Reviewer 1

This manuscript proposes a novel precipitation forecast verification (PFV) method, CLPFV, based on self-supervised contrastive learning. The study addresses a meaningful methodological issue, namely, how to develop a comprehensive verification method that is more tolerant of minor forecast errors, more sensitive to substantial errors, and better able to reflect different degrees of error. The basic idea of using data augmentations (displacement, intensity, and area size), together with an improved contrastive loss function, to train a neural network to learn the gradient of forecast errors is both scientifically sound and methodologically elegant. Overall, I find this manuscript valuable and potentially suitable for publication in GMD after minor revision.

Thank you so much for your comments and we are glad to read that you have positive feedback to our work. We have carefully addressed the concerns raised in your review and made the corresponding revisions to our manuscript. Our Responses are provided in green and revisions in the manuscript are highlighted as blue.

Major Comments

1. In the Introduction, there is a logical gap between the discussion of the limitations of spatial methods and the introduction of deep-learning-based image verification. Please explain more explicitly how the extraction of "high-level abstract features" directly helps address the spatial "double penalty" issue.

Thank you so much for pointing out this logical flaw between the limitations of spatial validation methods and deep-learning-based verification. We have revised the start of paragraph of introducing deep-learning-based verification in lines 66-72: The limitation of existing spatial verification methods essentially stems from their reliance on predefined parameters and rules, preventing them from truly capturing the spatial distributions of observed and forecast precipitation fields. Consequently, conducting PFV from an overall structural perspective promises more reliable results. Inspired by this, we propose a deep-learning-based PFV method that evaluates forecast performance by comparing the high-dimensional features of observed and forecasted precipitations. This approach leverages the exceptional capabilities of deep learning in simulating human cognitive processes and extracting complex features, as well as its remarkable success in image verification practices in recent years.

2. I suggest adding a short subsection, for example, "2.1 Basic Idea," to explicitly present the core logic behind the proposed solution to the scientific gap. Part of the second-to-last paragraph of the Introduction already seems to contain this basic idea.

Thanks a lot for this insightful suggestion, a short introduction of basic idea can help to better understand our proposed verification method. Therefore, we added a paragraph in the beginning of 2 Methodology in lines 88-95: In this section, we present the proposed verification method, named CLPFV, in detail. The core idea of CLPFV is to conduct the verification by shifting from grid-matching to an overall high-level structural similarity comparison through self-supervised contrastive learning. To be more specific, we first used multi-dimensional precipitation augmentations in CLPFV to create intrinsic supervisory signals from unlabeled data to address the scarcity of labeled samples. Subsequently, we designed an improved contrastive loss function that applies proportional penalties to forecast errors when extracting high-level precipitation features, thereby reasonably reflecting the gradient of errors. Finally, the result is directly calculated by comparing the high-dimensional features of observed and forecasted precipitations, achieving the verification from an overall structural perspective.

3. In Section 2, the conceptual framework is somewhat mixed with specific technical implementations (e.g., ResNet-18). In my view, the proposed verification framework does not strictly depend on ResNet-18 or InfoNCE. A brief discussion of the portability of this framework in the Discussion section, such as its applicability to other spatial modeling tasks, would further strengthen the methodological contribution of the paper.

We totally agree with this comment, which highlights the contribution of CLPFV. Just as reviewer suggested, we added a discussion in 4 Conclusion in lines 417-421: Notably, CLPFV also serves as a feasible framework when evaluating the prediction of other meteorological variables or environmental phenomena. Depending on the specific requirements of forecasting or prediction tasks, e.g., PM2.5 forecast and soil mapping, the elements of CLPFV (such as ResNet-18 deep learning model, InfoNCE loss function, and Cosine similarity calculation) can be changed to adapt with various architectures. This flexibility underscores CLPFV's methodological contribution as a generalizable paradigm of forecast verification in the broad earth science field.

4. The simplification of forecast errors into displacement, intensity, and area size is reasonable and useful. However, "area size" may not fully capture all structural errors in real precipitation forecasts. A brief acknowledgement of this limitation would improve the manuscript.

We appreciate the careful consideration behind this recommendation. In the paragraph of introducing displacement, intensity, and area size, we added a brief acknowledgement of this limitation and explained why we conducted the precipitation augmentation in these three aspects, in lines 111-113: While these three aspects fall short of providing a comprehensive depiction of the spatial morphology and structure of precipitation, they offer a straightforward and quantitative assessment of forecast errors and are commonly used in PFV studies (Ebert et al., 2009).

5. The rationale for the quadratic penalty in the improved loss function could be explained more clearly. The current explanation is understandable but somewhat brief. One or two additional sentences on why a quadratic penalty is appropriate here would make the design more transparent.

Thank you for pointing out this issue. We added a brief explanation of quadratic penalty in lines 194-197: Specifically, the quadratic function provides a smooth and differentiable penalty for model training and features extraction. It ensures the forecast verification remains tolerant of minor errors while applying increasingly strict penalties to larger errors, effectively distinguishing different ties of forecast quality.

Minor Comments

6. Several acronyms are used in the Abstract (POD, FAR, TS, FSS, SAL) without prior definition. Please ensure that all abbreviations are spelled out at their first occurrence.

Since it is not necessary to introduce the concrete compared verification methods in the Abstract, we directly deleted abbreviations of POD, FAR, TS, FSS, SAL. Then we thoroughly checked the manuscript, now we ensured that all abbreviations are spelled out at their first occurrence.