Response to Reviewer

Manuscript Title: Precipitation Nowcasting Based on Convolutional LSTM with Spatio-Temporal Information Transformation Using Multi-Meteorological Factors

Dear Editor and Reviewer,

We would like to express our deepest gratitude for your time and effort in reviewing our manuscript. We are highly encouraged by your positive assessment of our work, particularly your recognition of our model architecture, loss function, and overall experimental design. Your constructive feedback has been invaluable in helping us refine our narrative, clarify our experimental setups, and strengthen the overall rigor of the paper.

Below, we provide a detailed, point-by-point response to each of your comments, outlining the corresponding modifications made to the revised manuscript.

Part I: Response to General Comments

1. Introduction and Related Work sections: *Reviewer's Comment:* **These sections are somewhat descriptive and read more like a listing of previous studies. The manuscript would benefit from additional synthesis and transition and summary sentences. This would help guide readers more effectively and better motivate the proposed approach.**

- **Response:** We sincerely thank the reviewer for this constructive structural advice. We have revised both the Introduction and Related Work sections. Specifically, we added comprehensive synthesis paragraphs and improved the transitional phrasing between different methodological families. This new structure better summarizes the common limitations of existing models and provides a much clearer, logical progression that directly motivates our proposed approach.

2. Training with precipitation events only: *Reviewer's Comment:* **At line 259, the authors mention that events with no rainfall are excluded from training. This choice requires further discussion ... completely removing them may bias the training distribution...**

- **Response:** We owe you a profound apology for a critical typographical error in our original text. We mistakenly wrote that "we excluded events with no rainfall." In reality, from the original pool of approximately 13,000 SEVIR events, we filtered out events strictly due to missing sensor data or severe gaps in continuous time intervals. We did not filter out no-rain events. The total number of excluded events accounted for only about 2% of the dataset. We utilized 10,000 of the remaining complete events for our experiments. We have corrected this phrasing in Section 4.1 to accurately reflect our data quality control process and to reassure readers that the natural class balance of the operational environment was preserved.

3. Training with interpolated data: *Reviewer's Comment:* **...the stated motivation for interpolation is reduced training complexity, suggesting an efficiency-driven rather than accuracy-driven choice ... If efficiency is the main motivation, the authors could provide a direct comparison of computational cost, such as training time and memory usage, between the interpolated and original-resolution settings.**

- **Response:** We agree with your observation. To directly address the efficiency motivation, we have added a comprehensive comparison of computational costs to the manuscript. This new addition explicitly details the differences in training time and model parameter volume between the interpolated and original-resolution settings, thereby providing concrete evidence for our preprocessing choices.

4. Performance of the proposed model: *Reviewer's Comment:* **The conclusion drawn in lines 324-325 could be too strong ... Do authors think the proposed architecture is generally superior beyond the tested experimental conditions? For operational applications, a fair comparison between different models should allow each model to use its preferred input data, loss function, and training strategy.**

- **Response:** We appreciate your diligence in ensuring academic rigor.
    - First, we have carefully reviewed our manuscript and added necessary qualifiers to all conclusive statements, explicitly bounding our claims (e.g., stating that the model demonstrates "improved CSI for 1-hour VIL nowcasting within the SEVIR dataset").
    - Second, regarding general superiority beyond the tested conditions: theoretically, our architecture is designed to generalize. To validate this, we have proactively conducted an additional experiment using an external, real-world precipitation dataset for training and testing. The preliminary results confirmed the effectiveness of our model. We have included a discussion on this out-of-distribution generalization in the Future Work section. This experiment is not the main experiment of this paper, but only a verification experiment.
    - Finally, regarding fair comparison: to guarantee absolute fairness, all baseline models were trained using their official, optimal parameter settings and recommended training strategies. We made zero modifications to the underlying baseline architectures. Furthermore, both our model and the baselines were fed the exact same input data and evaluated on the same test sets, thereby ensuring a strictly level playing field for deep learning model benchmarking.

Part II: Response to Specific Comments

1. Line 29: *Reviewer's Comment:* **The phrase "... recent observational data more effectively capture precipitation at the next time step..." reads a bit confusing.**

- **Response:** We agree that the phrasing was awkward. We have rephrased this sentence in the revised manuscript to be much clearer and easier to understand.

2. Line 51: *Reviewer's Comment:* **As the acronym SOTA only appears once in the manuscript, the full name should be used instead.**

- **Response:** Corrected. We have replaced the acronym with "state-of-the-art".

3. Line 63: *Reviewer's Comment:* **The full name of LSTM should be given for the first time.**

- **Response:** Corrected. We now provide the full name, Long Short-Term Memory, at its first occurrence.

4. Lines 137-138: *Reviewer's Comment:* **The paper "Predicting future dynamics from short-term time series using an Anticipated Learning Machine" seems relevant to the discussion of using spatio-temporal information.**

- **Response:** We sincerely thank you for recommending this highly relevant literature. The Anticipated Learning Machine (ALM) paper introduces an elegant framework capable of extracting temporal evolution rules directly from short-term observations without relying on traditional numerical integration. This conceptual framework aligns perfectly with our discussion on leveraging high-dimensional spatiotemporal information to predict future dynamics. We have synthesized its core innovations and incorporated this important reference into our Related Work section to enrich the theoretical background of spatiotemporal modeling.

5. Lines 257-258: *Reviewer's Comment:* **Should the contribution of lightning data to precipitation forecasting be case-dependent, for example depending on rainfall type, convective intensity, and the availability of other data types?**

- **Response:** We completely agree with your nuanced perspective. The role of lightning data is indeed highly case-dependent and strongly correlated with intense convective events. We have revised our statement in Section 4.1 to acknowledge this. We clarified that our decision to exclude lightning was primarily due to its sparse point-event format. Aligning point data with our dense, rasterized radar framework causes spatial mismatch issues and introduces interpolation artifacts. This exclusion does not imply that lightning lacks meteorological value. Furthermore, as noted in reference[1] of our manuscript, incorporating lightning data within the SEVIR dataset did not yield a positive impact on VIL forecasting specifically.

6. Line 266: *Reviewer's Comment:* **What is the spatial resolution after interpolation?**

- **Response:** After interpolation, the spatial resolution of the data is 1.5 km. We have explicitly added this detail to the text.

7. Line 321: *Reviewer's Comment:* **I wonder whether the same loss functions were used across all models.**

- **Response:** To ensure a fair comparison, all baseline models were trained using the standard, optimal loss functions recommended in their respective official implementations. Conversely, our STI-DEDN was trained using our proposed ADGLoss. We have clarified this in the text.

8. Line 330: *Reviewer's Comment:* **The word "efficient" usually implies reduced computational cost. However, in this context, the authors seem to be referring mainly to the accuracy or effectiveness of the model in representing uncertainty.**

- **Response:** Thank you for pointing out this lexical inaccuracy. We have changed "efficient" to "effective" to accurately convey our meaning.

9. Figures 5 and 6: *Reviewer's Comment:* **The proposed method appears to be clearly separated from the other methods in terms of forecast skill... Could the authors discuss the main factors contributing to this separation?**

- **Response:** The significant separation in forecast skill is primarily attributable to the synergistic effect of our STI-driven dual encoder-decoder design and the ADGLoss function. Standard baseline models typically suffer from progressively smoothed results and severe error accumulation over longer time steps. In contrast, our spatiotemporal dual structure acts as a robust regularizer. Combined with the adaptive gradient penalty of the ADGLoss, the model strictly enforces spatiotemporal consistency during training. This significantly reduces uncertainty amplification and preserves structural fidelity much better than standard RNN/CNN architectures. We have added a detailed discussion of this phenomenon to the manuscript.

10. Figure 6: *Reviewer's Comment:* **Could the authors explain the jumps in the PastNet results at small spatial scales, shown by the purple lines?**

- **Response:** The abnormal jumps at small spatial scales for PastNet are likely due to its unique physical prior modules and internal spatial resampling steps. These mechanisms can inadvertently inject unphysical high-frequency noise (artifacts) when forecasting highly complex, non-smooth precipitation patterns. It is worth noting that deep learning models inherently possess stochasticity, and some degree of fluctuation is normal when forecasting highly non-linear precipitation, but PastNet's architectural traits make these artifacts more pronounced at small scales. We have added a brief explanatory note regarding this to the text.

11. Line 417: *Reviewer's Comment:* **The phrase "Overall, STI-DEDN is more applicable in large-scale extreme precipitation events" could be confusing.**

- **Response:** We apologize for the semantic confusion. We have corrected the phrasing to accurately reflect that our method is "capable of forecasting extreme precipitation more accurately at larger spatial scales," as you kindly suggested.

12. Figures 5, 7 and 9: *Reviewer's Comment:* **Rainymotion appears to preserve more small-scale texture than other models. Could the authors discuss the reason for this?**

- **Response:** Rainymotion preserves high-frequency, small-scale textures because it is a physics-based optical flow method, not a deep neural network. It essentially advects existing pixel intensities rigidly across the grid based on motion vectors. Consequently, it completely avoids the blurring and smoothing effects that are inherently caused by the convolutional downsampling and upsampling operations utilized in deep learning autoencoders. We have added this valuable insight to our discussion.

13. Section 5.4.2: *Reviewer's Comment:* **The meaning of "removing the STI framework" could be defined more explicitly. Does this refer specifically to removing the dual spatio-temporal and temporal-spatial converter structure?**

- **Response:** Yes, your understanding is exactly correct. We have explicitly clarified this definition in the text. "Removing the STI framework" means downgrading our architecture to a standard, single-stream encoder-decoder network by completely removing the temporal-spatial converter (the inverse mapping constraint) and its associated conjugate training mechanism.

Part III: Response to Technical Corrections

Technical Corrections 1-4: *Reviewer's Comment:* **Missing spaces before "(", lowercase "on" in titles, consistent capitalization, and a duplicated period at Line 261.**

- **Response:** We have meticulously implemented all technical corrections. We added the missing spaces, corrected the capitalization of "on" and other words in titles for consistency, and removed the duplicate period. We deeply appreciate your exceptionally thorough review of our manuscript.

We hope that these revisions fully address your questions and meet your high standards.

Sincerely,

The Authors

[1] Yang, N. and Li, X.: Lightweight AI-powered precipitation nowcasting, The Innovation Geoscience, 2, 100 066, https://doi.org/10.59717/j.xinn-geo.2024.100066, 2024.