Reply to the Reviewers

1. The claimed novelty of "dual encoder-decoder training framework" needs stronger differentiation from existing dual-learning architectures in literature.

Thank you for your valuable comment. The core innovation of our method is the use of the Spatio-Temporal Information (STI) equation to address the issue of low accuracy in short-term precipitation forecasting with short-term data. We employ dual learning to implement the STI equation, rather than considering dual learning as the primary innovation. In fact, our dual learning framework is fundamentally derived from the inherent conjugate duality of the STI equation itself. Compared with other methods, STI-DEDN, guided by theoretical principles, can solve the conjugate dual components of the STI system simultaneously. As far as we know, we have not found any other literature on short-term precipitation forecasting using the Spatio-Temporal Information equation.

We have revised the manuscript, and the specific changes are as follows:

- **Abstract, Page 1:** the present work proposes a dual encoder-decoder training framework based on the STI equation and the idea of dual learning, which can map multidimensional spatial features to the temporal prediction of future precipitation variables.
- **Introduction, Page 4:** Aiming at the prediction inaccuracies caused by existing STI equation solvers and the limitations of short-term data in forecasting short-duration heavy precipitation, we have established a novel dual encoder-decoder neural network that enables precise computation of STI equations and robust nonlinear mapping between high-dimensional meteorological and precipitation variables.

2. A 60-minute forecast lead time is far from sufficient and fails to effectively highlight the model's performance. In operational precipitation nowcasting, the focus is typically on 0–3 hour forecasts. While the authors present 2h and 3h results in Section 5.3, the model's performance in MSE and PSNR metrics is inferior to Transformer-based models and only comparable to RNN-based approaches (e.g.,

PredRNN and PhyDNet). This appears somewhat contradictory to the claim that the proposed model mitigates error accumulation in long-term forecasting. The ADGLoss also seems to have limited effectiveness in improving longer lead-time predictions. To strengthen the analysis, the authors should: 1) Include visualizations of typical cases for 2h and 3h forecasts to illustrate the performance limitations. 2) In the conclusion, propose future research directions to enhance the model's effectiveness for extended forecasting periods.

Thank you for your critical and insightful comments, which undoubtedly strengthen our experimental analysis. We fully agree with your statement that a 60-minute lead time for forecasting extreme precipitation from severe convection is insufficient and that the performance of long-term predictions needs further investigation. However, the focus of our model is not on extreme precipitation. Our primary research direction is short-term precipitation forecasting within a 60-minute lead time [1][2][3], which is a popular direction in the field of short-term precipitation forecasting in recent years [4] and is crucial for operational meteorology, including emergency response, agricultural planning, and traffic management. Timely information can lead to better decision-making and risk mitigation [5][7][8][9]. The performance of forecasts beyond two hours remains a highly challenging research area that requires further study [4]. Additionally, we conducted experimental analyses for 2-hour and 3-hour forecasts and for heavy precipitation in our experiments, to address these two aspects in the future.

As you requested, we will continue to supplement our experiments and have added a visualization analysis and discussion of 2-hour and 3-hour prediction cases in Section 5.3 of the revised manuscript, along with an explanation of the specific limitations of our model. Our model did not achieve the best performance in MSE/PSNR, indicating that Transformer-based models show higher clarity in image pixel processing. However, our task is precipitation forecasting, not purely in the image domain. Therefore, we focus more on the meteorological indicators in precipitation forecasting. Our method achieved the best performance in meteorological indicators (CSI and HSS), which means that our method is more accurate in the physical coherence and structure of precipitation distribution, as can be seen in Figure 1. The performance of ADGLoss is also similar in this regard, as we focus more on meteorological indicators. Moreover, to better summarize the innovations of this paper, we have revised the performance expression of ADGLoss to "Our proposed method alleviates the error amplification caused by the extension of prediction time." which can be proven in our loss function ablation experiments.

Finally, in response to your request, we have added a dedicated paragraph in the conclusion to outline future research directions, acknowledging that our model has shortcomings in long-term prediction. We will explore new methods, possibly combining the structure of STI-DEDN with Transformer-based long-range dependency models to improve the effectiveness of our method in long-term heavy precipitation forecasting.

We have revised the manuscript, and the specific changes are as follows:

- **Abstract, Page 1:** due to the rapid pace of climate change, long-term time series data are often inadequate for accurately addressing precipitation forecasting for extreme weather events in a short period of time, as past meteorological time series data may not accurately reflect current atmospheric conditions.
- **Abstract, Page 1:** Additionally, an adaptive weighted gradient loss (ADGLoss) is proposed to mitigate the error amplification caused by the extension of prediction time and rectify systematic underestimation of high-intensity precipitation regions.
- **1 Introduction, Page 2:** The rapid evolution of observational technologies has driven a surge in demand for short-term precipitation nowcasting, particularly for operational applications requiring precise forecasting in time windows from 30 minutes to 1-h. This is a popular direction in the field of short-term precipitation forecasting in recent years and is crucial for operational meteorology, including emergency response, agricultural planning, and traffic management. Timely information can lead to better decision-making and risk mitigation. The performance of forecasts beyond two hours remains a highly challenging research area that requires further study.
- **1 Introduction, Page 4:** This loss function dynamically adjusts error gradients based on rainfall magnitude, effectively mitigating the error amplification issue that models face as prediction time extends, while preserving the fine-scale precipitation structures.
- **1 Introduction, Page 4:** To improve the issue of precipitation underestimation

in heavy rainfall areas and the problem of increasing error in existing precipitation models as prediction time extends, we propose an adaptive weighted gradient loss function that can effectively enhance the model's accuracy in spatiotemporal prediction tasks.

- **3.3 Loss Function, Page 10:** which strengthens the GDL loss's ability to capture high-frequency precipitation patterns globally and simultaneously refines prediction accuracy for heavy precipitation areas, thereby reducing the error amplification issue in precipitation forecasting tasks.
- **5.3 Results of Long-Term Forecasting, Page 21:** Table 5 and Fig. 9 presents the forecasting results of our proposed method alongside other methods on the SEVIR dataset for the three-hour prediction interval.
- **5.3 Results of Long-Term Forecasting, Page 21:** In contrast, the Transformer-based model achieved the highest MSE and PSNR values, which corresponded to the clearest image resolution obtained by Earthformer in Fig. 9. However, the task we conducted was weather precipitation forecasting, which does not fully align with the objectives of image processing. We place greater emphasis on meteorological indicators.

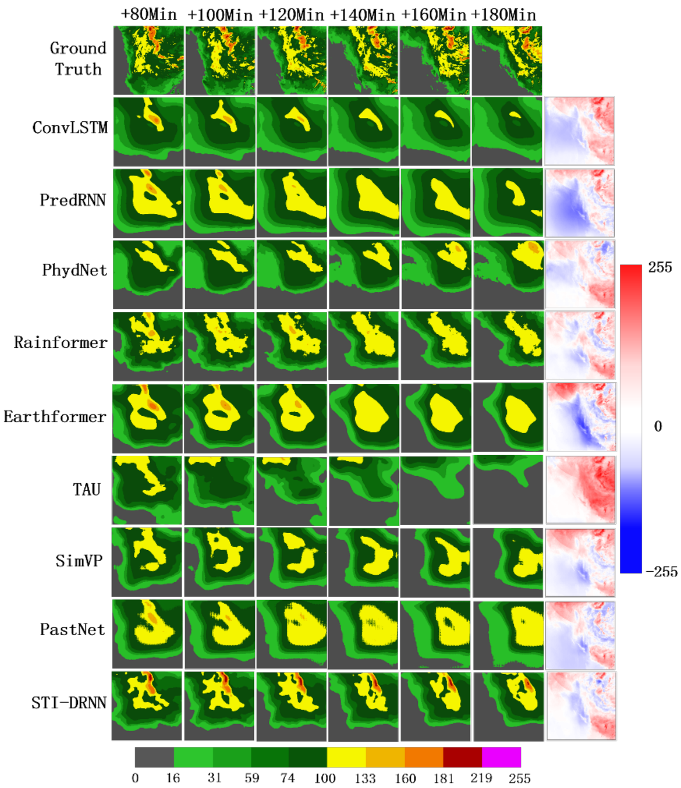

Figure. 1 Visualization of precipitation maps and error heatmaps for all methods in the long-term experiment of Case 3.

In summary, the Transformer-based model performed better in terms of producing smoother radar echo images in forecasting; however, it was not as effective as STI-DEDN in predicting the distribution and structure of precipitation areas, with STI-DEDN being able to predict relatively larger local precipitation amounts. The inherent limitations of RNN-based methods restrict STI-DEDN's ability to capture long-term dependencies, while the attention mechanism can effectively address long-term dependency issues at the pixel level of images. These points pave the way for our future research areas.

- **6 Conclusions, Page 27:** Future work will explore incorporating richer meteorological variables and addressing data imbalance issues and long-term forecasting error accumulation through improved augmentation strategies and effective attention mechanisms.

3. The computational requirements (Section 4.3) are underspecified. The authors should provide training time and inference speed.

We have added the computational requirements into Section 4.3 in response to the suggestion.

We have revised the manuscript, and the specific changes are as follows:

- **4.3 Environment Setup, Page 14:** The training batch size is set to 2. On a single NVIDIA GeForce RTX 3090 (24GB) GPU card, the average training time per epoch is about 2.55 hours, and the total training time for 100 epochs is about 255.55 hours. For inference, the model takes approximately 0.54 seconds to process a radar sequence (18 frames), corresponding to an inference speed of about 34 frames per second.

4. The exclusion of lightning data (Section 4.1) requires more rigorous justification given its potential relevance to extreme precipitation events. The authors should discuss how this exclusion might impact model performance.

Thank you for your valuable comment. Lightning does exhibit a certain correlation with convective activity and may, in some cases, help identify extreme precipitation events. To make it more clear, we provide a more detailed explanation of the reasons for not incorporating lightning data in this study and suggest potential directions for improvement in future research. We will ensure that these aspects are adequately addressed in the revised manuscript.

First, in the SEVIR dataset, lightning is stored as point events (with precise spatiotemporal coordinates), while other modalities such as infrared, visible light, and VIL are represented as gridded matrices. Converting these sparse point events into a gridded format requires additional interpolation and aggregation methods, which inevitably introduce more artificial patterns and noise. It will deviate from the true physical processes. This spatiotemporal mismatch will interfere with the model learning and reduce the prediction performance. Second, lightning is a byproduct of strong convective activity, however, strong convective activity does not necessarily accompany lightning events [6]. Therefore, in current short-term precipitation forecasting research, lightning is rarely used as a core predictive input. Additionally, the contribution of lightning data to precipitation forecasting is very low [7].

We have revised the manuscript, and the specific changes are as follows:

- **4.1 Dataset, Page 11:** Lightning data is stored as site events (precise spatiotemporal coordinates), while the other modalities (IR, VIS, VIL) are all rasterized images. Converting sparse point events into raster format requires additional interpolation and aggregation methods, which inevitably introduce artificial patterns and noise, deviating from the true physical process. This spatiotemporal mismatch can interfere with model learning and reduce the reliability of predictions. Lightning is a byproduct of strong convective activity, and strong convective activity does not necessarily accompany lightning events. Moreover, the contribution of lightning data to precipitation forecasting is very low.

5. The authors utilized bicubic interpolation to standardize the spatial resolution of various meteorological factors. While this approach resolves the resolution mismatch, please consider and discuss the following concerns: 1. Could this interpolation method introduce artificial errors into the dataset? 2. Might this processing lead to misalignment in the spatial representation of precipitation patterns

among the different factors?

We sincerely appreciate the reviewer's insightful comments regarding our use of bicubic interpolation in the preprocessing stage. In the revised manuscript, there is a dedicated paragraph to discuss this issue. First, we chose bicubic interpolation to standardize the spatial resolution for the balance between computational efficiency and output quality. Compared to nearest-neighbor or other traditional methods, bicubic interpolation produces smoother results , which is beneficial for maintaining the continuity of meteorological fields. Any interpolation method may introduce slight artifacts or smooth out fine-scale features, but the original SEVIR dataset has already been spatially and temporally aligned. Our interpolation process only alters the resolution and does not affect the spatial patterns among different variables. Furthermore, we have conducted additional experiments where we retrained and tested the model using the original resolution data without interpolation. There is virtually no difference, as summarized in the table below. This essentially confirms that our use of bicubic interpolation will not introduce significant artificial errors and spatial misalignment.

Table. 1 Comparison of results with and without interpolation for STI-DEDN on the SEVIR test set.

Method	MSE	SSIM	PSNR	FAR	CSI	HSS
Interpolation	2.3794	0.9567	46.5231	0.4686	0.4410	0.5595
Original resolution	2.3947	0.9569	46.507	0.4781	0.4388	0.5579

We have revised the manuscript, and the specific changes are as follows:

- **4.1 Dataset, Page 12:** Although bicubic interpolation may introduce slight artifacts or smooth fine-scale features in the data, the original SEVIR dataset has already been aligned in spatial location and time. Our interpolation only changes the pixel resolution and does not affect the spatial patterns between different variables. Moreover, bicubic interpolation can balance computational efficiency, producing smoother results than nearest-neighbor or bilinear methods, which helps maintain the continuity of meteorological fields.

Additionally, we conducted extra experiments to retrain and test the model without interpolation at the original resolution. As shown in Table 2, there was almost no difference.

6. The 0-3-hour nowcasting is aimed at sudden severe convective or heavy rainfall processes. Therefore, the first case study in this paper has essentially no practical significance and is insufficient to demonstrate the model performance. The selection of the second case is reasonable; however, in terms of the performance in this localized heavy precipitation case, STI-DEDN is inferior to operational models such as NowcastNet (Sheng et al. 2025). Such results are insufficient to support the publication of this paper in GMD.

Thank you for your thoughtful and constructive feedback on our manuscript. We need to clarify our work. Our method is not designed for long-term forecasting of heavy precipitation events, but rather focuses on short-term precipitation forecasting within a one-hour time frame. This has become a prominent focus in the field of short-term precipitation forecasting [4], and is crucial for meteorological applications such as emergency response, agricultural planning, and traffic management [5][7][8][9]. In our experiments, we used our model to achieve 2-3 hour short-term precipitation forecasts through recursive reasoning. The prediction performance did not achieve the best scores in MSE and PSNR due to accumulated errors, but the precipitation indicators still outperformed mainstream models. In the future, we will also gradually improve the prediction accuracy for longer lead times.

Regarding the first case study in our experiments, we believe it serves as a baseline for evaluating our model's forecasting performance under typical precipitation conditions. This is essential for establishing the reliability of our method in everyday scenarios. The second case study involves localized heavy precipitation, which highlights our model's capability in predicting severe weather events, even if its performance may not surpass that of larger models like NowcastNet [10]. We have studied the NowcastNet, which has made significant progress in the 0-3 hour forecasting. And we must emphasize the difference between our approach and this approach. Our model is a standard deep neural network. It is impractical to make direct comparisons between it and large model. Additionally, our primary focus is on 0-1 hour forecasting tasks, while their work primarily addresses 0-3 hour predictions. The training datasets used in both approaches also differ significantly, which can have a substantial impact on model performance. Thank you for your help to improve our work.

We have revised the manuscript, and the specific changes are as follows:

- **Abstract, Page 1:** due to the rapid pace of climate change, long-term time series data are often inadequate for accurately addressing precipitation forecasting for extreme weather events in a short period of time, as past meteorological time series data may not accurately reflect current atmospheric conditions.
- **1 Introduction, Page 2:** The rapid evolution of observational technologies has driven a surge in demand for short-term precipitation nowcasting, particularly for operational applications requiring precise forecasting in time windows from 30 minutes to 1-h. This is a popular direction in the field of short-term precipitation forecasting in recent years and is crucial for operational meteorology, including emergency response, agricultural planning, and traffic management. Timely information can lead to better decision-making and risk mitigation. The performance of forecasts beyond two hours remains a highly challenging research area that requires further study.
- **5.2.1 Case 1, Page 17:** We believe it serves as a baseline for evaluating our model's forecasting performance under typical precipitation conditions. This is essential for establishing the reliability of our method in everyday scenarios.
- **5.2.2 Case 2, Page 19:** The second case involves a localized heavy precipitation event, which highlights our model's capability in predicting severe weather events.

Finally, Thank you for your valuable feedback on our work. We will incorporate all the suggested modifications into our revised manuscript to better clarify our research focus and contributions. Your insights are greatly appreciated, and we are committed to enhancing the quality of our paper based on your recommendations. Once again, thank you for your constructive comments.

References

[1]. Ji C, Xu Y, Ke W, et al. PredAt-GAN: a new spatiotemporal forecast model for precipitation nowcasting with weather radar echo images[J]. Signal, Image and Video Processing, 2025, 19(1): 5.

[2]. Zhao Z, Dong X, Wang Y, et al. Advancing realistic precipitation nowcasting with a spatiotemporal transformer-based denoising diffusion model[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1-15.

[3]. Wu H, Liang Y, Xiong W, et al. Earthfarsser: Versatile spatio-temporal dynamical systems modeling in one model[C]//Proceedings of the AAAI conference on artificial intelligence. 2024, 38(14): 15906-15914.

[4]. An S, Oh T J, Sohn E, et al. Deep learning for precipitation nowcasting: A survey from the perspective of time series forecasting[J]. Expert Systems with Applications, 2025, 268: 126301.

[5]. ZHAO Q, ZHENG Y, JING Y, et al. Research Progress of Short-Duration Heavy Precipitation in China[J]. Advances in Earth Science, 2025: 1.

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

[7]. MA Z, ZHANG H, LIU J. A survey of precipitation nowcasting based on deep learning[J]. Computer Engineering & Science, 2023, 45(10): 1731.

[8]. Wei F, Lin P, Nan W, et al. Survey on the application of artificial intelligence in precipitation nowcasting[J]. Nanjing Xinxi Gongcheng Daxue Xuebao, 2020, 12(4): 406-420.

[9]. Ting W, YunPeng L, Chen D, et al. A survey of short-term impending precipitation forecasting methods and their applications[J]. Electronic World, 2019 (10): 11-13.

[10]. SHENG Jie, JIN Ronghua, ZHANG Xiaowen, DAI Kan, ZHANG Xiaoling, GUAN Liang, YANG Bo, ZHANG Yuchen, XING Lanxiang, LONG Mingsheng, WANG Jianmin,2025.Verification and Case Evaluation of the “Fenglei” V1 Meteorological Nowcasting Model[J].Meteor Mon,51(4):389-399.