ISSN 2959-6157

Learning to Predict the Unpredictable: Deep Models for Severe Convective Weather

Qiaoyue Fang

Data Science and Big Data Technology, North China University of Technology, Beijing, China
Corresponding author: 22190010313@mail.ncut.edu.cn

Abstract:
Breakthroughs in deep learning have ushered in transformative opportunities for interdisciplinary research in the emerging field of "AI + Meteorology." Among the most challenging and societally impactful problems in this domain is the prediction of severe convective weather, which is characterized by highly dynamic and complex atmospheric processes. This paper provides a comprehensive overview of recent theoretical advancements and methodological innovations in applying deep neural networks to severe convective weather forecasting. It systematically reviews the limitations of traditional numerical and statistical methods, discusses representative datasets and evaluation metrics, and emphasizes the integration of physical and data-driven modeling principles. The application and performance of various deep learning models—including recurrent and non-recurrent architectures, generative approaches, and large-scale meteorological models—are thoroughly analyzed. In addition, the paper highlights critical challenges such as long-tailed data distributions, model interpretability, and lack of physical consistency. Finally, it outlines prospective research directions and open questions, aiming to offer both theoretical insights and practical guidance for developing next-generation intelligent weather prediction systems.

Keywords: Deep Learning; Severe Convective Weather Prediction; Generative Adversarial Networks; Meteorological Large Models.

1. Introduction

The global climate system is undergoing dramatic changes with a significant warming trend. From 1880 to 2023, the global average temperature rise has reached 1.15±0.13℃, which has directly intensified the frequency and destructive intensity of severe convective weather [1]. Observation data indicate that

ISSN 2959-6157

global severe convective weather events have increased at a rate of 8.7% per decade between 1973 and 2022, and their characteristics of suddenness and locality pose severe challenges to disaster prevention and mitigation systems [2].

Severe convective weather is an umbrella term for hazardous weather phenomena caused by intense vertical air movement, including thunderstorm gales (wind speed ≥17m/s), short-duration heavy rainfall (≥20mm per hour), and hailstorms (diameter ≥5mm) [3]. Such weather, characterized by rapid variability and high destructiveness, remains a key challenge in weather forecasting.

Traditional prediction methods have notable limitations: Numerical Weather Prediction (NWP), which relies on solving physical equations, offers physical consistency but comes with high computational costs and insufficient accuracy in short-term forecasts (0–2 hours). Extrapolation methods, while fast and efficient as they extrapolate based on observation data, struggle to capture the formation and dissipation of systems [4].

As a representative data-driven approach, deep learning can uncover underlying patterns from massive historical observation data and has demonstrated unique advantages in fields like image processing and time-series prediction. Combined with data from remote sensing satellites, radar detection, and ground observations, the application of deep learning in severe convective weather prediction has become a research hotspot. This paper will conduct a systematic review of key methods, typical models, and application practices of deep learning in severe convective weather prediction, synthesize the current research status, summarize the technological evolution path, and discuss current challenges and future directions.

2. Traditional Methods for Severe Convective Weather Prediction

2.1 Numerical Weather Prediction

Numerical Weather Prediction (NWP) quantitatively predicts future atmospheric states by solving fluid dynamics and thermodynamics equations describing atmospheric motion using large-scale computers [4]. Its workflow includes data collection, data assimilation, model operation, and forecast output, with its physical basis rooted in the laws of thermodynamics and Newton's laws of motion, offering significant advantages in mathematical rigor and physical consistency.

NWP is widely applied in short-term and medium-term weather forecasting, covering predictions of pressure, temperature, humidity, wind, clouds, and precipitation, and has become the primary method in meteorological forecasting. Optimizations of high-resolution numerical models and fusion of multi-source data have driven progress in severe convection prediction: Research utilizing the Weather Research and Forecasting (WRF) model has markedly enhanced the simulation precision of thunderstorms, hailstorms, and other such systems by optimizing cloud microphysical parameterization schemes and turbulent diffusion mechanisms;the Zhejiang Rapid Update Assimilation System (ZJWARRS), with a spatial resolution of 0.03°×0.03° and a 3-hour update frequency, has enhanced the spatiotemporal refinement capability of short-term forecasts.

However, NWP has inherent limitations: Errors in the initial field accumulate over time, and complex terrain interferes with boundary layer processes, leading to forecast biases in regions such as mountainous areas; it incurs high computational costs and requires time integration for initialization and simulation, resulting in poor performance in 0–2 hour nowcasting [5, 6]. To address these shortcomings, scholars have integrated machine learning with NWP; for example, the XGBoost algorithm optimizes the prediction of severe convective precipitation areas through ensemble decision trees, increasing the Threat Score (TS) by approximately 15% compared to traditional methods.

2.2 Extrapolation Forecasting Methods

Extrapolation methods, based on the assumption that "weather systems move continuously," predict future states using radar and satellite observation data. Characterized by small data volume and fast computation, they serve as an important supplement to NWP [7]. Mainstream methods include cross-correlation algorithms and optical flow methods:

The cross-correlation algorithm (TREC) ascertains the motion vector features of echoes through computing the optimal spatial coherence of distinct regions in radar echoes at adjacent time points, thereby projecting future locations and rendering it a commonly employed tracking method [8]. Improved TREC-based algorithms such as COTREC (optimizing vector continuity) and DITREC (incorporating differential images) have further enhanced the stability of precipitation forecasts [9]. However, TREC and its extensions can only predict positional changes of precipitation systems and fail to characterize trends in their intensity, leading to high failure rates in tracking intense precipitation echoes.

Optical flow techniques compute the optical flow field of radar echoes to acquire the motion vector field, offsetting the limitations of cross-correlation methods and enhancing the performance of convective nowcasting systems

[10]. Nevertheless, optical flow methods still struggle to cope with the rapid changes of severe convective weather, have limited ability to capture long- and short-term features, and cannot effectively simulate the formation and dissipation of echoes.

Extrapolation methods also face multiple challenges in application: The chaotic nature of severe convective weather introduces inherent uncertainty in long-term predictions; relying on historical data to infer future trends makes them unable to reflect new patterns caused by climate change, potentially producing misleading results; the local characteristics of severe convection may be diluted in large-scale datasets, leading to prediction biases in regions with sparse or insufficiently representative data.

3. Datasets and Evaluation Metrics

3.1 Key Datasets

High-quality datasets form the foundation of deep learning modeling. Common datasets for severe convection prediction include radar data, multimodal data, and reanalysis data:

Among radar datasets, IowaRain and RYDL provide precipitation observations with resolutions below 1km, suitable for short-term nowcasting tasks of precipitation. The Multi-Radar Multi-Sensor System (MRMS) provides multi-radar precipitation data covering the United States with a 2-minute temporal resolution, and is widely utilized in analyzing and predicting short-duration heavy rainfall events [11]. The OPERA dataset focuses on the European region, providing radar precipitation data with a 2km resolution to support research on large-scale precipitation prediction.

Represented by SEVIR, multimodal datasets integrate radar, satellite, and lightning observations, offering multimodal meteorological data that covers the complete observation of storm lifecycles, making them suitable for research on multi-source data fusion [12].

Reanalysis datasets are comprehensive reconstructed data of global or regional meteorological elements. As the most widely used numerical weather prediction reanalysis data globally, ERA5 provides data with a 0.25° spatial resolution and a 1-hour temporal step, serving as one of the standard datasets for deep learning in weather prediction tasks. Currently, mainstream meteorological large models are all trained on this dataset [13]. CMA-RA V1.5 is a high-resolution global reanalysis dataset developed by the China Meteorological Administration (CMA), offering two spatial resolutions (10km and 25km) and a 1-hour temporal resolution. It integrates 40 years of historical observation data and data from 48 global satellites, adopting advanced assimilation techniques such as Hybrid-4DEnVar and EnKF. Compared with CMA-RA V1.0, the root mean square error of 500hPa geopotential height is reduced by 22%. Its high spatiotemporal resolution and multi-variable characteristics make it an important data source for training meteorological large models, short-term weather prediction, and long-term climate analysis [14].

3.2 Evaluation Metric System

Evaluating model performance requires constructing an indicator system from multiple dimensions, including global accuracy, binary accuracy, downscaling accuracy, and clarity:

Global accuracy metrics measure the overall proximity between predicted results and observed data, covering all pixels, and play a key role in verifying overall prediction performance.Mean Squared Error (MSE) and Mean Absolute Error (MAE) serve as key metrics for assessing the comprehensive model structure, where lower values signify better performance of the prediction model. The Pearson Correlation Coefficient (PCC) is commonly employed to gauge the statistical association between observed and forecast outcomes; the nearer its value to 1, the more robust the linear association between the two.

Binary accuracy assesses precipitation intensity on a pixel-by-pixel basis using a confusion matrix derived from precipitation thresholds. It focuses on determining whether each pixel's precipitation volume and intensity align with predefined criteria (such as the 20mm/h benchmark for short-duration heavy rainfall). Typical metrics in this category include the Critical Success Index (CSI), Accuracy, F1-score, Precision, Recall, Probability of Detection (POD), and False Alarm Rate (FAR). In precipitation evaluation, confusion matrices should be computed for complete batches rather than individual small subsets. Special attention must be paid to non-precipitation cases to prevent assessment distortions arising from sample imbalance.Downscaling Accuracy (DSA) reflects the consistency between the conversion of low-resolution data to high-resolution data by the model and the real high-resolution data. Calculations must be performed on multi-scale validation sets to avoid single-resolution biases, and invalid data regions must be explicitly excluded. If high-resolution ground truth is missing, it should be marked as NA instead of zero to ensure evaluation accuracy. Clarity refers to the sensitivity of generated prediction images and is a key criterion for evaluating the quality of precipitation region boundaries and blurring issues. Since extreme precipitation events typically involve small-scale convective features, the clarity of generated images direct-

ISSN 2959-6157

ly affects the accuracy of characterizing severe convective system structures. Frequently used metrics encompass Gradient Difference Loss (GDL), Learned Perceptual Image Patch Similarity (LPIPS), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM). That said, these approaches have constraints and fail to fully indicate whether clarity aligns with physical patterns. As a result, they should be integrated with other assessment methods to comprehensively gauge model performance.

4. Comparative Analysis of Deep Learning Models in Severe Convection Prediction

4.1 Classical Deep Learning Models

Recurrent models capture temporal dependencies through recursive computation, primarily used for modeling short-term weather evolution. Their core lies in accumulating and learning temporal dynamics through hidden states, performing strongly in prediction tasks on shorter time scales (e.g., radar echo extrapolation, short-term precipitation prediction). CNN-based recurrent methods incorporate Convolutional Neural Networks (CNN) into traditional RNN structures, with ConvLSTM as a typical representative. They replace fully connected operations in LSTM with convolutional operations, enhancing spatial awareness when processing gridded meteorological data (e.g., radar images, satellite cloud images) while reducing the number of parameters. In ConvLSTM, the input x, cell state c, hidden state H, and their gating units (input gate i, forget gate f, output gate o) are all represented as 3D tensors, with state updates performed using local neighborhood information.

PredRNN optimizes temporal modeling capabilities through multi-level memory units on this basis. E3D-LSTM enhances the correlation of spatiotemporal features. MIM adopts cascaded memory modules combined with hidden state difference operations to improve the ability to model high-order non-stationarity. SwinLSTM incorporates the window attention mechanism of Swin Transformer, breaking through the limitation of the local receptive field of convolution and enhancing the capture of global spatial dependencies. Its core calculation formula simplifies and integrates the gating mechanism into a single gate.

Non-recurrent models enhance computational efficiency and mitigate the vanishing gradient problem in information propagation by eliminating recursive computation and adopting a parallel computing structure. They primarily rely on architectures such as CNN, Transformer, and frequency-domain transforms for spatiotemporal modeling, and demonstrate significant advantages in computational efficiency, scalability, and training stability. The development of non-recurrent methods can be divided into two stages: before and after SimVP. Models prior to SimVP (e.g., RainNet, STConvS2S) removed recurrent units but still relied on stepwise rolling inference for multi-step prediction, leading to error accumulation. The proposal of SimVP marked the entry of non-recurrent methods into the end-to-end multi-step prediction stage. Through an encode-transform-decode architecture, it generates complete spatiotemporal prediction sequences in one go, fundamentally optimizing the application of non-recurrent methods in weather prediction. MIMO-VP, based on Transformer's multi-input multi-output mode, captures long temporal dependencies via the self-attention mechanism and outperforms SimVP on multiple benchmark datasets. As a standardized spatiotemporal prediction framework, OpenSTL introduces the MetaFormer structure to connect the fields of computer vision and meteorology, transferring visual models such as ViT and Swin Transformer to precipitation forecasting to enhance cross-scale feature fusion capability.

4.2 Generative Models and Meteorological Large Models

Generative models capture the uncertainty of weather systems through modeling probability distributions, mainly including Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and diffusion models, which have demonstrated significant advantages in tasks such as radar echo prediction, precipitation forecasting, and extreme weather prediction. GANs consist of a generator and a discriminator. Through adversarial training, the generator improves the authenticity of generated samples. In weather prediction, they generate possible future weather fields by learning the statistical characteristics of historical data. MoCoGAN performs motion modeling by combining motion-content decomposition with RNN to generate weather evolution sequences; TsGAN adopts a two-stage GAN structure and combines ConvLSTM to optimize the spatiotemporal consistency of precipitation prediction ; PID-GAN introduces a physics-constrained discriminator to ensure that the generated weather fields conform to physical laws and reduce the risk of "plausible but implausible" results.Diffusion models generate data through progressive noise addition and reversing the denoising process, enabling the generation of smoother, more stable, and high-resolution weather fields. PreDiff combines latent variable diffusion with the Earthformer-UNet structure to enhance the uncertainty quantification capability of precipitation probability prediction; DiffCast decomposes global deterministic motion trends and local random variations, and combines GTUNet for temporal modeling, effectively improving the stability of precipitation prediction on long-time scales.

Meteorological large models leverage massive data and advanced architectures to achieve efficient global/regional forecasting. Adopting an end-to-end learning paradigm and integrating deep neural networks, physics-guided mechanisms, and self-supervised learning, they demonstrate strong generalization capabilities in long-time-scale and large-scale weather prediction. FourCastNet employs Adaptive Fourier Neural Operators (AFNO) for global weather prediction, trained on ERA5 data (0.25° resolution, 6-hour intervals). Its core innovation lies in combining Fourier Neural Operators (FNO) to model the evolution of large-scale weather systems in the frequency domain, with a computational speed nearly 10,000 times faster than traditional NWP, enabling completion of 10-day global predictions in less than 2 seconds. GraphCast uses Graph Neural Networks (GNNs) for global weather state modeling, trained on ERA5 data (0.25° resolution, 6-hour time steps, 227 variables). Its key innovation is the "Encode-Process-Decode" structure, which enables efficient information propagation through multi-layer GNNs. It outperforms traditional numerical models in over 90% of meteorological variables, particularly excelling in tasks such as tropical cyclone track prediction and modeling of extreme weather events (e.g., atmospheric rivers). Pangu-Weather utilizes a 3D Earth-specific Transformer (3DEST) for global weather prediction, trained on ERA5 data (0.25° resolution, 13 atmospheric variables and 4 surface variables). Its core innovation is integrating 3D Transformer to capture interactions between different pressure layers, with lower RMSE than ECMWF IFS for key variables such as 500hPa geopotential height and 2m temperature, and 7-day global forecasts taking less than 2 seconds.

5. Conclusion

While deep learning has demonstrated remarkable potential in severe convective weather prediction—with recurrent models excelling in short-term forecasting, non-recurrent models boosting long-term efficiency, generative models strengthening uncertainty characterization, and large meteorological models enabling global, efficient prediction—it still faces practical challenges. On the data front, extreme weather samples are scarce, and observations in regions such as plateaus and oceans remain insufficient, limiting models' ability to identify and capture the evolution of extreme events.

References

[1] World Meteorological Organization (WMO). State of the Global Climate 2024. Geneva: WMO Publications, 2024: 23-45.

[2] IPCC. Climate Change 2023: The Physical Science Basis [M]. Cambridge: Cambridge University Press, 2023.

[3] Meteorological Industry Standard. Severe Convective Weather Grades (QX/T 336-2016). Beijing: China Standards Press, 2016.

[4] Palmer T. A Vision for Numerical Weather Prediction in 2030 [C]||Proceedings of the American Geophysical Union Fall Meeting, 2020: A196-01.

[5] Yan Q, Ji F X, Miao K C, et al. Convolutional Residual Attention: A Deep Learning Approach for Precipitation Nowcasting. Advances in Meteorology, 2020(5): 1-12.

[6] REINOSO-RONDINEL R, REMPEL M, SCHULTZE M, et al. Nationwide Radar-Based Precipitation Nowcasting—A Localization Filtering Approach and Its Application for Germany. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15: 1670-1691.

[7] BROWNING K A. Review Lecture: Local Weather Forecasting. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 1980, 371(1745): 179-211.

[8] TIAN Jiyang. Land-Atmosphere Coupled Hydrological Forecasting Supported by Multi-Source Weather Radar Data Assimilation. Beijing: China Institute of Water Resources and Hydropower Research, 2017.

[9] CAO Weihua, CHEN Mingxuan, GAO Feng, et al. Study on Nowcasting by Fusing Thunderstorm Area Tracking Vectors and Thunderstorm Cell Tracking Vectors. Acta Meteorologica Sinica, 2019, 77(6): 1015-1027.

[10] CAO Chunyan, CHEN Yuanzhao, LIU Donghua, et al. Optical Flow Method and Its Application in Nowcasting. Acta Meteorologica Sinica, 2015, 73(3): 471-480.

[11] Zhang J, Howard K, Langston C, et al. Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. Bulletin of the American Meteorological Society, 2016, 97(4): 621-638.

[12] VEILLETTE M, SAMSI S, MATTIOLI C. SEVIR: A Storm Event Imagery Dataset for Deep Learning Applications in Radar and Satellite Meteorology [J]. Advances in Neural Information Processing Systems, 2020, 33: 22009–22019.

[13] EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS. ERA5 Fifth Generation Reanalysis Data. Copernicus Climate Change Service (C3S). [2024-01-15].

[14] National Meteorological Information Center, China Meteorological Administration. CMA Global Atmospheric Reanalysis Dataset. Beijing: National Meteorological Information Center, 2024.