

# LightNet: A Dual Spatiotemporal Encoder Network Model for Lightning Prediction

Yangli-ao Geng

Qingyong Li

Tianyang Lin

Lei Jiang

Beijing Key Lab of Traffic Data  
Analysis and Mining, Beijing Jiaotong  
University  
Beijing, China

{gengyla,liqy,tyanglin,LeiJiang}@bjtu.edu.cn  
{xult,zhengdong,yaowen,wlyu}@cma.gov.cn

Liangtao Xu

Dong Zheng

Wen Yao

Weitao Lyu

State Key Laboratory of Severe  
Weather, Chinese Academy of  
Meteorological Sciences  
Beijing, China

Yijun Zhang

Department of Atmospheric and  
Oceanic Sciences, Fudan University  
Shanghai, China  
zhangyijun@fudan.edu.cn

## ABSTRACT

Lightning as a natural phenomenon poses serious threats to human life, aviation and electrical infrastructures. Lightning prediction plays a vital role in lightning disaster reduction. Existing prediction methods, usually based on numerical weather models, rely on lightning parameterization schemes for forecasting. These methods, however, have two drawbacks. Firstly, simulations of the numerical weather models usually have deviations in space and time domains, which introduces irreparable biases to subsequent parameterization processes. Secondly, the lightning parameterization schemes are designed manually by experts in meteorology, which means these schemes can hardly benefit from abundant historical data. In this work, we propose a data-driven model based on neural networks, referred to as LightNet, for lightning prediction. Unlike the conventional prediction methods which are fully based on numerical weather models, LightNet introduces recent lightning observations in an attempt to calibrate the simulations and assist the prediction. LightNet first extracts spatiotemporal features of the simulations and observations via dual encoders. These features are then combined by a fusion module. Finally, the fused features are fed into a spatiotemporal decoder to make forecasts. We conduct experimental evaluations on a real-world North China lightning dataset, which shows that LightNet achieves a threefold improvement in equitable threat score for six-hour prediction compared with three established forecast methods.

## CCS CONCEPTS

- Applied computing → Earth and atmospheric sciences;
- Computing methodologies → Neural networks; Spatial and physical reasoning;
- Information systems → Spatial temporal systems.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD '19, August 4–8, 2019, Anchorage, AK, USA

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6201-6/19/08...\$15.00

<https://doi.org/10.1145/3292500.3330717>

## KEYWORDS

Spatiotemporal data mining; Lightning Prediction; Deep learning; Time series prediction; Convolutional Neural Network

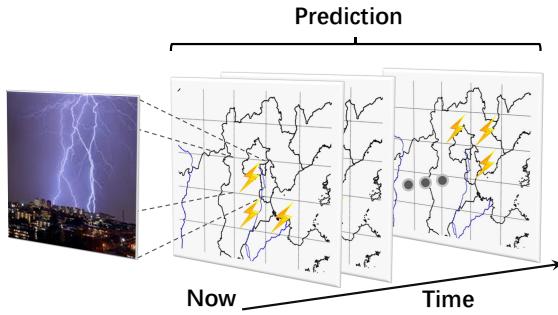
### ACM Reference Format:

Yangli-ao Geng, Qingyong Li, Tianyang Lin, Lei Jiang, Liangtao Xu, Dong Zheng, Wen Yao, Weitao Lyu, and Yijun Zhang. 2019. LightNet: A Dual Spatiotemporal Encoder Network Model for Lightning Prediction. In *The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '19)*, August 4–8, 2019, Anchorage, AK, USA. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3292500.3330717>

## 1 INTRODUCTION

Lightning is a typical feature of severe weather that is a significant source of concern for public safety and a wide range of weather sensitive operations [24]. According to an investigation conducted by Curran et al. [5], lightning is near the top of the list of all types of weather-related deaths. On the other hand, lightning also poses a serious threat to the physical environment, such as wildfires, power failures, air traffic disruption, and damage to telecommunication infrastructure [17].

The undoubtedly importance of lightning has driven significant interest in the prediction of this natural hazard [7]. Preliminary studies focus on understanding the electrification mechanism of lightning [1, 12, 18]. These researches provide basic theory and routes for the prediction of lightning. On the other hand, the development of computing technology gave rise to a modern prediction tool, numerical weather prediction (NWP) system. Given initial conditions, an NWP system can simulate atmospheric evolution for the future. Though this tool can provide simulated atmospheric parameters such as ice mixing ratio and snow mixing ratio [8], it does not directly support lightning prediction. Therefore, researchers proposed lightning parameterization schemes to convert the simulated parameters into the probability of lightning occurrence [14, 16, 24]. In recent years, with the success of deep neural networks (DNNs), some researches have applied DNNs to precipitation [21] and radar echo [32] nowcast. These methods can nowcast weather distribution for a short period (e.g. three hours) based on recent weather observations, but their performance may drop for longer prediction.



**Figure 1:** A sketch map of the spatiotemporal prediction for lightning.

Despite the extensive research into lightning, the spatiotemporal prediction for it (see Figure 1) still poses a challenge. The difficulties come from three aspects. Firstly, given the inconstancy of thunderstorms, the performance of nowcasting methods based on recent weather observations degenerate for long-period prediction. Secondly, the simulation of the NWP systems often shows deviations in the time and space domain, which introduce irreparable biases to prediction methods based on the simulation. For example, the simulated parameters of the southeast for 7:00 may be the real situation of the south for 8:00. Thirdly, the errors in the simulation demonstrate different patterns in different spatiotemporal locations, which make it hard to calibrate the simulation. For instance, the simulated weather parameters in high latitude area may have a positive bias, but the bias may be negative in low latitude area. Furthermore, the sign of bias may change with time. Though recent lightning observation data is capable of calibrating the simulation, it is hard to bridge their relation due to complex spatiotemporal correspondences between them.

To meet the above-mentioned challenges, we propose LightNet, a data-driven lightning prediction model based on deep neural networks. Unlike the existing methods, which exploit either simulation data (commonly used in conventional prediction methods) or recent weather observation data (often used in extrapolation methods), LightNet makes predictions based on both of them. LightNet extracts spatiotemporal features of the simulation and the observation via dual encoders. Then, these features are combined by a fusion module. Finally, the fused features are fed into a spatiotemporal decoder to make forecasts. LightNet has demonstrated superior performance for lightning prediction without complex feature selection process and tuning trick. To summarize, our contributions include four aspects:

- (1) We propose a data-driven LightNet model for lightning prediction, which outperforms the established prediction methods that fail to mine knowledge from historical data.
- (2) We introduce dual spatiotemporal encoders to extract simulation based and observation based features, which offers new insights into the general problem of weather forecast.
- (3) We combine simulation based and observation based features by a fusion module, which proves effective in withstanding deviations in the simulation.

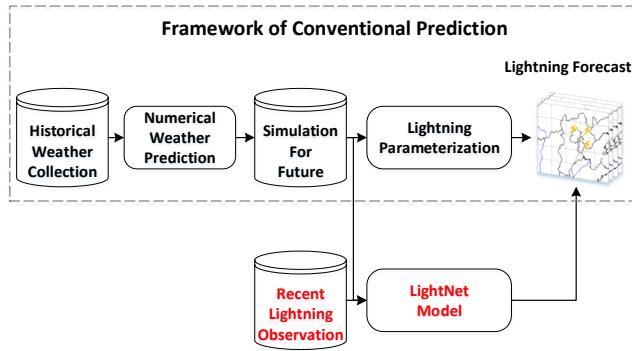
- (4) In our experiment, we used a real-world lightning of North China dataset to evaluate LightNet, which achieved threefold improvements in equitable threat score compared with three established prediction methods.

The rest of this paper is organized as follows. In Section 2, we discuss the related works. Section 3 introduces the preparatory work. Section 4 details the proposed LightNet framework. Experimental results are demonstrated in Section 5. Finally, we conclude this paper and suggest future work in Section 6.

## 2 RELATED WORKS

**Lightning Prediction.** Generally speaking, modern lightning prediction methods fall into three categories [31]: numerical diagnosis prediction based on synoptic background filed statistical relations; the numerical weather model coupled with the explicit electrification and lightning parameterization schemes; and flash rate parameterization. Here, we focus on prediction methods based on the flash rate parameterization, which are widely used in the real world and served as our baseline. As a pioneering work, Price et al. [16] bridged the relation between lightning frequency and maximum vertical velocity (or convective cloud top height), and propose the well-known PR92 lightning parameterization scheme. Michalon et al. [14] proposed a new parameterization by combining PR92 and cloud droplet concentration. Deierling et al. [6] used ground-based dual-polarimetric radar and total lightning data to investigate the relationship between total lightning activity and precipitation ice mass on a storm scale. Jr et al. [24] proposed two approaches based on the upward fluxes of precipitating ice hydrometeors in the mixed-phase region and the vertically integrated amounts of ice hydrometeors, respectively. They found that the first method is more capable of capturing time-dependent features of lightning while the second one owns superiority in forecasting lighting areas. Though lightning prediction has been researched extensively, its accuracy leaves much to be desired.

**Weather Pattern Mining Based on Deep Learning.** With its recent success in many walks of life, deep learning has boosted great interest in weather pattern mining. Some works focusing on weather pattern mining are relevant to our model. As pioneers in this field, Klein et al. [11] proposed a new neural network layer called "dynamic convolutional layer" and applied it to short-term forecasting of location and intensity of rain and snow. To solve the problem of spatiotemporal dependency, Shi et al. [21] developed the conventional LSTM and propose convolutional LSTM (ConvLSTM). Their experiments demonstrate that ConvLSTM outperforms the previous works on the grid-wise precipitation nowcasting task. Furthermore, they also proposed the trajectory GRU model [22] that improves ConvLSTM by actively learning the recurrent connection structure. Liu et al. [13] applied deep convolutional neural networks to detect extreme weather in large historical climate datasets. Wang et al. [32] presented a predictive recurrent neural network (PredRNN) in the light of the idea that spatiotemporal predictive learning should memorize both spatial appearances and temporal variations in a unified memory pool. Their model has achieved state-of-the-art performance on a radar echo forecasting task. Our work focuses on lightning prediction, which is more challenging given the unstable development trend of thunderstorms.



**Figure 2: Framework of conventional lightning prediction methods (inside the dotted box). Unlike conventional prediction, our **LightNet model** introduces recent lightning observation data to assist prediction besides simulation data.**

**Spatiotemporal Prediction Model in Deep Learning.** There are also some deep learning models focusing on spatiotemporal prediction tasks such as 3D ConvNets [29], PredCNN [27], DCRNN [26] and StepDeep [20]. To better model spatiotemporal information, 3D ConvNets [29] extend conventional 2-dimensional convolution, which slides only in the spatial dimension, to 3-dimensional convolution, which moves in both spatial and temporal dimensions. PredCNN [27] designs a cascade multiplicative unit and models the dependencies between the next frame and the spatiotemporal inputs by an entirely CNN based architecture. Traffic forecasting is one canonical example of spatiotemporal tasks. DCRNN [26] models traffic flow as a diffusion process on a directed graph and introduces diffusion convolutional recurrent neural network for traffic forecasting. StepDeep [20] formulates the problem of mobility event prediction as a spatiotemporal task and then designs a spatial-temporal progress cell to handle it. In summary, these approaches make predictions based on recent weather observations. By contrast, our model introduces simulation data besides the observation data, in a bid to enhance the prediction for a longer period.

### 3 PRELIMINARY

We first review the framework of conventional prediction methods and contrast it with the proposed LightNet. Secondly, we introduce the Weather Research and Forecasting (WRF) model, which serves as our numerical weather prediction system. Thirdly, we detail the input data of LightNet. Finally, we define the problem to be solved.

#### 3.1 Framework of Conventional Prediction

The framework of conventional lightning prediction methods is shown in the dotted box of Figure 2. The historical weather collection includes daily-updated meteorological data measured by weather detection equipment such as satellites, radars and sounding aircraft. At first, the numerical weather prediction system is initialized by the data from the historical weather collection. After initialization, this system will compute and simulate the weather condition for the future. The simulation has a structure of grids, where each grid represents a spatial local area. Lastly, a lightning

parameterization scheme is exploited to grid-wise convert the simulation to lightning forecasts.

As shown in Figure 2, unlike conventional prediction methods, our LightNet model additionally introduces recent lightning observation data besides simulation data, with the goal of calibrating the simulation data and assisting the prediction. Notice that, due to the complex computation rules, the numerical weather prediction system cannot directly assimilate the observation data and calibrate itself.

#### 3.2 Weather Research and Forecasting

The Weather Research and Forecasting (WRF<sup>1</sup>) Model [23] is a next-generation numerical weather prediction system designed for both atmospheric research and operational forecasting applications. The WRF model serves as a numerical weather prediction system to supply us the simulation data. In the present study, the WRF model, with a version 3.4.1, is initialized using data from the National Centers for Environmental Prediction (NCEP) global final (FNL) analysis data at a resolution of 1° and six-hour intervals.

#### 3.3 Data Description

Once the WRF model is initialized, it can run and simulate the weather condition for the future, where the weather condition is described by numerous meteorological parameters with a structure of grids. Here, we introduce the simulated parameters used in this work.

**Simulated Micro-Physical Parameters.** According to the recommendation in [24], three types of microphysical parameters, ice mixing ratio (QICE), snow mixing ratio (QSNOW) and graupel mixing ratio (QGRAUP), are selected in our study. QICE, QSNOW and QGRAUP present the ratio of the mass of ice, snow and graupel, respectively, per unit mass of dry air. Since all three parameters share THE same data structure, we only use QICE to illustrate. The simulated data of QICE can be described as a fourth-order tensor  $I$ , with three spatial indexes  $x, y, z$  (corresponding to longitude, latitude and altitude, respectively) and a temporal index  $t$ . Specifically,  $I_t^{x,y,z}$  stores the simulated value of QICE in a grid with a spatial coordinate  $(x, y, z)$  and at a time scale  $t$ . For convenience, we use  $I_t$  (a third-order tensor) to denote the slice at time  $t$ . Similarly, the simulated tensor of QSNOW and QGRAUP are denoted as  $S$  and  $G$ , respectively.

**Simulated Radar Reflectivity.** Radar reflectivity is a measure of the efficiency of a meteorological target (e.g. cloud, precipitation particles) in intercepting and returning radio energy. It is proportional to the number of drops per unit volume and the sixth power of drops' diameter and is thus used to estimate the rain or snow intensity [19]. We introduce the radar reflectivity simulated by the WRF model to improve the lightning prediction. We denote this data by  $R$ , which shares the same structure with the microphysical parameters.

**Simulated Maximum Vertical Velocity.** Maximum vertical velocity is the maximum wind velocity in a vertical direction. Price and Rind [16] found a strong relationship between the maximum vertical velocity and lightning activities. The simulated maximum

<sup>1</sup>Full formulation and documentation can be accessed through UCAR at: <https://www.mmm.ucar.edu/weather-research-and-forecasting-model>.

vertical velocity is described by a third-order tensor  $V$ , with two spatial indexes  $x, y$  (the dimension  $z$  vanished due to the maximum operation) and a time index  $t$ .

**Real-World Lightning Observations.** In addition to the WRF simulated data, we also access real-world lightning observations, provided by China National Lightning Detection Networks [25]. We denote this data by  $L$ , which is structured in the same way as the simulated maximum vertical velocity  $V$ . Specifically,  $L_t^{x,y}$  is a binary value indicating whether lightning happened during the period of  $[t, t + 1]$ .

### 3.4 Problem Definition

Given the WRF simulation data for next time slots and the lightning observation data for recent time slots, our goal is to predict the lightning activities for next time slots. We formally define the problem as below. Suppose the current moment is  $t = 0$ . We have access to the WRF simulation  $[P_t]_{t=0}^h$  ( $P_t = [I_t, S_t, G_t, R_t, V_t]$ ) and the recent lightning observations  $[L_t]_{t=-s}^{-1}$ . Our task is to predict  $[\hat{L}_t]_{t=0}^h$  and make them as close as possible to  $[L_t]_{t=0}^h$ , which is the real lightning observations for next time slots (i.e.  $[0, 1], [1, 2], \dots, [h, h + 1]$ ). Specifically, our goal is to find a mapping  $f$  such that

$$\begin{aligned} \min_f & \text{loss}([\hat{L}_t]_{t=0}^h, [L_t]_{t=0}^h) \\ \text{s. t. } & [\hat{L}_t]_{t=0}^h = f([P_t]_{t=0}^h, [L_t]_{t=-s}^{-1}) \end{aligned} \quad (1)$$

## 4 METHOD

In this section, we first introduce the convolutional long short-term memory (ConvLSTM) networks, which serve as a backbone of our model. Then we detail the proposed LightNet for spatiotemporal lightning prediction.

### 4.1 Convolutional Long Short Term Memory Networks

Convolutional long short-term memory (ConvLSTM) network is an extension of the popular fully connected long short-term memory (FC-LSTM) network, with a goal to overcome the drawbacks of FC-LSTM in handling spatial-temporal data such as videos [21]. Specifically, in the ConvLSTM network, the fully-connected gates of the FC-LSTM module are replaced by convolutional gates thereby making it capable of encoding spatial-temporal features. Following the conventional notations, we let  $X_t, H_t, C_t$  and  $Y_t$  denote the input, hidden state, cell state, and output of the ConvLSTM network at the  $t$ -th moment, respectively. The equations characterizing the ConvLSTM are shown as follows:

$$\begin{aligned} Z_t^i &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + B_i), \\ Z_t^f &= \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + B_f), \\ C_t &= Z_t^f \circ C_{t-1} + Z_t^i \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + B_c), \\ Y_t &= \sigma(W_{xy} * X_t + W_{hy} * H_{t-1} + W_{cy} \circ C_t + B_y), \\ H_t &= Y_t \circ \tanh(C_t) \end{aligned} \quad (2)$$

where  $*$ ,  $\circ$  and  $\sigma(\cdot)$  denote the convolution operator, the Hadamard product and the logistic sigmoid function, respectively.  $W$  and  $B$  are trainable parameters whose sizes match the corresponding tensors.

Note that all the gate activation and the hidden states in ConvLSTM are three-dimensional tensors in contrast to vectors in FC-LSTM, which lose the natural spatial information.

### 4.2 Model Architecture

Figure 3 illustrates the overall architecture of LightNet, comprising a WRF encoder, an observation encoder, a fusion module and a prediction decoder. Since the WRF simulation (for next six hours) and the lightning observations (for recent three hours) come from different time domains, LightNet introduces the WRF encoder and the observation encoder to convert the raw data into feature codes in a unified space for further combination. Then in this space, the WRF feature code and the observation feature code are fused via the fusion module. Finally, the prediction decoder is utilized to recursively decode the lightning predictions from the fused feature. We detail each component as follows.

**WRF Encoder.** The WRF encoder is responsible for encoding the WRF simulation data. We stack the microphysical parameters ( $S, I, G$ ), the radar reflectivity ( $R$ ) and the maximum vertical velocity ( $V$ ) along the  $z$  direction for each time slot. These parameters constitute the input of the WRF encoder. Specifically, we denote the input  $P = [P_t]_{t=0}^h$ , where  $P_t$  is defined as

$$P_t = \text{stack}^{z\dagger}(S_t, I_t, G_t, R_t, V_t). \quad (3)$$

$P$  has a large size due to the high resolution in  $x$ - $y$ - $z$  space. However, adjacent grids in this space usually share very similar value, which makes  $P$  somewhat redundant and increases the burden on model training. Therefore, we reduce the spatial resolution of  $P_t$  through a convolution layer with a ReLU activation:

$$\bar{P}_t = \text{ReLU}(W * P_t + B) = \text{Conv}_1(P_t). \quad (4)$$

Then, the compact input  $\bar{P} = [\bar{P}_t]_{t=0}^h$  is sent into a convLSTM network to obtain two WRF feature tensors:

$$\left[ C^{wrf}, H^{wrf} \right] = \text{convLSTM}^{wrf}(\bar{P}), \quad (5)$$

where  $C^{wrf}$  and  $H^{wrf}$  denote the cell state and the hidden state of the last moment, respectively. The detailed calculation steps of (5) were presented in (2).

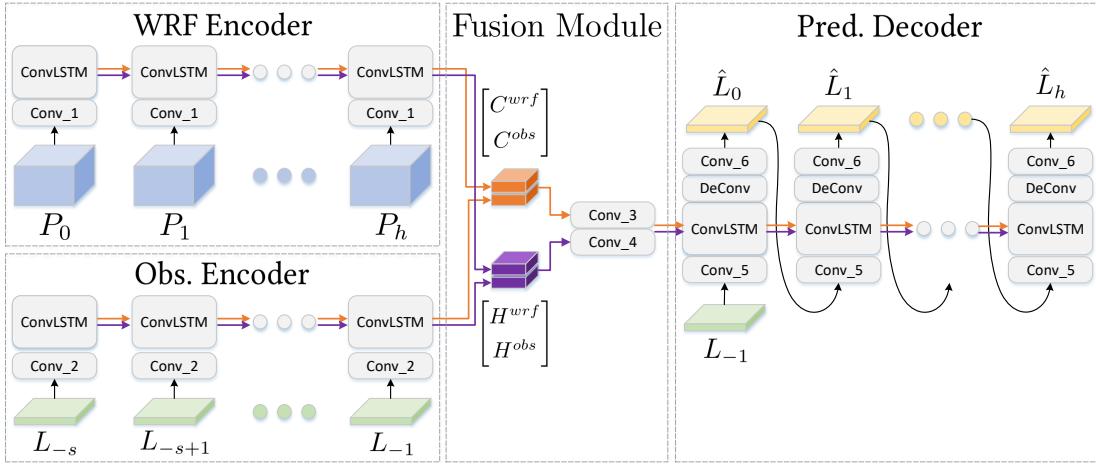
**Observation Encoder.** The observation encoder is in charge of extracting the information within lightning observations for recent  $s$  periods ( $L = [L_t]_{t=-s}^{-1}$ ). To be consistent with the WRF encoder, we also use a convolution layer to reduce the horizontal resolution of  $L$ :

$$\bar{L}_t = \text{Conv}_2(L_t). \quad (6)$$

Then the compact representation  $\bar{L} = [\bar{L}_t]_{t=-s}^{-1}$  is fed to a convLSTM network and it returns two observation feature tensors:

$$\left[ C^{obs}, H^{obs} \right] = \text{convLSTM}^{obs}(\bar{L}). \quad (7)$$

**Fusion Module.** The WRF simulation data can provide long-term weather information for the future, but it usually shows deviation in space and time domains. On the other hand, the observation data can offer the lightning distribution estimation for short-term prediction (e.g. less than three hours), whereas its effect degenerates for long-term prediction. Therefore, we introduce the fusion module in order to improve the short-term prediction and calibrate



**Figure 3: The diagram of LightNet.** LightNet consists of four parts: WRF encoder, observation (Obs.) encoder, fusion module and prediction (Pred.) decoder. The WRF encoder and the observation encoder are responsible for encoding the spatiotemporal features of WRF simulation ( $P = [P_t]_{t=0}^h$ ) and recent lightning observations ( $L = [L_t]_{t=-s}^{-1}$ ), respectively. Then the fusion module combines the WRF feature ( $\{C^{wrf}, H^{wrf}\}$ ) and observation feature ( $\{C^{obs}, H^{obs}\}$ ) so as to improve short-term prediction and calibrate long-term simulation. Finally, the prediction decoder receives the fused feature prediction and recursively makes forecasts ( $\hat{L} = [\hat{L}_t]_{t=0}^h$ ).

long-term simulation. Especially, the WRF features ( $\{C^{wrf}, H^{wrf}\}$ ) and observation features ( $\{C^{obs}, H^{obs}\}$ ) are fused as

$$\begin{aligned} C^{fuse} &= \text{Conv}_3 \left( \text{stack}^{z\uparrow}(C^{wrf}, C^{obs}) \right), \\ H^{fuse} &= \text{Conv}_4 \left( \text{stack}^{z\uparrow}(H^{wrf}, H^{obs}) \right). \end{aligned} \quad (8)$$

**Prediction Decoder.** The prediction decoder receives the fused features and makes predictions for the future. Considering chronological relation among these predictions, we employ a convLSTM network as the configuration of the prediction decoder. In addition, inspired by the method from machine translation [2], the prediction for the moment  $t$  will be sent as the input of forecasting for the moment  $t + 1$ . Specifically, at the moment  $t = 0$ , all variables are initialized as follows:

$$\begin{aligned} C_{-1} &= C^{fuse}, \\ H_{-1} &= H^{fuse}, \\ X_0 &= \text{Conv}_5(L_{-1}). \end{aligned} \quad (9)$$

Notice that we introduce the last lightning observation  $L_{-1}$  as the initial input to provide a good “starting point” for the subsequent predictions. After loading the initial values, the prediction decoder will run following the calculation as shown in (2). In particular, the input and the prediction at a moment  $t > 0$  are specified as follows:

$$\begin{aligned} X_t &= \text{Conv}_5(\hat{L}_{t-1}), \\ \hat{L}_t &= \text{softmax}((\text{Conv}_6(\text{DeConv}(Y_t))), \end{aligned} \quad (10)$$

where a deconvolution (DeConv) layer [15] is used to recover the horizontal resolution of the prediction. The prediction decoder will recursively run  $h$  times and give predictions  $\hat{L} = [\hat{L}_t]_{t=0}^h$ . The settings (the kernel size and stride) of each component of LightNet are detailed in Table 1.

**Table 1: The detailed settings of LightNet.**

Module	Notation	Size	stride
WRF Encoder	Conv <sub>1</sub>	[7 × 7, 64]	2
	ConvLSTM	[5 × 5, 128]	1
Obs. Encoder	Conv <sub>2</sub>	[7 × 7, 4]	2
	ConvLSTM	[5 × 5, 8]	1
Fusion Module	Conv <sub>3</sub>	[1 × 1, 64]	1
	Conv <sub>4</sub>	[1 × 1, 64]	1
Pred. Decoder	Conv <sub>5</sub>	[7 × 7, 4]	2
	ConvLSTM	[5 × 5, 64]	1
	DeConv	[7 × 7, 64]	2
	Conv <sub>6</sub>	[1 × 1, 1]	1

### 4.3 Implementation Details

The proposed neural networks are implemented with Keras<sup>2</sup>. All network parameters are initialized with a normal distribution.

**Weighted Loss Function.** In our dataset, the number of positive samples (grids with lightning) is much smaller than the number of negative samples. To meet the challenge of sample imbalance, we select the weighted cross entropy as the loss function. Formally, given the ground truth  $L = [L_t]_{t=0}^h$  and the predictions  $\hat{L} = [\hat{L}_t]_{t=0}^h$ , the weighted cross entropy between them is computed as

$$\text{loss}(L, \hat{L}) = - \sum_{t=0}^h \sum_{x,y} w L_t^{xy} \log(\hat{L}_t^{xy}) + (1 - L_t^{xy}) \log(1 - \hat{L}_t^{xy}), \quad (11)$$

where  $w$  is the weight of positive loss. We tune this hyper-parameter in {4, 8, 16, 32, 64} and find setting  $w = 16$  achieves the best performance.

<sup>2</sup><https://keras.io/>

**Training Details.** We use Adaptive Moment Estimation (Adam) [10] method to adapt learning rates. Adam essentially combines RMSProp [28] with momentum. The initial learning rate is set to 0.0001 without dropout. We train our model for 50 epochs and select the model with the lowest validation loss.

## 5 EXPERIMENTS

In this section, we evaluate the proposed model on a real lightning dataset of North China. In Section 5.1, we introduce the details of the dataset. Section 5.2 introduces several baseline methods coming from meteorology and machine learning. Performance metrics are described in Section 5.3. Section 5.4 compares LightNet with baseline approaches. In Section 5.5, we visualize six representative examples for further analysis. Finally, we analyze the effect of different configuration on our model in Section 5.6. We conduct experiments on a 64-bit Ubuntu 16.04 computer with 18 Intel 2.68GHz CPUs, 256 GB memory, and 8 NVIDIA TITAN X GPUs.

### 5.1 Dataset

**Prediction Region.** We access the data of a sub-region of North China from Chinese Academy of Meteorological Sciences<sup>3</sup>. Specifically, this region is a square with a center at 40°N and 116.2°E, which is divided into 159 × 159 grids and each grid has a resolution of 4km × 4km.

**WRF Simulation Data.** Due to the limitation of computing power, the WRF system can only provide four simulations in a day, at 0:00, 6:00, 12:00 and 18:00. Each simulation contains the weather parameter grids for the next 18 hours, with a time span of one hour. Each grid has nine vertical levels. We use the simulations for the first six hours since they have higher quality. The data covers June to September 2015 time period, and May to September in both 2016 and 2017.

**Observation Data.** The observation data shares the same spatiotemporal resolution and time period with the WRF simulation data.

Our task is to make lightning forecasts for future six hours (i.e., output  $\{\hat{L}_t\}_{t=0}^5$ ) based on six-hour WRF data ( $\{P_t\}_{t=0}^5$ ) and three-hour observation data ( $\{L_t\}_{t=-3}^{-1}$ ). We divide the total dataset by a ratio of 11:1:2 for training, validation and testing.

### 5.2 Baselines

To evaluate the effectiveness of our framework, we compare LightNet with seven baselines. Among them, PR92, LF1 and LF2 are conventional prediction methods in meteorology, and others are data-driven models.

**PR92.** The PR92 method [30] is the most well documented and commonly used lightning parameterization scheme in meteorology. It relates lightning flash rates to the fifth-power of vertical updraft velocities as follows:

$$F = 5.7 \times 10^{-6} v_{\max}^{4.5} \quad (12)$$

Here,  $v_{\max}$  is the maximum vertical velocity and  $F$  is lightning frequency (unit:  $\text{min}^{-1}$ ).

**LF1.** The LF1 method developed by McCaul et al. [4] forecasts flash rate as a linear function of graupel flux at -15°C. The graupel

<sup>3</sup><http://www.camscma.cn/>.

**Table 2: The detailed information for three performance metrics.**

Name	Equation	Range	Explanation
POD	$\frac{N_1}{N_1+N_3}$	[0,1]	The ratio of the number of hit lightnings to the number of observed lightnings.
FAR	$\frac{N_2}{N_1+N_2}$	[0,1]	The ratio of the number of false alarm lightnings to the number of foretasted lightnings.
ETS	$\frac{N_1-R}{N-N_4-R}$	$[-\frac{1}{3}, 1]$	The ratio of the number of hit lightnings to the number of events except for the correct rejections, and removed the contribution from hits by chance in random forecasts.

flux is the product of the upward vertical velocity and graupel mixing ratio. Specifically, the flash rate in LF1 is calculated as

$$F = 0.042(v_{-15}g_{-15}), \quad (13)$$

where  $v_{-15}$  and  $g_{-15}$  respectively denote the upward vertical velocity and graupel mixing ratio at -15°C.

**LF2.** The LF2 method [4] builds the relation of flash rate to graupel, snow, and cloud ice mixing ratios ( $\text{kg kg}^{-1}$ ) within a grid column. Its expression is

$$F = 0.02 \int \rho(g_z + s_z + i_z) dz, \quad (14)$$

where  $g_z$ ,  $s_z$  and  $i_z$  respectively denote the graupel, snow, and cloud ice mixing ratios at height  $z$ .

**GBDT.** Gradient Boost Decision Tree (GBDT) is widely used in data mining applications these years. LightGBM [9] is a state-of-art GBDT implementation used for supervised learning problems and has a good performance in many areas. We select LightGBM 2.0.4 as a baseline to predict the lightning.

**StepDeep.** StepDeep [20] is a prediction framework based on 3D ConvNets (3-dimensional convolutional networks), which are proposed for spatiotemporal feature learning using 3-dimensional (introducing time dimension) convolutional kernels. It demonstrated superiority for spatiotemporal mobility event prediction. We first extract the features of the WRF simulation data utilizing StepDeep, and then send the WRF features and the recent lightning observations into an additional CNN module to make the predictions.

**LightNet-W.** An incomplete LightNet in which observation encoder is removed.

**LightNet-O.** An incomplete LightNet in which WRF encoder is removed.

### 5.3 Performance Metric

We evaluate the prediction results by three metrics: probability of detection (POD), false alarm ratio (FAR) and equitable threat score (ETS), which have been widely applied in meteorology [3]. Let  $N$  denote the total number of grids. Let  $N_1$ ,  $N_2$ ,  $N_3$  and  $N_4$  denote the number of true-positive, false-positive, false-negative and true-negative grids, respectively. Define the expectation of the number

**Table 3: Illustration for quantitative results (POD(%), FAR(%), ETS). The best performance is reported using bold red, and the second best is reported using bold blue. The term “Neighb.” is the abbreviations of “Neighborhood”.**

Method	First-Hour Cumulative Score						First-Three-Hour Cumulative Score						Six-Hour Cumulative Score					
	Strict Metric			Neighb.-Based Metric			Strict Metric			Neighb.-Based Metric			Strict Metric			Neighb.-Based Metric		
	POD	FAR	ETS	POD	FAR	ETS	POD	FAR	ETS	POD	FAR	ETS	POD	FAR	ETS	POD	FAR	ETS
PR92	7.1	99.0	0.006	31.4	94.9	0.042	18.0	96.9	0.019	52.1	88.5	0.094	23.2	95.2	0.027	54.7	86.1	0.110
LF1	6.0	96.8	0.002	15.8	93.6	0.035	15.6	97.1	0.018	45.3	86.2	0.100	18.8	93.8	0.037	47.4	88.0	0.109
LF2	6.6	97.9	0.013	22.3	91.9	0.060	15.0	94.5	0.035	37.7	82.8	0.126	19.1	92.0	0.048	39.5	80.5	0.139
GBDT	11.3	94.7	0.034	33.6	81.2	0.134	19.3	90.0	0.065	50.0	61.8	0.264	22.6	86.1	0.084	50.0	61.8	0.264
StepDeep	67.5	<b>81.4</b>	<b>0.168</b>	87.9	<b>50.5</b>	<b>0.459</b>	42.7	<b>74.1</b>	<b>0.187</b>	67.4	<b>40.2</b>	<b>0.458</b>	26.3	<b>70.9</b>	0.152	48.1	<b>35.2</b>	0.373
LightNet-W	8.7	82.2	0.041	24.4	75.3	0.138	9.8	83.6	0.062	23.1	58.1	0.171	7.8	76.9	0.058	18.2	45.5	0.154
LightNet-O	<b>72.3</b>	82.1	0.164	<b>89.7</b>	51.8	0.452	<b>59.1</b>	<b>77.5</b>	<b>0.189</b>	<b>79.8</b>	<b>46.6</b>	<b>0.462</b>	<b>41.7</b>	73.7	<b>0.182</b>	<b>63.9</b>	41.6	<b>0.428</b>
LightNet	<b>71.4</b>	<b>81.9</b>	<b>0.166</b>	<b>89.0</b>	<b>51.5</b>	<b>0.453</b>	<b>59.8</b>	77.7	<b>0.187</b>	<b>80.2</b>	47.4	<b>0.458</b>	<b>46.5</b>	<b>73.3</b>	<b>0.194</b>	<b>68.0</b>	<b>41.3</b>	<b>0.449</b>

of hit lightnings in random forecasts as  $R = (N_1 + N_2)(N_1 + N_3)/N$ . The equations for the three metrics are detailed in Table 2.

The above described three measures are called strict metrics since they consider that a lightning is hit (i.e. predicted correctly) only if the lightning strictly falls in the predicted grid. By contrast, neighborhood-based metrics relax the criteria for “hits” by considering grid points within a specified radius  $r$ . Specifically, a prediction is called an  $r$ -neighborhood-based hit if a lightning falls in the area within the radius  $r$  centered at the predicted grid. Accordingly, denote  $\tilde{N}_1$ ,  $\tilde{N}_2$ ,  $\tilde{N}_3$  and  $\tilde{N}_4$  as the number of  $r$ -neighborhood-based true-positive, false-positives, false-negative and true-negative grids, respectively. The  $r$ -neighborhood-based POD, FAR and ETS are computed by the equations as shown in Table 2, where  $\{N_i\}_{i=1}^4$  is replaced by  $\{\tilde{N}_i\}_{i=1}^4$ . In our experiments, we set the radius  $r = 1$  (i.e. the eight neighborhoods). Notice that the strict metrics can be seen as a special case of the neighborhood-based metrics with  $r = 0$ . Clark et al. [3] have shown that the neighborhood-based metrics demonstrate better properties than the strict metrics in assessing the prediction for high-resolution grids.

#### 5.4 Comparison with Baselines

Table 3 presents the performance of LightNet compared with the three meteorologic methods and four data-driven baselines. We evaluate their prediction results for three periods, i.e. the first hour, first three hours and all six hours. For each period, we demonstrate their scores of both strict metrics and neighborhood-based metrics (with radius  $r = 1$ ). The upper part shows the results of the three meteorologic baselines. The lower part reports the performance of all data-driven models including LightNet.

From Table 3, we observe that the three methods from meteorology show poor performance compared with data-driven methods. This result is caused by two factors. Firstly, they do not exploit the recent observation data and are completely based on the simulation of WRF, which usually have deviations in the space and time domain. Secondly, these methods are designed manually by meteorological experts, which hardly benefits from massive historical data. In contrast, data-driven methods achieve improvements thanks to the experience mined from training data.

For data-driven models, we observe that LightNet and LightNet-O achieve comparable performance on the first hour and first three hours, which will be explained in Subsection 5.6. StepDeep obtains lower scores but is still competitive. In contrast, GBDT and

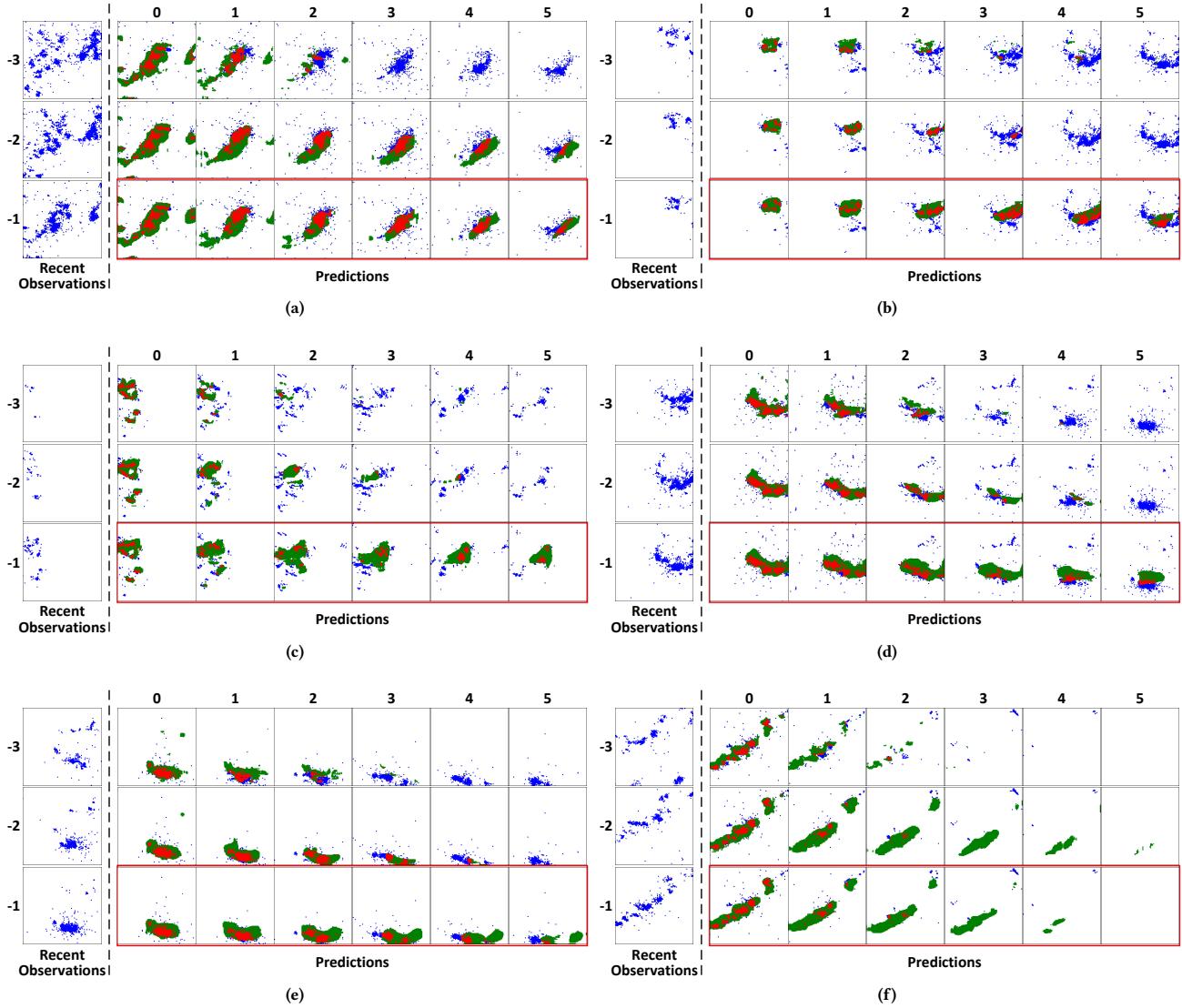
LightNet-W show similar poor performance among data-driven methods, since they are entirely based on the WRF simulation data, which usually shows deviations in space and time domain. Next, we focus on the three competitive methods (StepDeep, LightNet-O and LightNet) and compare their performance for different time periods as follows:

- (1) First-hour prediction. LightNet-O achieves the highest POD values for both strict and neighborhood-based metrics, which are slightly higher than the second place, LightNet. Though StepDeep is not very competitive in the POD measure, it records the highest ETS due to its lower FAR. For neighborhood-based metrics, we find that their performance gaps are small (below 0.01), since all their predictions are mainly based on the same information provided by  $L_{-1}$ .
- (2) First-three-hour prediction. LightNet and LightNet-O share similar strict POD values and have an obvious advantage (about 10%) than the third place, StepDeep. However, higher POD values also cause higher FARs, which are 77.7% and 77.5% for LightNet and LightNet-O, respectively. As a result, LightNet, LightNet-O and StepDeep obtain similar ETSs. The performance of these methods on neighborhood-based metrics follows what they do on the strict metrics.
- (3) Six-hour prediction. LightNet achieves the highest scores on strict and neighborhood-based POD metrics, which are higher than the second place (LightNet-O) by 5% and 4%, respectively. Furthermore, LightNet has an advantage on the strict ETS measure, which is more obvious on the neighborhood-based measure. On the other hand, though StepDeep has the lowest FAR, its poor POD leads to its low ETS.

#### 5.5 Visualization Results

Figure 4 visualizes six representative cases for StepDeep, LihgtNet-O and LightNet. From this figure, we observe that all methods make a good prediction for the first hour, which benefits from the information provided by the last-hour observation. On the other hand, StepDeep tends to predict no lightnings for the last three hours of the forecast period, which is consistent with its performance on quantitative metrics shown in Table 3. Next, we focus on comparing the results of LightNet-O and LightNet.

For the case shown in Figure 4a, both LightNet-O and LightNet correctly predict the main area of lightning for six hours, and the difference between their predictions is small. However, for the



**Figure 4: Visualization of six representative cases for six-hour prediction.** In each example, the left part is observations for recent three hours, and the right part is predictions for future six hours, where the predictions from top to bottom are made by StepDeep, LightNet-O and LightNet. The observation, prediction and hit are marked by blue ( $\bullet$ ), dark green ( $\bullet$ ) and red ( $\bullet$ ), respectively. Best viewed on a color screen.

cases shown in Figures 4b, 4c and 4d, LightNet-O tends to predict no lightnings for the last several hours. This is because the observations for recent hours provide insufficient information about the development trend of thunderstorms. In contrast, LightNet captures the approximate area of thunderstorms thanks to the knowledge extracted from the WRF data. For the case shown in Figure 4e, LightNet captures the core thunder zone for the first three hours, but the prediction gradually deviates from the right track for the last three hours. Figure 4f shows a case where thunderstorms gradually disappear. It seems that this trend is predicted in all methods.

However, from the above cases, we speculate that the trend predicted by StepDeep and LightNet-O is more likely to be an inertial behavior. By comparison, LightNet is more possible to anticipate the disappearing trend in light of the WRF simulation.

## 5.6 Model Analysis

To demonstrate the effect of different configuration on our model, we also evaluate the prediction performance of LightNet-W, LightNet-O and LightNet for the last three hours and show their scores in Table 4. The prediction scores for the first three hours have been shown in Table 3.

**Table 4: Quantitative results (POD(%), FAR(%), ETS) for last-three-hour prediction. The best performance is reported using bold dark. The term “Neighb.” is the abbreviations of “Neighborhood”.**

Method	Last-Three-Hour Cumulative Score					
	Strict Metric			Neighb.-Based Metric		
	POD	FAR	ETS	POD	FAR	ETS
LightNet-W	1.8	87.0	0.015	5.9	59.3	0.053
LightNet-O	12.3	83.8	0.071	26.9	60.0	0.187
LightNet	<b>22.3</b>	<b>82.8</b>	<b>0.102</b>	<b>42.3</b>	<b>59.1</b>	<b>0.257</b>

For LightNet-W, we observe that neither the first-three-hour prediction nor the last-three-hour prediction is acceptable, which illustrates that only exploiting WRF data is insufficient to make a good prediction. On the other hand, for LightNet-O, the first-three-hour prediction is very competitive, but its performance rapidly degenerates for the last three hours. This is because the recent observations can hardly provide the development trend of thunderstorms for such long hours. In contrast, LightNet not only achieves comparable performance on the first-three-hour prediction, but also has an obvious advantage for the last three hours. This demonstrates the necessity of combining WRF simulation data and the recent observation data to make a good prediction.

## 6 CONCLUSIONS

In this paper, we presented a model (named LightNet) for lightning prediction. LightNet is an end-to-end deep neural network, which overcomes the drawback of conventional prediction methods which fail to mine knowledge from historical data. LightNet exploits dual encoders to extract spatiotemporal features of WRF simulation data and recent lightning observation data. Furthermore, these features are combined by a fusion module, which proves effective in withstanding the deviations in the simulation and enhancing the prediction. We conducted the experiments on a real-world lightning dataset of North China. The results show that LightNet achieves a threefold improvement in equitable threat score for the six-hour prediction compared with three established forecast methods.

In future research, we will extend the current work in three aspects. Firstly, we will continue to improve the forecast quality and further predict the intensity of lightning. Secondly, we will expand the forecast duration from six hours to over twelve hours. Thirdly, we will generalize our framework and employ it to predict other meteorological disasters such as windstorm and hailstone.

## ACKNOWLEDGMENTS

This work is partly supported by the National Key Research and Development Program of China (No. 2017YFC1501503), Beijing Municipal Education Commission Research Program (No. SM20191001107, PXM2019\_014213\_000007) and National Social Science Foundation of China (No. 18CSH019).

## REFERENCES

- [1] Louis J. Battan. 1964. Some Observations of Vertical Velocities and Precipitation Sizes in a Thunderstorm. *Journal of Applied Meteorology* 3, 4 (1964), 415–420.
- [2] Kyunghyun Cho *et al.* 2014. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In *Proc. EMNLP*. Association for Computational Linguistics, Doha, Qatar, 1724–1734.
- [3] Adam J Clark, William A Gallus Jr, and Morris L Weisman. 2010. Neighborhood-based verification of precipitation forecasts from convection-allowing NCAR WRF model simulations and the operational NAM. *Weather and Forecasting* 25, 5 (2010), 1495–1509.
- [4] Kristin Ashley Cummings. 2017. *An Investigation of the Parameterized Prediction of Lightning in Cloud-Resolved Convection and the Resulting Chemistry*. Ph.D. Dissertation. University of Maryland, College Park.
- [5] E Brian Curran, Ronald L Holle, and Raúl E López. 2000. Lightning casualties and damages in the United States from 1959 to 1994. *Journal of Climate* 13, 19 (2000), 3448–3464.
- [6] Wiebke Deierling *et al.* 2008. The relationship between lightning activity and ice fluxes in thunderstorms. *Journal of Geophysical Research: Atmospheres* 113, D15 (2008).
- [7] Theodore M Giannaros, Vassiliki Kotoni, and Konstantinos Lagouvardos. 2015. Predicting lightning activity in Greece with the Weather Research and Forecasting (WRF) model. *Atmospheric Research* 156 (2015), 1–13.
- [8] Todd S Glickman and Walter Zenk. 2000. *Glossary of meteorology*. American Meteorological Society.
- [9] Guolin Ke *et al.* 2017. Lightgbm: A highly efficient gradient boosting decision tree. In *Proc. NIPS*. 3146–3154.
- [10] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2014).
- [11] Benjamin Klein, Lior Wolf, and Yehuda Afek. 2015. A dynamic convolutional layer for short range weather prediction. In *Proc. IEEE CVPR*. 4840–4848.
- [12] J Latham and Basil John Mason. 1961. Generation of electric charge associated with the formation of soft hail in thunderclouds. *Proc. Royal Society London Series A. Mathematical and Physical Sciences* 260, 1303 (1961), 537–549.
- [13] Yunjie Liu *et al.* 2016. Application of deep convolutional neural networks for detecting extreme weather in climate datasets. *arXiv preprint arXiv:1605.01156* (2016).
- [14] N *et al.* Michalon. 1999. Contribution to the climatological study of lightning. *Geophysical research letters* 26, 20 (1999), 3097–3100.
- [15] Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han. 2015. Learning Deconvolution Network for Semantic Segmentation. In *Proc. IEEE ICCV*. Washington, DC, USA, 1520–1528.
- [16] Colin Price and David Rind. 1992. A simple lightning parameterization for calculating global lightning distributions. *Journal of Geophysical Research: Atmospheres* 97, D9 (1992), 9919–9933.
- [17] Elisabetta Renni, Elisabeth Krausmann, and Valerio Cozzani. 2010. Industrial accidents triggered by lightning. *Journal of hazardous materials* 184, 1–3 (2010), 42–48.
- [18] SE Reynolds, M Brook, and Mary Foulks Gourley. 1957. Thunderstorm charge separation. *Journal of Meteorology* 14, 5 (1957), 426–436.
- [19] R. R. Rogers and Man Kong, Yau. 1989. *A short course in cloud physics* (3rd ed. ed.). Pergamon Press Oxford ; New York.
- [20] Bilong Shen *et al.* 2018. StepDeep: A Novel Spatial-temporal Mobility Event Prediction Framework based on Deep Neural Network. In *Proc. KDD*. ACM, 724–733.
- [21] Xingjian SHI *et al.* 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Proc. NIPS*. 802–810.
- [22] Xingjian Shi *et al.* 2017. Deep learning for precipitation nowcasting: A benchmark and a new model. In *Proc. NIPS*. 5617–5627.
- [23] W. C. Skamarock *et al.* 2008. *A description of the Advanced Research WRF Version 3*. Technical Report. National Center for Atmospheric Research. 1–113 pages.
- [24] McCaul Jr *et al.* 2009. Forecasting lightning threat using cloud-resolving model simulations. *Weather and Forecasting* 24, 3 (2009), 709–729.
- [25] Meng Qing *et al.* 2006. Development of national lightning detection network and its application in China.
- [26] Yaguang Li *et al.* 2018. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. In *Proc. ICLR*.
- [27] Ziru Xu *et al.* 2018. PredCNN: Predictive learning with cascade convolutions. In *Proc. IJCAI*. 2940–2947.
- [28] Tijmen Tieleman and Geoffrey Hinton. 2012. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning* 4, 2 (2012), 26–31.
- [29] Du Tran *et al.* 2015. Learning spatiotemporal features with 3d convolutional networks. In *Proc. IEEE ICCV*. 4489–4497.
- [30] Ying Wang, Yi Yang, and Shuanglong Jin. 2018. Evaluation of lightning forecasting based on one lightning parameterization scheme and two diagnostic methods. *Atmosphere* 9, 3 (2018), 99.
- [31] Haoliang Wang *et al.* 2017. Progress in lightning forecast by using numerical weather models and model outputs. *Advances in Earth Science* 32, 1 (2017), 44–55.
- [32] Yunbo Wang *et al.* 2017. Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. In *Proc. NIPS*. 879–888.