预击穿脉冲序列的统计分析对比[J]. 大气科学, 2014, 38(1): 21–31. DOI: 10.3878/j.issn.1006-9895.2013.12167. [WANG Y, QIE X S, WANG D F, et al. Comparisons of preliminary breakdown pulse trains in positive and negative cloud-to-ground lightning flashes[J]. Chinese Journal of Atmospheric Sciences, 2014, 38(1): 21–31. DOI: 10.3878/j.issn.1006-9895.2013.12167.]
- [9] SMITH D A, SHAO X M, HOLDEN D N, et al. A distinct class of isolated intracloud lightning discharges and their associated radio emissions[J]. Journal of Geophysical Research: Atmospheres, 1999, 104(D4): 4189–4212. DOI: 10.1029/1998jd200045.
- [10] NAG A, RAKOV V A, TSALIKIS D, et al. On phenomenology of compact intracloud lightning discharges[J/OL]. Journal of Geophysical Research: Atmospheres, 2010, 115(D14)[2024-03-27]. <https://cir.nii.ac.jp/crid/1360855569471474688>. DOI:10.1029/2009jd012957.
- [11] 吴亭, 董万胜, 李良福, 等. 基于电离层反射的袖珍云闪(CID)三维定位研究[J]. 地球物理学报, 2012, 55(4): 1095–1103. DOI: 10.6038/j.issn.0001-5733.2012.04.004. [WU T, DONG W S, LI L F, et al. 3D location of compact intracloud discharge based on its ionospheric reflection pair[J]. Chinese Journal of Geophysics, 2012, 55(4): 1095–1103. DOI: 10.6038/j.issn.0001-5733.2012.04.004.]
- [12] XIAO L, CHEN W, WANG Y, et al. Toward an Interpretable CNN Model for the Classification of Lightning-Produced VLF/LF Signals[J/OL]. Journal of Geophysical Research: Atmospheres, 2023, 128(22): e2023JD039517. DOI:10.1029/2023JD039517.
- [13] ZHU Y, RAKOV V A, TRAN M D, et al. A study of National Lightning Detection Network responses to natural lightning based on ground truth data acquired at LOG with emphasis on cloud discharge

- activity[J]. Journal of Geophysical Research: Atmospheres, 2016, 121(24). DOI: 10.1002/2016jd025574.
- [14] 李鹏, 宋立军, 韩超, 等. 基于 AR 模型与神经网络的核爆与闪电电磁脉冲信号识别[J]. 强激光与粒子束, 2010, 22(12): 3052–3056. DOI: 10.3788/HPLPB20102212.3052. [LI P, SONG L J, HAN C, et al. Recognition of NEMP and LEMP signals based on auto-regression model and artificial neural network[J]. High Power Laser and Particle Beams, 2010, 22(12): 3052–3056. DOI: 10.3788/HPLPB20102212.3052.]
- [15] WANG Y, GU S, FANG Y, et al. Compact electric field change meter and its application in lightning detection and fault analysis for power grids[C/OL]//2020 IEEE International Conference on High Voltage Engineering and Application (ICHVE). Beijing, China: IEEE, 2020: 1-4[2024-04-03]. <https://ieeexplore.ieee.org/document/9279853/>. DOI:10.1109/ICHVE49031.2020.9279853.
- [16] 肖力郎, 陈维江, 王宇, 等. 基于多通道卷积神经网络的 VLF/LF 雷电辐射电场波形分类方法[J]. 高电压技术, 2024-02-23. DOI:10.13336/j.1003-6520.hve.20231990. [XIAO L L, CHEN W J, WANG Y, et al. A VLF/LF Lightning Radiation Electric Field Waveform Classification Method Based on a Multi-channel Convolutional Neural Network [J]. High Voltage Engineering, 2024-02-23. DOI:10.13336/j.1003-6520.hve.20231990.

论文引用格式

王宇, 肖力郎, 贺恒鑫, 等. VLF 至 MF 频段雷电电场变化波形分类数据集[J/OL]. 中国科学数据, 2024, 9(4). (2024-12-22). DOI: 10.11922/11-6035.csd.2024.0070.zh.

数据引用格式

WANG Yu. VLF 至 MF 频段雷电电场变化波形分类数据集[DS/OL]. V1. Science Data Bank, 2024. (2024-10-24). DOI: 10.57760/sciencedb.j00001.01124.

A dataset of lightning electric field change waveforms ranging from VLF to MF frequency band

WANG Yu^{1,2*}, XIAO Lilang³, HE Hengxin³, FU Zhong⁴, CHENG Chen⁴,
GU Shanqiang^{1,2}, CHEN Weijiang⁵

1. State Grid Electric Power Research Institute Wuhan NARI Limited Liability Company, Wuhan 430206, P.R. China
2. State Grid Electric Power Research Institute Co., Ltd., Nanjing 211199, P.R. China
3. School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan 430032, P.R. China
4. State Grid Anhui Electric Power Co., Ltd., Electric Power Research Institute, Hefei 230022, P.R. China
5. State Grid Corporation of China, Beijing 100031, P.R. China

*Email: wangyu16@sgepri.sgcc.com.cn

Abstract: Lightning, a complex atmospheric discharge phenomenon, generates various electric field change waveforms during its discharge process. The accurate identification of waveforms is crucial for enhancing the performance of lightning monitoring systems and having a deeper understanding of the physical processes behind lightning. Based on 13 sets of lightning electric field change measuring equipment deployed in Hefei, Anhui province, from 2022 to 2023, we collected electric field change waveform data during thunderstorm activities in Hefei and its surrounding areas for the study. The data span from March to October, covering the main lightning activity season in Hefei and extending to the entire city and its surrounding areas, ensuring the representativeness and comprehensiveness of the data. To ensure the accuracy and reliability of the dataset, strict data screening and preprocessing procedures were implemented. An advanced convolutional neural network (CNN) model was used to systematically classify the waveform data, successfully identifying four types of electric field change waveforms during the lightning discharge process, including Return Stroke (RS, 15290 cases), Preliminary Breakdown (PB, 3946 cases), Narrow Bipolar Events (NBE, 3919 cases), and Intra-cloud Discharge (IC, 4111 cases). The classification model achieved an accuracy exceeding 99%, demonstrating its efficient recognition capabilities. This paper has established a large-scale, high-quality classified dataset of lightning electric field change waveforms. This dataset can support the development of more advanced lightning waveform classification algorithms and promotes the advancement of technologies in lightning monitoring and nuclear explosion electromagnetic pulse monitoring. Moreover, it is expected to provide an important data foundation for a deeper understanding of the physical processes behind lightning.

Keywords: lightning electric field change waveforms; Cloud-to-ground lightning flash; Intracloud flash; Preliminary breakdown process; Narrow bipolar events

Dataset Profile

Title	A dataset of lightning electric field change waveforms ranging from VLF to MF frequency band
Data corresponding author	WANG Yu (wangyu16@sgepri.sgcc.com.cn)
Data authors	WANG Yu, XIAO Lilang, HE Hengxin, FU Zhong, CHENG Chen, GU Shanqiang, CHEN Weijiang
Time range	2022–2023
Geographical scope	30°51'N–32°51'N, 116°16'E–118°16'E
Data volume	2.02 GB
Data format	*.csv and *.png
Data service system	< https://doi.org/10.57760/sciencedb.j00001.01124 >
Source of funding	State Grid Corporation of China Science and Technology Project (5500-202120583A-0-5-SF)
Dataset composition	The dataset includes a total of 4 folders, where: the folder labeled as RS contains cloud-to-ground lightning flash waveform data, with a total of 15,290 entries, including

	*.csv and *.png files, each *.csv corresponding to a *.png; the folder labeled as PB contains preliminary breakdown process waveform data, with a total of 3,946 entries; the folder labeled as NB contains narrow bipolar pulse event waveform data, with a total of 3,919 entries; the folder labeled as IC contains general intracloud flash waveform data, with a total of 4,111 entries.
--	--