Lightweight AI-powered precipitation nowcasting

Nan Yang[1] and Xiaofeng Li[1,*]

[1]Key Laboratory of Ocean Circulation and Waves, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China

*Correspondence: lixf@qdio.ac.cn

Received: September 19, 2023; Accepted: April 11, 2024; Published Online: April 18, 2024; https://doi.org/10.59717/j.xinn-geo.2024.100066

Citation: Yang N. and Li X. (2024). Lightweight AI-powered precipitation nowcasting. The Innovation Geoscience 2(2): 100066.

Dear Editors,

Recently, we have witnessed a surge in dangerous incidents, from the severe precipitation in Central China in summer 2023 to the extensive mountain fires now affecting the U.S. and Canada. Such events lead to substantial human and property damage, underscoring the critical need for accurate forecasting. Latest research in Nature (2023)[1] reveals that human activity and greenhouse warming will lead Hazardous Meteorological Events (HME) to become more and more frequent across the globe.

Conventional weather forecasting typically relies on operational Numerical Weather Prediction (NWP) models, such as NOAA's Global Forecast System (GFS) and the European Centre for Medium-Range Weather Forecasts (ECMWF). Recent advancements have seen the emergence of artificial intelligence (AI)-driven large-scale meteorological models.[2,3] These global models have shown notable success in synoptic-scale forecasting of atmospheric conditions, but their effectiveness in forecasting HME remains limited due to the complexities of these phenomena and the relatively coarse resolution of the models. For more localized forecasting of small to meso-scale meteorological events, high-resolution community models like the Weather Research and Forecasting Model (WRF) are commonly employed, using initial and boundary conditions provided by the NWP models. Simulating these events with WRF is not a trivial task and typically requires a nested grid configuration, various atmospheric boundary layer and microphysics schemes, and accurate local topographical data, often necessitating supercomputing capabilities and specialized numerical modeling expertise. In response to the global need for more urgent, fast, and accurate forecasts for HME, one of the most dynamically researched areas has emerged, centering on short-term (0-3 hours) HME predictions through lightweight AI technology— the subject of this letter.

Lightweight AI Short-term forecasting is made possible through two key mechanisms: (1) In the realm of Earth observation, the growing availability of multi-source satellite data offers image sequences of HMEs, as shown in Figure 1A; (2) Within the AI development community, continuous progress is being made in the area of lightweight forecasting systems that utilize image-to-image methodologies.[4,5,6,7] The technology transforms short-term forecasting into a video prediction task, simplifying the intricate inferring inherent in NWP or WRF models' complex physical processes. For example, DeepMind[8] has created the Deep Generative Model of Radar (DGMR), a system that uses sequences of ground-based radar images for precipitation nowcasting. However, DGMR relies solely on a single-channel image sequence and employs an isolated factor to estimate the probability of short-term forecasting for future time steps. Contrarily, environmental satellites offer multi-channel data with global coverage. Yet, the presence of information overlap and autocorrelation among these channels complicates the feature extraction process, ultimately reducing the efficacy of forecasting. Consequently, a question arises: how to effectively utilize multiple environmental factors and understand their influence on forecasting outcomes? It remains a significant challenge in lightweight AI-powered short-term forecasting.

Preliminary: We first define 2 key terms, the terminology refers to "environmental factors" as "raw data from satellite observations" and "channels" as "feature channels" that represent diverse features extracted by different convolution kernels in the neural network.

The core ideas of this letter are threefold: (1) We propose a novel, lightweight, and adaptable Front-End Processor (FEP) design for dynamic, multi-channel information extraction, as shown in Figure 1A. The goal is to leverage and analyze multiple environmental factors to achieve accurate forecasting; (2) Using satellite precipitation image sequences as an example, we conduct design experiments with various combinations of channel inputs to identify key measured environmental factors and optimize the use of multi-channel data; (3) Highly accurate precipitation forecasting achieved using the factors with the highest contribution identified in steps (1) and (2), and a Back-End Processor (BEP), as shown in Figure 1A.

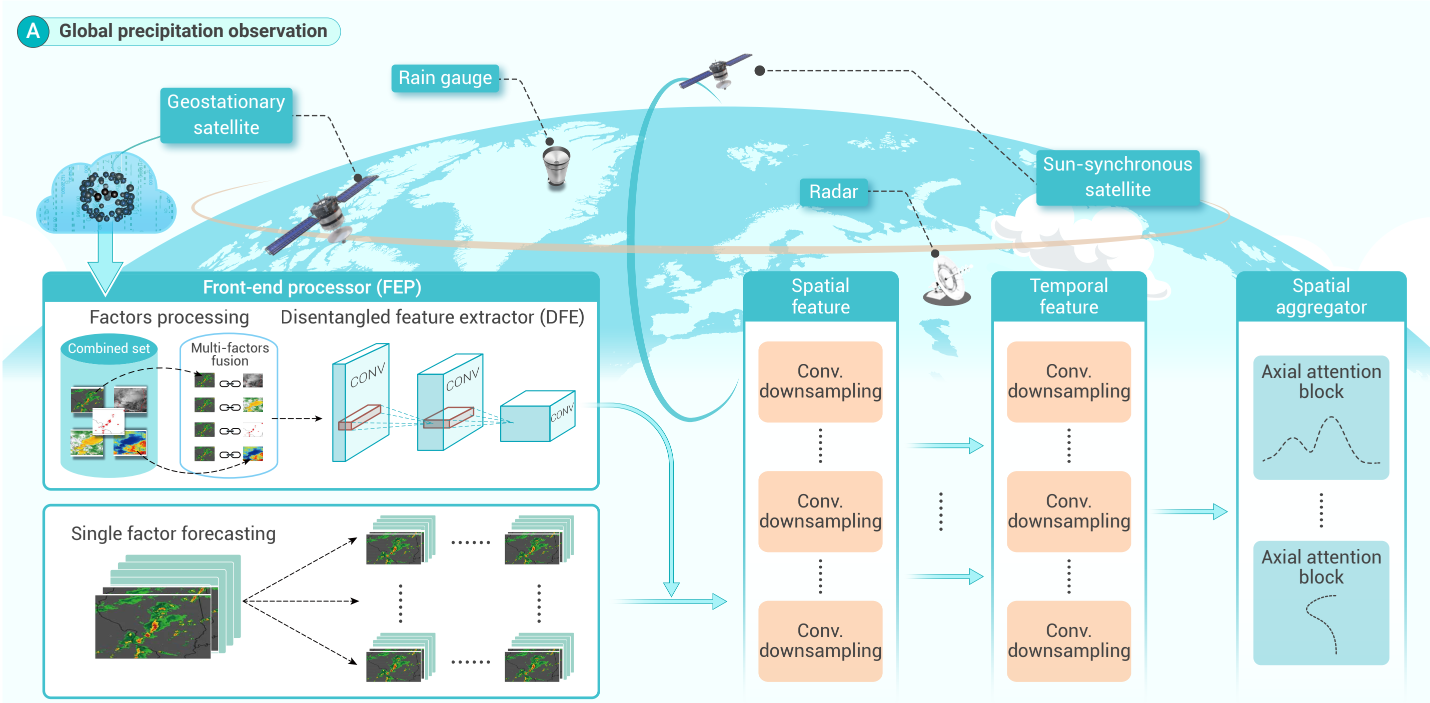
Novel FEP Conceptual Design (Figure 1A): The proposed FEP concept adaptively receives factors of different scales and achieves lightweight AI-powered forecasting with multi-factor analysis. FEP consists of convolutional layers that can be seen as multiple interfaces to extract diverse feature channels, enabling dynamically receiving images with varying spatial resolutions, and taking multiple environmental factors into image sequences, such as surface temperature, vegetation cover, precipitation, and aerosol concentration. These sequences are then processed into multiple combinations that feed into the Disentangled Feature Extractor (DFE), accomplishing the feature learning of multiple precipitation factors, and integrating them into a combined feature map. The feature map is then fed into the single-factor forecasting module. The module encompasses components such as the convolutional and down-sampling layers to extract temporal and spatial features. The input is then directed to the spatial aggregation module to generate the forecasting image for the subsequent N time steps.

The FEP is portable and allows seamless integration of observed factors into short-term forecasting algorithms with minimal fine-tuning. FEP simplifies the traditional physical process modeling and improves the inferring efficiency of short-term forecasting.

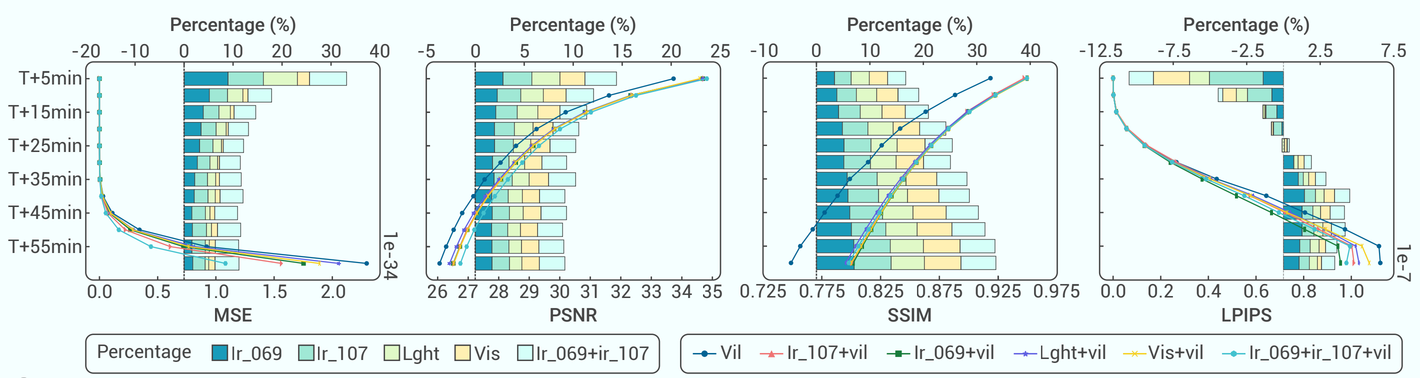
Dominant Factor Analysis (Figure 1B): We provide an example to verify the FEP-based forecasting model to extract and integrate precipitation features of various scales using Storm Event Imagery (SEVIR).[9] SEVIR is a collection of temporally and spatially aligned image sequences depicting weather events captured over the contiguous US (CONUS) by GOES-16 satellite and the mosaic of NEXRAD radars. SEVIR contains five image types: GOES-16 0.6 μm visible satellite channel (VIS), 6.9 μm and 10.7 μm infrared channels (Ir_069; Ir_107), a radar mosaic of vertically integrated liquid (VIL) derived by NEXRAD, and total lightning flashes collected by the GOES-16 geostationary lightning mapper (GLM) (Lght). Precipitation forecasting with a single factor as baseline, we use 13 VIL images with a spatial resolution of 384 × 384 km, sampling at 5-minute intervals, as input to the forecasting model. The model is trained to generate VIL images for the next 12 time steps, corresponding to the precipitation distribution in the subsequent hour. Therefore, the next-generation weather radar (NEXRAD), recognized as a standard product rigorously tested for precipitation measurement, is employed as the ground truth value. Employing a larger image sequence time step, such as 10 minutes, allows us to extend our forecasting range to 2 hours.

The analysis includes six sets of experiments: Single-factor analysis using VIL. 2-5) Two-factor analyses pairing VIL with each of Ir_069, Ir_107, Lght, and VIS. Three-factor analysis combining Ir_069, Ir_107, and VIL. The forecasting accuracy is evaluated using 4 metrics:[10] Mean Square Error (MSE), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS). Among these metrics, MSE quantifies the error between the forecasting and the ground truth, while the other three metrics are employed to assess the quality of the output images. Here, we use the increments of these metrics to gauge the improvement or decline in accuracy, defining this change as a "contribution" (e.g., ΔMSE).

How do we analyze factors with the highest contribution? We follow the following steps: (1) For the precipitation forecasting task, we initially train a radar-to-radar forecasting model, i.e., a pre-trained back-end, recording the accuracy of each metric as a baseline. (2) Establish binary groups of different environmental factors and radar precipitation observation images. (3) Fix the network parameters of the pre-trained back-end and train the network



MSE (1e-4)							PSNR					
	Vil	Ir107_vil	Ir069_vi	Lght_vil	Vis_vil	Ir069_ir107_vil	Vil	Ir107_vil	Ir069_vi	Lght_vil	Vis_vil	Ir069_ir107_vil
T+5min	7.10	6.46	6.59	6.60	6.92	6.56	33.73	34.68	34.72	34.69	34.59	34.82
T+15min	19.07	18.32	18.46	18.63	18.92	18.22	30.19	30.83	30.85	30.83	30.87	31.02
T+25min	29.70	28.76	28.90	29.18	29.60	28.47	28.57	29.13	29.11	29.07	29.12	29.32
T+35min	38.68	37.83	37.67	38.03	38.39	36.95	27.54	28.07	28.05	28.01	28.08	28.30
T+45min	46.73	45.99	45.44	46.29	46.28	44.55	26.80	27.27	27.27	27.18	27.29	27.50
T+55min	53.82	52.95	52.35	53.38	53.16	51.25	26.28	26.73	26.73	26.62	26.74	26.95
SSIM							LPIPS					
T+5min	0.918	0.950	0.947	0.950	0.950	0.949	0.168	0.170	0.174	0.170	0.172	0.171
T+15min	0.863	0.899	0.898	0.899	0.903	0.901	0.256	0.257	0.258	0.256	0.256	0.256
T+25min	0.826	0.869	0.868	0.866	0.869	0.868	0.299	0.298	0.300	0.298	0.298	0.298
T+35min	0.799	0.845	0.844	0.842	0.845	0.844	0.323	0.320	0.322	0.322	0.322	0.321
T+45min	0.777	0.826	0.825	0.822	0.826	0.824	0.337	0.332	0.334	0.335	0.334	0.333
T+55min	0.757	0.810	0.808	0.804	0.809	0.806	0.344	0.340	0.342	0.342	0.342	0.341



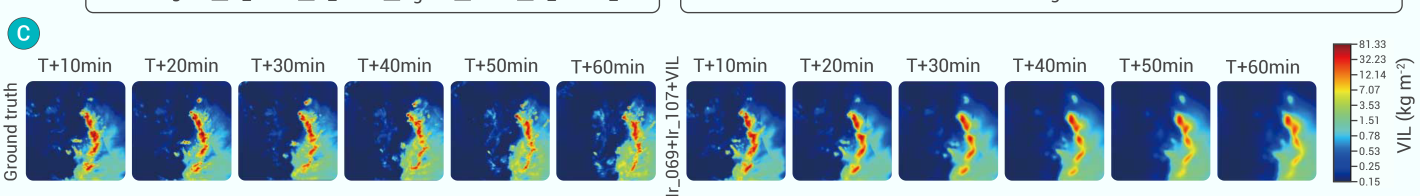


Figure 1. Current global precipitation observation tools and the proposed conceptual design, and quantitative and qualitative experimental results (A) Global precipitation observations, including geostationary, sun-synchronous satellites, weather radar, rain gauges, and meteorological observation stations, with their corresponding observational results. The developed Front-End Processor (FEP) conceptual design and its corresponding working pipeline, including multi-factor fusion and forecasting with a single factor. (B) The accuracy evaluation results of 4 metrics, including MSE, PSNR, SSIM, and LPIPS for 12 future time steps, where red and green indicate the winner and runner-up, respectively. Note that the source input data has a time interval of 5 minutes. The displayed metrics and images in the figure have a time interval of 10 minutes, so only 6 metrics and images are shown. The comparative experimental results on 4 metrics are shown with a line chart, denoting the accuracy and image quality at each time step. The stacked bar chart represents the percentage improvement at each future time step for the multi-input experiment compared to the single-input experiment. (C) High-accuracy precipitation forecasting achieved using the factors with the highest contribution identified by FEP.

parameters of the Front-End Processor(FEP). After FEP convergences, record the accuracy of each metric. (4) Repeat the training with different tuples, calculating increments in accuracy for different metrics under different tuple combinations. (5) Determine the contribution of different environmental factors through incremental comparisons.

Figure 1B comprehensively evaluates six different factor combinations using four evaluation metrics for future forecasting. Identified as the channels with the highest contribution through FEP, Experiment 6 resulted in the most significant accuracy improvement. We plot the stacked bar chart and the line graph to quantify the percentage increase of four factors in each time step (Figure 1B). The results show that the factors with the highest contribution, i.e., Ir_069 and Ir_107 with VIL, can improve the accuracy of precipitation forecasting across all time steps, e.g., PSNR, MSE, and SSIM, the average increases reached 2.7%, 4.8%, and 5.3%, respectively. With the increase of the forecasting time step, the contribution of multiple environmental factors to LPIPS also gradually increases, consistent with the experimental results in Figure 1B. These results underscore the versatility, accuracy, and robustness of our developed FEP, with its benefits in multi-factor forecasting becoming increasingly evident over time. From the analysis, we draw three important conclusions: (1) Every channel contributes, but the degree of their contributions is different, as seen in metrics like SSIM increase of 5.42% in Experiment 2. (2) Although distinct metrics reflect diverse evaluation criteria, the contribution ranking of each channel on different metrics is consistent through comprehensive analysis. (3) More input channels are not always better, such as channel Lght in Experiment 4, which only resulted in a negligible 0.00005 improvement in the MSE.

Analysis suggests that it is reasonable and effective for researchers to input multiple factors when conducting similar precipitation forecast experiments. FEP within the portability, lightweight, and generalization can easily analyze the ranking of different environmental factors' contributions please note that there is relative to large models with complex parameters and numerical models containing multiple physical mechanisms), screen out valuable channel data, reveal the interaction between the multiple factors, solve the feature entanglement problem in forecasting, and help researchers integrate lightweight AI with HME to utilize multi-channel data's effectiveness fully.

Accurate Short-Term Forecasting (Figure 1C): The figure showcases the high-accuracy precipitation forecasting results obtained by utilizing the factors with the highest contribution determined by FEP's automatic multi-factor analysis technology. BEP utilized for precipitation forecasting is the Simple Video Predictive (SimVP) model. SimVP is a Convolutional Neural Network (CNN)-based architecture designed for encoding, translating, and decoding features to accomplish efficient sequence-to-sequence forecasting.

As illustrated, the forecasting results demonstrate notable accuracy in the forthcoming 15-20 minutes, with SSIM values surpassing 0.9. Results emphasize FEP's proficiency in generating high-quality forecasting imagery. Furthermore, the deep generative model DGMR,[4] proposed in Nature, also notes that the physics-based model (PySTEPS) tends to overestimate rainfall intensity over time, deviating from realistic observations and inadequately covering the spatial extent of rainfall. Despite slightly decreasing model accuracy as the forecasting lead time extends, the SSIM metric remains above 0.8 at T+55 minutes. Notably, even in experiments employing individual input factors, the SSIM surpasses 0.75. FEP+BEP (SimVP) short-term precipitation forecasting model achieves remarkable accuracy for immediate forecasts and sustains reliability for extended periods.

CONCLUSIONS

FEP has brought essential technological innovation to short-term forecasting, providing an effective solution for application scenarios requiring an immediate response. Accurate short-term forecasting can be accomplished by aggregating various environmental factors and analyzing the applicative worth of data from multiple channels. With continuous development and improvement of FEP, AI-powered models will continue to play a revolutionary role in HME forecasting.

Overall, the developed conceptual design can be applied to any image-to-image forecasting model to dynamically adapt to multiple input factors, breaking the limitation of traditional AI models that only analyze one factor. The FEP-based short-term forecasting model greatly simplifies the physical process and plays into the effectiveness of multi-channel data and lightweight AI-based models. Experimental analysis confirms the effectiveness of multi-factor input in improving forecasting accuracy and revealing the complex relationship between environmental factors. With the injection of more advanced lightweight AI-based models and multi-channel data, the research results will further enhance the accuracy and reliability of weather forecasting, leading to a far-reaching impact on HME forecasting.

REFERENCES

1. Ham, Y.-G., Kim, J.-H., Min, S.-K., et al. (2023). Anthropogenic fingerprints in daily precipitation revealed by deep learning. Nature 622(7982): 301–307. DOI: 10.1038/s41586-023-06474-x.
2. Bi, K., Xie, L., Zhang, H., et al. (2023). Accurate medium-range global weather forecasting with 3D neural networ ks. Nature 619(7970): 533–538. DOI: 10.1038/s41586-023-06185-3.
3. Wu, H., Zhou, H., Long, M., et al. (2023). Interpretable weather forecasting for worldwide stations with a unifie d deep model. Nature Machine Intelligence 5(6): 602–611. DOI: 10.1038/s42256-023-00667-9.
4. Zhang, Y., Long, M., Chen, K., et al. (2023). Skilful nowcasting of extreme precipitation with NowcastNet. Nature 619(7970): 526–532. DOI: 10.1038/s41586-023-06184-4.
5. Espeholt, L., Agrawal, S., Sønderby, C., et al. (2022). Deep learning for twelve hour precipitation forecasts. Nature communications 13(1): 1–10. DOI: 10.1038/s41467-022-32483-x.
6. Ham, Y.-G., Kim, J.-H., and Luo, J.-J.. (2019). Deep learning for multi-year ENSO forecasts. Nature 573(7775): 568–572. DOI: 10.1038/s41586-019-1559-7.
7. Pang, B., Zha, K., Cao, H., et al. (2020). Complex sequential understanding through the awareness of spatial and temporal concepts. Nature Machine Intelligence 2(5): 245–253. DOI: 10.1038/s42256-020-0168-3.
8. Ravuri, S., Lenc, K., Willson, M., et al. (2021). Skilful precipitation nowcasting using deep generative models of radar. Nature 597(7878): 672–677. DOI: 10.1038/s41586-021-03854-z.
9. Veillette, M., Samsi, S., and Mattioli, C.. (2020). Sevir: A storm event imagery dataset for deep learning applications in radar and satellite meteorology. Advances in Neural Information Processing Systems 33: 22009–22019. DOI: 10.5555/3495724.3497570.
10. Zhang, R., Isola, P., Efros, A.A., et al. (2018). The unreasonable effectiveness of deep features as a perceptual metric. Proceedings of the IEEE conference on computer vision and pattern recognition . DOI: 10.1109/CVPR.2018.00068.

FUNDING AND ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (42306214), Shandong Province Postdoctoral Innovative Talents Support Program (SDBX2022026), China Postdoctoral Science Foundation(2023M733533), Special Research Assistant Project of the Chinese Academy of Sciences in 2022.

DECLARATION OF INTERESTS

The authors declare no competing interests.